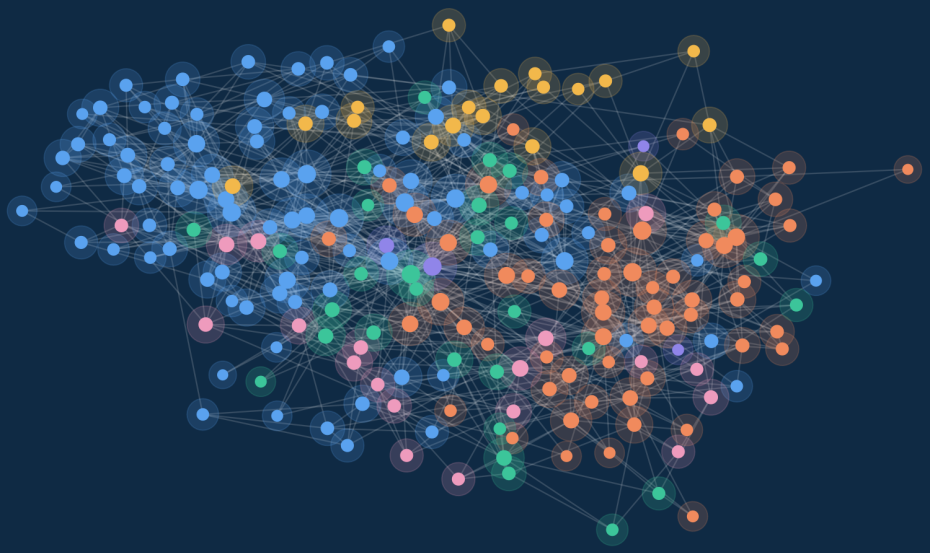
SOCIAL SCIENCE THEORY

A Field Guide to the Theories, Effects, and Frameworks Behind Research and Business



MIKE NGUYEN

150+ theories · open access

Social Science Theory

A Field Guide to the Theories, Effects, and Frameworks Behind Research and Business

Mike Nguyen

Last updated: October 27, 2026

Open access at https://mikenguyen13.github.io/social-theory/

DOI: 10.5281/zenodo.23708720

Licensed under CC BY 4.0

Contents

Welcome **2**
- Who this book is for . . . 2
- How each chapter is built . . . 3
- How to read this book . . . 3
- How to cite . . . 3
- Contributing . . . 4
- License . . . 4
- About the author . . . 4
- Acknowledgments . . . 4

1 Introduction **5**
- 1.1 Why theory matters: in research and in business . . . 5
- 1.2 A map of the theories . . . 6
- 1.3 Finding the right theory for your question . . . 6
- 1.4 How to read this book . . . 7
- 1.5 Conventions . . . 7

I Foundations **10**

2 What Is a Theory? **11**
- 2.1 Why Theory Matters . . . 12
- 2.2 What Theory Is Not . . . 13
- 2.3 The Building Blocks of Theory . . . 14
- 2.4 Types of Theory . . . 19
- 2.5 What Makes a Theory Good? . . . 21
- 2.6 Levels of Analysis . . . 23
- 2.7 How Theories Change . . . 27
- 2.8 Chapter Summary . . . 28
- 2.9 Discussion Questions . . . 29
- 2.10 Further Reading . . . 29

3 Using Theory in Research and Practice **31**
- 3.1 From Theory to Hypotheses . . . 32
- 3.2 Constructs, Measures, and Nomological Networks . . . 34
- 3.3 Mediation and Moderation . . . 35

3.4 Choosing Among Theories . . . 39
3.5 Theoretical Contribution and Modes of Reasoning . . . 41
3.6 Causal Diagrams as Explicit Theory . . . 43
3.7 The Replication Crisis and Research Integrity . . . 45
3.8 Theory in Practice . . . 47
3.9 Chapter Summary . . . 49
3.10 Discussion Questions . . . 50
3.11 Further Reading . . . 50

II Organizations and Strategy 51

4 Resources and Competitive Advantage 52
4.1 The Resource-Based View . . . 53
4.2 The VRIO Framework . . . 61
4.3 The Knowledge-Based View . . . 64
4.4 Resource-Advantage Theory . . . 66
4.5 Market-Based Assets . . . 68
4.6 Chapter Summary . . . 71
4.7 Discussion Questions . . . 72
4.8 Further Reading . . . 72

5 Dynamic Capabilities, Learning, and Renewal 74
5.1 Dynamic Capabilities Framework . . . 75
5.2 Absorptive Capacity . . . 84
5.3 Exploration and Exploitation . . . 86
5.4 Organizational Ambidexterity . . . 89
5.5 Chapter Summary . . . 92
5.6 Discussion Questions . . . 93
5.7 Further Reading . . . 93

6 Governance, Contracts, and Leadership 95
6.1 Agency Theory . . . 96
6.2 Stewardship Theory . . . 103
6.3 Transaction Cost Economics . . . 106
6.4 Upper Echelons Theory . . . 112
6.5 Chapter Summary . . . 115
6.6 Discussion Questions . . . 116
6.7 Further Reading . . . 117

7 Organizations and Their Environments 118
7.1 Resource Dependence Theory . . . 119
7.2 Institutional Theory . . . 124
7.3 Legitimacy . . . 128
7.4 Stakeholder Theory . . . 129
7.5 Organizational Ecology . . . 133
7.6 Contingency Theory . . . 135
7.7 Corporate Social Responsibility Pyramid . . . 138

7.8 Creating Shared Value . . . 140
7.9 Chapter Summary . . . 143
7.10 Discussion Questions . . . 144
7.11 Further Reading . . . 145

8 Information, Signals, and Markets 146
8.1 Information Asymmetry . . . 147
8.2 Signaling Theory . . . 149
8.3 Screening Theory . . . 153
8.4 Costly Signaling . . . 154
8.5 Identity Signaling . . . 155
8.6 Countersignaling . . . 156
8.7 Chapter Summary . . . 157
8.8 Discussion Questions . . . 157
8.9 Further Reading . . . 158

9 Social Exchange and Networks 159
9.1 Social Exchange Theory . . . 160
9.2 Network Exchange Theory . . . 164
9.3 Strength of Weak Ties and Embeddedness . . . 167
9.4 Structural Holes and Brokerage . . . 171
9.5 Social Capital . . . 174
9.6 Commitment-Trust Theory of Relationship Marketing . . . 177
9.7 Chapter Summary . . . 179
9.8 Discussion Questions . . . 180
9.9 Further Reading . . . 181

10 Innovation Diffusion and Technology Adoption 182
10.1 Diffusion of Innovations . . . 183
10.2 The Bass Diffusion Model . . . 188
10.3 Theories of Reasoned Action and Planned Behavior . . . 192
10.4 Technology Acceptance Model . . . 193
10.5 Unified Theory of Acceptance and Use of Technology . . . 195
10.6 Information Systems Success Model . . . 200
10.7 Task-Technology Fit . . . 202
10.8 Chapter Summary . . . 205
10.9 Discussion Questions . . . 206
10.10 Further Reading . . . 207

11 Strategy, Competition, and Industries 208
11.1 Structure-Conduct-Performance . . . 209
11.2 Strategic Groups . . . 213
11.3 Competitive Dynamics . . . 216
11.4 Game Theory . . . 219
11.5 First-Mover Advantage . . . 224
11.6 Real Options Theory . . . 227
11.7 Behavioral Theory of the Firm . . . 231
11.8 Attention-Based View . . . 235

11.9 Chapter Summary . . . 237
11.10 Discussion Questions . . . 238
11.11 Further Reading . . . 239

12 Innovation and Entrepreneurship 240
12.1 Creative Destruction . . . 241
12.2 Disruptive Innovation . . . 244
12.3 Profiting from Innovation . . . 247
12.4 Architectural Innovation . . . 251
12.5 Dominant Design and Technology Cycles . . . 254
12.6 Open Innovation . . . 257
12.7 Effectuation . . . 260
12.8 Opportunity Discovery and Creation . . . 262
12.9 Bricolage . . . 264
12.10 Liability of Newness . . . 265
12.11 Chapter Summary . . . 268
12.12 Discussion Questions . . . 269
12.13 Further Reading . . . 270

13 International Business 271
13.1 Uppsala Model . . . 272
13.2 OLI Paradigm . . . 276
13.3 Internalization Theory . . . 280
13.4 Liability of Foreignness . . . 283
13.5 Institutional Voids . . . 286
13.6 Born Globals . . . 288
13.7 Chapter Summary . . . 291
13.8 Discussion Questions . . . 292
13.9 Further Reading . . . 293

14 Organizations, Teams, and Decision Processes 294
14.1 Garbage Can Model . . . 295
14.2 Escalation of Commitment . . . 298
14.3 Organizational Learning . . . 303
14.4 Knowledge Creation . . . 307
14.5 Organizational Identity . . . 309
14.6 Organizational Culture . . . 311
14.7 Paradox Theory . . . 314
14.8 Psychological Safety . . . 316
14.9 Attraction-Selection-Attrition . . . 319
14.10 Social Information Processing . . . 320
14.11 Sensemaking . . . 321
14.12 Chapter Summary . . . 324
14.13 Discussion Questions . . . 326
14.14 Further Reading . . . 326

15 Work, Jobs, and Leadership 327
15.1 Job Characteristics Model . . . 328

15.2 Job Demands-Resources Theory . . . 333
15.3 Conservation of Resources Theory . . . 337
15.4 Affective Events Theory . . . 339
15.5 Leader-Member Exchange . . . 340
15.6 Transformational Leadership . . . 342
15.7 Path-Goal Theory . . . 345
15.8 Psychological Contract Theory . . . 347
15.9 Organizational Justice . . . 349
15.10 Organizational Commitment . . . 352
15.11 Person-Environment Fit . . . 353
15.12 Human Capital Theory . . . 354
15.13 Tournament Theory . . . 356
15.14 Chapter Summary . . . 359
15.15 Discussion Questions . . . 360
15.16 Further Reading . . . 361

III Markets, Finance, and Economics 362

16 Finance and Corporate Policy 363
16.1 Portfolio Theory . . . 364
16.2 Capital Asset Pricing Model . . . 368
16.3 Efficient Market Hypothesis . . . 372
16.4 The Modigliani-Miller Theorems . . . 375
16.5 Trade-off and Pecking Order Theories . . . 379
16.6 Free Cash Flow Theory . . . 383
16.7 Limits to Arbitrage . . . 386
16.8 Behavioral Finance . . . 390
16.9 Chapter Summary . . . 394
16.10 Discussion Questions . . . 395
16.11 Further Reading . . . 395

17 Economics of Markets and Institutions 396
17.1 Coase Theorem . . . 397
17.2 Collective Action . . . 401
17.3 Property Rights Theory . . . 404
17.4 Search Theory . . . 407
17.5 Efficiency Wages . . . 410
17.6 Network Effects . . . 414
17.7 Two-Sided Markets . . . 417
17.8 Auction Theory . . . 421
17.9 New Institutional Economics . . . 424
17.10 Social Preferences . . . 428
17.11 Chapter Summary . . . 431
17.12 Discussion Questions . . . 432
17.13 Further Reading . . . 433

IV Individuals and Groups **434**

18 Culture and Cross-Cultural Theory **435**
18.1 What Do We Mean by Culture? . . . 436
18.2 Hofstede's Cultural Dimensions Theory . . . 436
18.3 Power Distance Theory . . . 442
18.4 Individualism, Collectivism, and Self-Construal . . . 445
18.5 The GLOBE Project . . . 447
18.6 Schwartz's Theory of Basic Values . . . 449
18.7 Tight and Loose Cultures . . . 452
18.8 Culture in Practice: Negotiation and Multicultural Teams . . . 453
18.9 Critiques and the Research Frontier . . . 454
18.10 Chapter Summary . . . 455
18.11 Discussion Questions . . . 456
18.12 Further Reading . . . 456

19 Motivation **457**
19.1 Content and Process Theories . . . 458
19.2 Expectancy Theory . . . 459
19.3 Equity Theory . . . 462
19.4 Goal-Setting Theory . . . 465
19.5 The Goal-Gradient Hypothesis . . . 468
19.6 Self-Determination Theory . . . 470
19.7 Losing Leads to Winning . . . 473
19.8 Integrating the Theories . . . 475
19.9 Chapter Summary . . . 475
19.10 Discussion Questions . . . 476
19.11 Further Reading . . . 477

20 Self-Regulation and Construal **478**
20.1 Regulatory Focus Theory . . . 479
20.2 Regulatory Fit . . . 483
20.3 Construal Level Theory . . . 486
20.4 Self-Control and Intertemporal Choice . . . 490
20.5 Integrating the Theories . . . 495
20.6 Chapter Summary . . . 495
20.7 Discussion Questions . . . 496
20.8 Further Reading . . . 496

21 Attitudes, Attribution, and Social Learning **498**
21.1 Cognitive Dissonance Theory . . . 499
21.2 Self-Perception Theory . . . 505
21.3 Attribution Theory . . . 508
21.4 Balance Theory . . . 513
21.5 Social Cognitive Theory . . . 515
21.6 Reactance Theory . . . 519
21.7 Chapter Summary . . . 523
21.8 Discussion Questions . . . 524

21.9 Further Reading 524

22 Social Comparison, Identity, and Influence **525**
22.1 Social Comparison Theory 526
22.2 Social Identity Theory 528
22.3 Optimal Distinctiveness Theory 530
22.4 Conformity 533
22.5 Social Norms 535
22.6 Social Facilitation and Inhibition 538
22.7 Distraction-Conflict Theory 541
22.8 Spiral of Silence 543
22.9 Chapter Summary 546
22.10 Discussion Questions 547
22.11 Further Reading 547

23 Judgment and Decision Making **548**
23.1 Bounded Rationality 549
23.2 Heuristics and Biases 552
23.3 Prospect Theory in Brief 554
23.4 Endowment Effect 555
23.5 Mental Accounting 557
23.6 Dual-Process Theory 559
23.7 Feelings-as-Information Theory 561
23.8 Processing Fluency 563
23.9 Discrepancy-Attribution Hypothesis 565
23.10 Choice Architecture and Defaults 567
23.11 Chapter Summary 570
23.12 Discussion Questions 571
23.13 Further Reading 571

24 Prospect Theory and Reference-Dependent Preferences **572**
24.1 Expected Utility Theory 573
24.2 Prospect Theory: The 1979 Model 577
24.3 Cumulative Prospect Theory 582
24.4 Loss Aversion 588
24.5 Reference-Dependent Preferences 593
24.6 Status Quo Bias 598
24.7 Myopic Loss Aversion 600
24.8 Disposition Effect 603
24.9 Further Applications 606
24.10 Replication and Robustness 610
24.11 Chapter Summary 611
24.12 Discussion Questions 612
24.13 Further Reading 613

25 Emotion and Well-Being **614**
25.1 Broaden-and-Build Theory 615
25.2 Hedonic Adaptation 618

25.3 The Affect Circumplex . . . 621
25.4 Emotion and Virality . . . 623
25.5 Appraisal Theory . . . 625
25.6 Money and Happiness . . . 628
25.7 Peak-End Rule . . . 630
25.8 Chapter Summary . . . 632
25.9 Discussion Questions . . . 632
25.10 Further Reading . . . 633

26 Persuasion and Communication 634
26.1 The Elaboration Likelihood Model . . . 635
26.2 The Persuasion Knowledge Model . . . 641
26.3 Word of Mouth and Social Transmission . . . 643
26.4 Microblogging and Undirected Communication . . . 648
26.5 Agenda Setting . . . 649
26.6 Framing Theory . . . 651
26.7 Uses and Gratifications . . . 654
26.8 Chapter Summary . . . 656
26.9 Discussion Questions . . . 657
26.10 Further Reading . . . 657

V Consumers and Marketing 658

27 Consumer Behavior and Branding 659
27.1 Expectancy-Disconfirmation Theory . . . 661
27.2 Means-End Chain Theory . . . 665
27.3 Customer-Based Brand Equity . . . 667
27.4 Brand Personality . . . 671
27.5 Consumer-Brand Relationships . . . 674
27.6 Self-Congruity Theory . . . 677
27.7 Schema Congruity Theory . . . 680
27.8 Hedonic and Utilitarian Consumption . . . 683
27.9 Psychological Ownership . . . 687
27.10 Variety Seeking . . . 689
27.11 Reference Price Theory . . . 691
27.12 Consumer Culture Theory . . . 694
27.13 Chapter Summary . . . 697
27.14 Discussion Questions . . . 698
27.15 Further Reading . . . 698

28 Marketing Strategy and Services 700
28.1 Market Orientation . . . 701
28.2 Service-Dominant Logic . . . 704
28.3 The Gaps Model of Service Quality . . . 708
28.4 Customer Equity and Customer Lifetime Value . . . 713
28.5 Hierarchy of Effects . . . 717
28.6 Product Life Cycle . . . 720

28.7 Marketing Mix . . . 723
28.8 Chapter Summary . . . 725
28.9 Discussion Questions . . . 726
28.10 Further Reading . . . 727

VI Effects, Phenomena, and Frameworks 728

29 Named Effects in Behavioral Science 729
29.1 Mere Exposure Effect . . . 730
29.2 Warm-Glow Heuristic . . . 733
29.3 Goldilocks Effect . . . 734
29.4 Coolidge Effect . . . 736
29.5 Bandwagon, Snob, and Veblen Effects . . . 737
29.6 Glass Cliff Effect . . . 739
29.7 IKEA Effect . . . 742
29.8 Decoy Effect . . . 743
29.9 Anchoring Effect . . . 746
29.10 Beauty in Averageness . . . 748
29.11 Pratfall Effect . . . 750
29.12 Using Named Effects Responsibly . . . 751
29.13 Chapter Summary . . . 751
29.14 Discussion Questions . . . 752
29.15 Further Reading . . . 753

30 Social Phenomena 754
30.1 Similarity Asymmetry . . . 755
30.2 Homophily . . . 757
30.3 Assortative Mating . . . 760
30.4 Social Magnet . . . 761
30.5 Wisdom of the Crowd . . . 762
30.6 Groupthink . . . 766
30.7 Pack Journalism . . . 767
30.8 Matthew Effect . . . 769
30.9 Social Loafing . . . 771
30.10 Bystander Effect . . . 772
30.11 Tragedy of the Commons . . . 774
30.12 Scent and Emotion . . . 776
30.13 Bullwhip Effect . . . 777
30.14 Chapter Summary . . . 781
30.15 Discussion Questions . . . 782
30.16 Further Reading . . . 782

31 Models and Frameworks 783
31.1 Information Processing Model . . . 784
31.2 Geneplore Model . . . 788
31.3 Combinatorial Creativity . . . 790
31.4 Porter's Five Forces . . . 793

31.5 Stimulus-Organism-Response Framework 797
31.6 A Mechanism Toolkit for Social Science Theory 799
31.7 Chapter Summary 803
31.8 Discussion Questions 804
31.9 Further Reading 804

VII Reference **806**

32 Theory Atlas **807**
32.1 Searchable atlas 807
32.2 Theories by mechanism 814
32.3 Glossary 815

References **818**

Welcome

Social Science Theory is an open-access field guide to the ideas that social scientists and evidence-minded managers use to explain behaviour: of people, teams, firms, markets, and societies. It covers more than 150 theories, and more than 200 entries in all once named effects, phenomena, and frameworks are included, drawn from management, marketing, finance, economics, psychology, sociology, communication, and information systems.

Every chapter is written to be useful twice: once for **research** (what the theory claims, why, under which conditions, how it has been tested, and how to derive hypotheses from it) and once for **practice** (where the theory shows up in real business decisions, and what it implies for strategy, marketing, finance, operations, and people management).

Table 1. What is inside the book.

Part	Chapters	Entries	What you will find
Foundations	1-3	4	What theory is, how to evaluate it, and how to turn it into hypotheses
Organizations & Strategy	4-15	88	Resources, capabilities, governance, institutions, information, networks, innovation, strategy, international business, organizations, and leadership
Markets, Finance & Economics	16-17	18	Finance, capital markets, institutions, platforms, auctions, and social preferences
Individuals & Groups	18-26	60	Culture, motivation, self-regulation, attitudes, social influence, decisions, prospect theory, emotion, and persuasion
Consumers & Marketing	27-28	19	Consumer behavior, branding, services, customer value, and marketing strategy
Effects, Phenomena & Frameworks	29-31	29	Named effects, social phenomena, classic models, and a cross-theory mechanism toolkit

Who this book is for

- **Doctoral students and early-career researchers** who need a systematic map of the theories their field "just knows", and a fast way to find the right lens for a research question.
- **Faculty** looking for a free, citable companion text for seminars in management, marketing, organizational behaviour, and consumer behaviour.
- **Practitioners**, strategists, marketers, product managers, analysts, HR leaders, and consultants, who want to understand *why* interventions work, not just *that* they work.

How each chapter is built

Theories are presented in a consistent format so they can be compared at a glance:

THEORY CARD

Core claim: the theory in one or two sentences. **Origins**, who proposed it, when, and in which discipline. **Level of analysis**: individual, dyad, group, firm, market, or institution. **Key constructs** and **boundary conditions**. **Related theories**: links to neighbouring entries.

Each card is followed by the theory's **core logic and propositions**, a **mechanism diagram**, a review of **evidence and extensions**, a balanced discussion of **critiques** (including replication status where relevant), one or more **business cases**, and a **research toolkit** with sample hypotheses, measures, designs, and data sources. Chapters close with key takeaways, a comparison table, discussion questions, and annotated further reading.

Look for these boxes throughout:

BUSINESS CASE

Applications drawn from strategy, marketing, finance, operations, HR, and policy.

RESEARCH TOOLKIT

Sample hypotheses, operationalizations, research designs, and data sources.

DEBATE AND EVIDENCE

Critiques, contested findings, and replication evidence.

COMMON PITFALL

The most common ways a theory is misunderstood or misapplied.

How to read this book

You do not need to read it in order. If you have a specific theory in mind, use the search bar or jump straight to the Theory Atlas, a searchable table of every entry with its level of analysis, mechanism, key source, and business use. If you want the big picture, start with the Introduction and the interactive theory map. If you are writing a paper, Chapters 2 and 3 show how to move from a theory to testable hypotheses.

How to cite

If this book is useful in your work, please cite it:

> Nguyen, M. (2026). *Social Science Theory: A Field Guide to the Theories, Effects, and Frameworks Behind Research and Business*. Zenodo. https://doi.org/10.5281/zenodo.23708720

```
@book{nguyen_social_theory,
  author    = {Nguyen, Mike},
  title     = {Social Science Theory: A Field Guide to the Theories, Effects,
               and Frameworks Behind Research and Business},
  year      = {2026},
  publisher = {Zenodo},
  doi       = {10.5281/zenodo.23708720},
  url       = {https://mikenguyen13.github.io/social-theory/}
}
```

The DOI always resolves to the latest edition. Each archived release on Zenodo also has its own DOI if you need to cite a specific version.

Contributing

This is a living book. Corrections, new theories, better examples, and translations are welcome. Every page has an **Edit this page** link that opens the source on GitHub. You can also open an issue to suggest a theory. Contributors should follow the style guide. If the book helps you, starring the GitHub repository helps others find it.

License

The text, figures, and tables of this book are licensed under the Creative Commons Attribution 4.0 International License (CC BY 4.0). You may share and adapt the material for any purpose, including commercially, provided you give appropriate credit, link to the license, and indicate if changes were made. The code that builds the book is released under the MIT License.

About the author

Mike Nguyen is the author of the open-source *A Guide on Data Analysis*, a companion text on statistical methods and causal inference for social science research. This book began as his own systematic notes on the theories that researchers in business and the social sciences are expected to know, and has grown into a reference for students, scholars, and practitioners.

Acknowledgments

The book is written in R Markdown and built with **bookdown** (Xie 2025a), **knitr** (Xie 2025b), and **rmarkdown** (Allaire et al. 2025). Figures use **ggplot2** (Wickham et al. 2025) and **DiagrammeR** (Iannone and Roy 2024). Tables use **kableExtra** (H. Zhu 2024). Interactive maps and tables use **visNetwork** (Almende B.V. and Contributors and Thieurmel 2025) and **DT** (Xie et al. 2025).

Chapter 1

Introduction

Doing social science research requires knowing a great many theories, and knowing which one fits which problem. For a new doctoral student this can be overwhelming. Senior colleagues seem to speak theory as a second language: "that's an agency problem", "it's a signaling story", "sounds like optimal distinctiveness". The standard advice is that theories are absorbed gradually, one paper at a time. That advice is sound, but slow and unsystematic, and it leaves gaps that only become visible when a reviewer points them out.

This book takes the opposite approach. It organizes the theories most often used in business and the social sciences into a single, consistent reference, so that a reader can (1) find the right theory quickly, (2) understand its logic well enough to use it correctly, and (3) see how it connects to its neighbours. It began as the author's personal notes and grew into a field guide meant for anyone who needs theory to do serious work: in a dissertation, a journal article, a classroom, or a boardroom.

1.1 Why theory matters: in research and in business

A **theory** is a set of statements about how and why a set of constructs are related, and under what conditions (Chapter 2). Theory does three jobs that data alone cannot do:

1. **It tells you where to look.** Every dataset contains more correlations than anyone could test. Theory narrows the search to relationships that should exist *if* a particular mechanism operates.
2. **It explains why.** An observed effect without a mechanism cannot be generalized. Knowing that a warranty raises sales is useful. Knowing that it does so because it is a costly signal (Section 8.2) tells you when it will stop working (when competitors can match it cheaply).
3. **It travels.** Mechanisms transfer across settings. The logic of adverse selection (Section 8.1) explains used cars, insurance, hiring, and venture capital alike.

Managers use theory constantly, usually without naming it. A pricing manager who offers a "good-better-best" menu is applying screening (Section 8.3). An HR director worried that a bonus will undermine intrinsic interest is invoking self-determination theory (Section 19.6). A utility that tells households how their energy use compares with their neighbours' is using social norms (Section 22.5). Making these theories explicit lets decision makers anticipate side effects and boundary conditions *before* they become expensive surprises.

BUSINESS CASE

Theory as a decision tool. Consider a subscription business deciding whether to offer a free trial. Four theories in this book make different predictions: the endowment effect (Section 23.4) suggests trial users will value the product more once they "own" it. Hedonic adaptation (Section 25.2) warns that novelty fades. Signaling theory (Section 8.2) suggests that a long free trial signals confidence in quality, and choice architecture (Section 23.10) highlights that whether the trial converts *by default* will dominate the outcome. A manager who knows all four can design an experiment that tells them apart, and a researcher can turn the same contrast into a paper.

1.2 A map of the theories

The theories in this book are not isolated islands. Many share mechanisms (for example, *information*, *incentives*, *identity*, or *power*), and many were developed in dialogue with one another. Figure 1.1 shows every entry in the book as a node, coloured by part of the book. Edges connect theories that cross-reference each other in the text. In the online edition the map is interactive: hover to see a theory's core claim, click to highlight its neighbours, and use the menu to select a theory.

Figure 1.2 cross-tabulates the entries by **level of analysis** and **dominant mechanism**. Two patterns stand out. First, most organization-level theories explain outcomes through *resources*, *incentives*, *power*, or *legitimacy (Section @ref(legitimacy (Section 7.3)))*, while individual-level theories rely on *cognition*, *emotion*, *motivation*, and *identity*. Second, a handful of mechanisms, especially information and social comparison, work at every level, which is why theories built on them (signaling, social comparison, social norms) travel so widely across disciplines.

1.3 Finding the right theory for your question

Researchers usually start from a phenomenon, not a theory. Table 1.1 works backwards from common research and managerial questions to the families of theory that address them.

Table 1.1. From research question to theory: a quick-start guide.

If your question is about...	Start with	Chapter
Why some firms persistently outperform others	Resource-based view, dynamic capabilities	4-5
Whether to make, buy, or partner	Transaction cost economics, resource dependence	6-7
Aligning managers or employees with owners	Agency theory, stewardship theory	6
Why firms in a field look alike	Institutional theory, legitimacy	7
Communicating quality that buyers cannot see	Signaling, screening, information asymmetry	8
Who gets information, power, and opportunities	Social exchange, weak ties, structural holes	9
How new products and technologies spread	Diffusion of innovations, Bass model, TAM/UTAUT	10
How firms compete and respond to rivals	Structure-conduct-performance, competitive dynamics, game theory	11
Why incumbents fail and start-ups succeed	Disruptive innovation, profiting from innovation, effectuation	12
How firms expand across borders	Uppsala model, OLI paradigm, institutional voids	13
How organizations decide, learn, and make sense of events	Garbage can, escalation of commitment, sensemaking	14

Table 1.1. From research question to theory: a quick-start guide. *(continued)*

If your question is about...	Start with	Chapter
Job design, engagement, and leadership	Job demands-resources, LMX, transformational leadership	15
Capital structure, valuation, and investor behaviour	Modigliani-Miller, efficient markets, behavioral finance	16
Platforms, externalities, and fairness in markets	Network effects, two-sided markets, social preferences	17
Adapting strategy or messages across countries	Hofstede, GLOBE, self-construal	18
Designing incentives, goals, and rewards	Expectancy, equity, goal setting, self-determination	19
Framing messages for different audiences	Regulatory focus and fit, construal level	20
Changing attitudes and explaining blame	Cognitive dissonance, attribution, reactance	21
Using peers, norms, and identity to change behaviour	Social comparison, social norms, optimal distinctiveness	22
Pricing, defaults, and seemingly irrational choices	Mental accounting, choice architecture, heuristics	23
Gains, losses, risk, and reference points	Prospect theory, loss aversion, status quo bias	24
Customer experience, well-being, and virality	Peak-end rule, hedonic adaptation, emotion and virality	25
Persuasion, word of mouth, and media effects	ELM, persuasion knowledge, social transmission	26
Satisfaction, brands, and consumer identity	Expectancy-disconfirmation, brand equity, self-congruity	27
Customer value, services, and marketing strategy	Market orientation, service quality, customer equity	28

1.4 How to read this book

- **Looking up one theory?** Search the book or use the Theory Atlas. Each entry opens with a *theory card* that summarises the essentials in under a minute.
- **Learning a field?** Read a whole chapter. Chapters group theories that are usually compared with each other and end with a comparison table and discussion questions suitable for seminars.
- **Writing a paper?** Read Chapters 2 and 3 first, then use the *research toolkit* box in each entry for sample hypotheses, measures, and designs.
- **Making a decision?** Go straight to the *business case* boxes, then read the *pitfall* and *debate* boxes to understand where the theory breaks down.

> **COMMON PITFALL**
>
> **Theory shopping.** It is tempting to pick a theory *after* the results are in and choose whichever one "explains" them. This is a form of HARKing, hypothesizing after the results are known, and it produces fragile findings (Chapter 3). Choose theories for their mechanisms before collecting data, and, where possible, pit two theories with different predictions against each other.

1.5 Conventions

Citations follow author-year style and link to the full reference list. Figures labelled *Illustrative* or *Stylised* are built from simulated or schematic data to show a theory's logic. They are not empirical results. All code used to make the figures is in the book's GitHub repository, so every figure can be

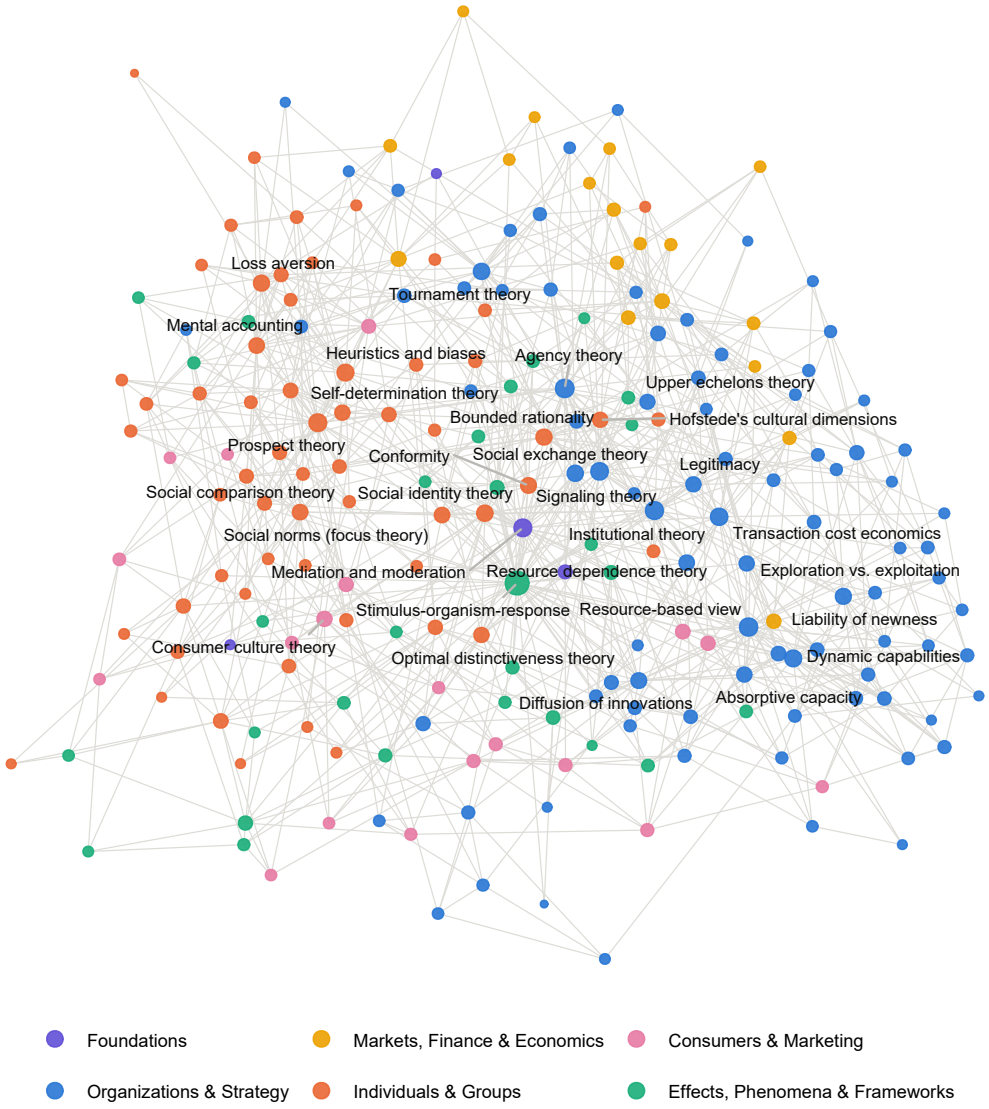


Figure 1.1. The theory map: each node is an entry in the book, coloured by part. Edges connect entries that cross-reference each other. The 30 most connected entries are labelled. The online edition is interactive.

reproduced and adapted.

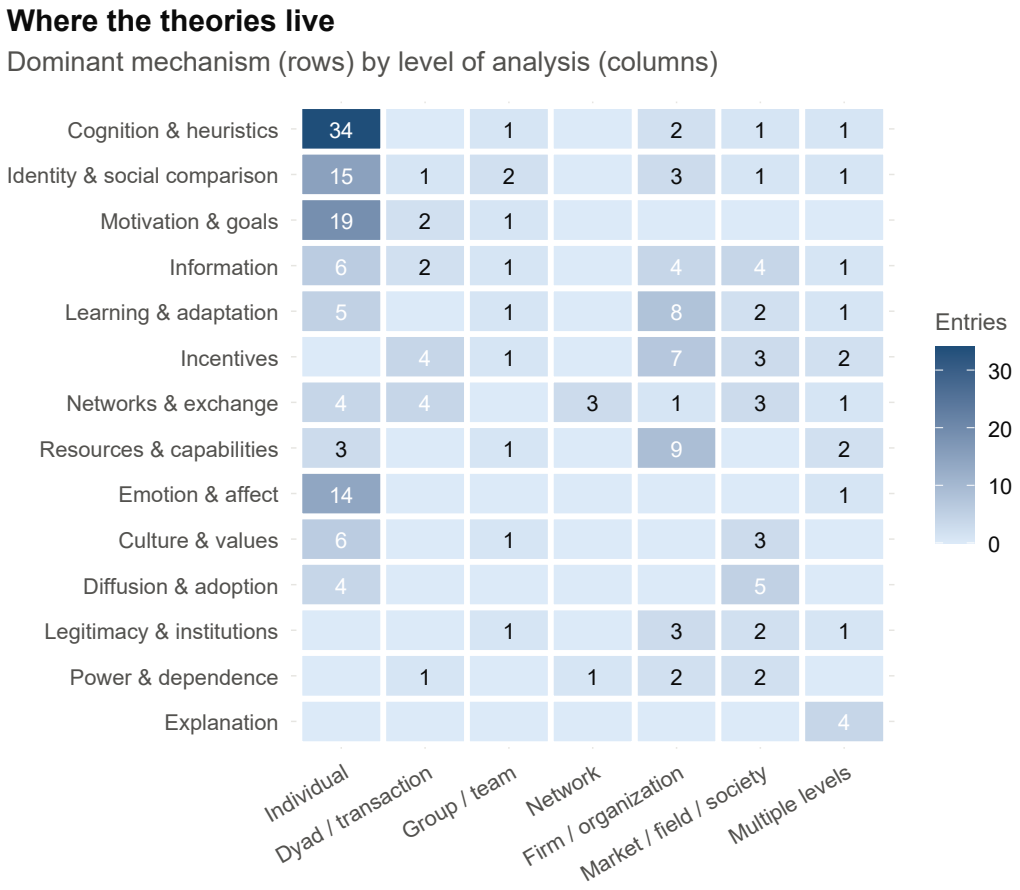


Figure 1.2. Entries in the book by level of analysis and dominant mechanism. Darker cells contain more theories.

Part I

Foundations

Chapter 2

What Is a Theory?

A regional grocery chain notices that its best-performing stores employ more staff per square metre than its worst-performing stores. The operations team proposes adding staff everywhere. The finance team objects: perhaps busy stores were *given* more staff because they were already busy. A third executive suggests that the real driver is the store manager: good managers both hire well and sell more. A fourth points out that the comparison mixes urban and rural stores, which differ in almost everything. Four people, one correlation, four different accounts of *what causes what and why*. Each account is a theory, and the decision the chain makes will depend on which theory it acts on.

This chapter is about what such accounts are, what distinguishes a good one from a bad one, and how scholars and practitioners can tell the difference. We begin with what theory is, and, just as importantly, what it is not. We then take theories apart into their **building blocks** (constructs, relationships, mechanisms, and boundary conditions), distinguish **types** of theory, set out the **criteria** by which theories are judged, examine the **levels of analysis** at which theories operate, and close with how theories change over time. Along the way we argue a simple thesis: managers do not choose between "theory" and "practice". They choose between theories that are explicit, tested, and bounded and theories that are implicit, untested, and unbounded (Table 2.1).

Table 2.1. Chapter at a glance: the anatomy and evaluation of theory.

Concept	Core question	Key sources
What theory is (and is not)	What separates a theory from data, lists, diagrams, or hypotheses?	Sutton & Staw (1995), Weick (1995), DiMaggio (1995)
Building blocks	What, how, why, and who/where/when?	Whetten (1989), Bacharach (1989), Dubin (1978)
Mechanisms	Through what process does a cause produce its effect?	Hedstrom & Swedberg (1998), Coleman (1990)
Types of theory	Does the theory analyse, explain, predict, or prescribe?	Merton (1968), Gregor (2006)
Criteria for good theory	Is it falsifiable, useful, parsimonious, general, and interesting?	Popper (1959), Davis (1971), Corley & Gioia (2011)
Levels of analysis	At what level do constructs live, and can findings cross levels?	Rousseau (1985), Klein, Dansereau & Hall (1994)
Theory change	How do theories and paradigms rise and fall?	Kuhn (1962), Lakatos (1970)

2.1 Why Theory Matters

The word *theory* suffers in everyday speech. "That's just a theory" suggests speculation. "In theory, yes, but in practice …" suggests irrelevance. In the sciences the word means something closer to the opposite: a theory is the most disciplined form of understanding available, an account that has been stated precisely enough to be checked and has survived attempts to show it wrong.

DEFINITION

Definition 2.1. Theory. A coherent set of statements about the relationships among constructs, together with an explanation of *why* those relationships hold and a specification of the conditions under which they are expected to hold. In Bacharach's formulation, a theory is a system of constructs and variables in which constructs are related by propositions and variables by hypotheses, the whole bounded by the theorist's assumptions (Bacharach 1989). In everyday managerial language, Christensen and Raynor (2003) offer a compact equivalent: a theory is a statement of what causes what, and why.

2.1.1 Managers run on implicit theories

Every managerial decision embeds a causal claim. A sales incentive presumes that pay shapes effort. A rebrand presumes that perceptions drive choice. A reorganization presumes that structure shapes coordination. McGregor (1960) made the point vividly half a century ago: managers who assume that people dislike work and must be controlled ("Theory X") design very different organizations from managers who assume that people seek responsibility under the right conditions ("Theory Y"). Neither manager thinks of himself or herself as a theorist, yet each acts on a theory of human motivation, and each theory tends to produce the behaviour it predicts.

Christensen and Raynor (2003) argue that executives should care about management theory precisely because they cannot avoid using it. Their critique targets the "best-practice" habit: copying what successful firms do without understanding the circumstances in which those practices work. Early attempts at human flight, they observe, imitated the *attributes* of birds, feathers and flapping wings, rather than the underlying mechanism of lift. Management advice built on attributes ("successful companies have flat structures") fails in the same way when it is transplanted to circumstances where the mechanism does not operate. Good theory, in their account, progresses from describing phenomena, to sorting them into categories defined by *circumstances* rather than surface attributes, to stating causal relationships that hold within each category, and it improves fastest when someone finds an anomaly the existing theory cannot explain.

2.1.2 Evidence-based management

The **evidence-based management** movement makes the same argument from a different direction. Rousseau (2006) defines it as translating principles based on the best available evidence into organizational practice, and contrasts it with decisions grounded in personal preference, imitation, and unsystematic experience. Pfeffer and Sutton (2006) document how often popular management practices, forced ranking, merger strategies, pay schemes, are adopted on the strength of anecdote or fashion rather than evidence. Evidence-based management is not anti-theoretical. It depends on theory to tell managers which evidence is relevant to their circumstances and why a finding from one setting should (or should not) generalize to another.

BUSINESS CASE

Open-plan offices and the theory of collaboration. Many firms moved to open-plan offices on the strength of an implicit theory: removing walls increases chance encounters, encounters increase face-to-face interaction, and interaction improves collaboration. Bernstein and Turban (2018) tested the second link directly. Using wearable sensors and electronic communication records in two corporate headquarters that switched to open-plan layouts, they found that face-to-face interaction *fell* sharply after the change while email and messaging rose. A plausible mechanism is that employees seek privacy when visual exposure increases. The case illustrates a general lesson: an intuitively appealing causal chain is a theory, and like any theory it can be wrong in one link even when the other links are plausible.

2.2 What Theory Is Not

A useful way to understand theory is through its imitations. In a widely cited editorial essay, R. I. Sutton and Staw (1995) identified five things that are often presented as theory in management manuscripts but are not, on their own, theory. Each is a legitimate *part* of a theoretical argument. None can substitute for it (Table 2.2).

Table 2.2. Five things that are not theory on their own, following Sutton and Staw (1995).

Not theory on its own	Why it falls short	What turns it into theory
References	Citing prior work signals which literature a paper belongs to but does not explain anything	State the logic the cited work supplies and how it applies here
Data	Data describe which patterns were observed, not why they arose	Explain the process that generated the pattern
Lists of variables or constructs	Naming predictors says what matters, not how or why	Specify relationships and the mechanism linking them
Diagrams	Boxes and arrows show structure but rarely the reasoning behind each arrow	Accompany each arrow with an explicit causal argument
Hypotheses or predictions	Hypotheses state what is expected, not why	Derive hypotheses from the underlying logic and its assumptions

The common thread is the missing *why*. A paper can be well referenced, rich in data, and full of diagrams and hypotheses, yet still fail to tell the reader the causal process that connects one construct to another, or why that process should operate in some settings but not others. Strong theory, in Sutton and Staw's account, delves into underlying processes, explores the connections among phenomena, and tells a story about why acts, events, structures, and thoughts occur.

DEBATE AND EVIDENCE

Theory as product or as process? Sutton and Staw's essay prompted two influential replies in the same issue of *Administrative Science Quarterly*. Weick (1995b) argued that theory is better seen as a continuum than as a binary: references, data, lists, diagrams, and hypotheses are often the visible traces of **theorizing**, an ongoing process whose products are approximations. Judging only finished theory, he warned, may discourage the interim work from which theory grows. DiMaggio (1995) distinguished three conceptions of theory, theory as *covering laws* (general statements that subsume particulars), theory as *enlightenment* (a way of seeing

that reveals what was hidden), and theory as *narrative* (an account of how a process unfolds), and argued that the most valuable theories combine elements of all three. The exchange remains a useful reminder that standards for theory depend partly on the stage of a research programme. Weick's earlier essay describes theory construction itself as "disciplined imagination": generating many conjectures and selecting among them with explicit criteria (Weick 1989).

2.3 The Building Blocks of Theory

If theory is not a list, a diagram, or a hypothesis, what is it made of? The most widely used answer in management comes from Whetten (1989), who proposed that a complete theory answers four questions. Earlier, Dubin (1978) had laid out a similar architecture of units, laws of interaction, boundaries, and system states, and Bacharach (1989) sharpened the distinction between the abstract and the observable parts of a theory.

2.3.1 What, how, why, and who/where/when

Table 2.3 presents the whetten's (1989) four elements of a complete theory, illustrated with resource-based theory (Section 4.1) (Wernerfelt, 1984, Barney, 1991).

Table 2.3. Whetten's (1989) four elements of a complete theory, illustrated with resource-based theory (Wernerfelt, 1984, Barney, 1991).

Element	Question	What it supplies	Illustration: resource-based theory
What	Which factors (constructs) logically belong in the explanation?	Comprehensiveness balanced against parsimony	Firm resources and capabilities, competitive advantage
How	How are the factors related?	Direction, form, and often causality of relationships	Resources that are valuable, rare, and costly to imitate lead to sustained advantage
Why	What psychological, economic, or social dynamics justify the relationships?	The mechanism, the theory's explanatory core	Imperfect imitability (causal ambiguity, path dependence, social complexity) blocks competitive erosion
Who / where / when	To which actors, settings, and times does the theory apply?	Boundary conditions, limits of generalizability	Markets with resource heterogeneity and imperfect factor mobility. Advantage can erode when environments shift

Whetten placed particular weight on the *why*. The *what* and *how* together describe a theory's domain, they could be read off a diagram, but the *why* supplies the reasons a reader should believe the diagram, and it is the element reviewers most often find missing. Resource-based theory illustrates the four elements well: Wernerfelt (1984) and Jay Barney (1991) identify the constructs (resources and competitive advantage), the relationship (certain resources produce sustained advantage), the mechanism (imitation is blocked), and the conditions (resource heterogeneity and immobility). Later work refined each element (J. Barney and A. Arikan 2001, J. B. Barney, Ketchen, and Wright 2011, Kozlenkova, Samaha, and Palmatier 2013), a pattern typical of productive theories.

2.3.2 Constructs and variables

DEFINITION

Definition 2.2. Construct. An abstract concept, deliberately invented or adopted for a scientific purpose, that cannot be observed directly (e.g., *trust*, *brand equity*, *absorptive capacity (Section 5.2)*). **Variable.** An observable, measurable operationalization of a construct (e.g., a five-item trust scale, a price premium, R&D intensity). **Measure.** The concrete instrument or data source that produces values of the variable.

Bacharach's distinction between the two levels of a theory is essential for reading and writing research (Bacharach 1989). **Constructs** are related to one another by **propositions**. **Variables** are related to one another by **hypotheses**. A proposition ("customer trust increases loyalty") is a claim about the world at the level of concepts. A hypothesis ("respondents' scores on the trust scale predict repurchase within twelve months") is a claim about measured data. Because the link from construct to variable is never perfect, every empirical test is simultaneously a test of the theory *and* of the measurement assumptions: a point that returns in Chapter 3 under the heading of construct validity.

Suddaby (2010) argues that many theories in management suffer from poor **construct clarity**: constructs are defined loosely, overlap with neighbouring constructs, or shift meaning across studies. Clear constructs have precise definitions, a stated scope (the conditions under which the construct applies), explicit relationships to other constructs, and coherence with the theory's broader logic. Table 2.4 illustrates the ladder from construct to measure across business functions.

Table 2.4. From construct to variable to measure: the two levels of a theory (following Bacharach, 1989).

Function	Construct	Variable	Example measure
Marketing	Brand equity	Price premium over a private-label equivalent	Scanner-panel prices for matched products
Strategy	Absorptive capacity	R&D intensity, prior related knowledge	R&D spend divided by sales
HR / OB	Psychological safety	Team members' perception that interpersonal risk is safe	Multi-item survey scale aggregated to the team
Finance	Information asymmetry	Dispersion of analysts' earnings forecasts	Standard deviation of forecasts for a firm-year
Operations	Process flexibility	Number of products a plant can switch among	Plant-level product-assignment records
Information systems	Perceived usefulness	User beliefs that a system improves performance	Validated survey items

2.3.3 Propositions, hypotheses, and assumptions

Propositions and hypotheses differ in abstraction, not in importance. A theory paper may advance only propositions. An empirical paper translates them into hypotheses that a particular dataset can test. Both rest on **assumptions**: statements the theorist takes as given rather than tests. Bacharach (1989) emphasizes that the theorist's implicit values and the spatial and temporal bounds of a theory are among its most consequential assumptions. Agency theory (Section 6.1), for instance, assumes self-interested agents and goal conflict. Stewardship theory relaxes the first assumption and therefore predicts different governance arrangements (see Agency theory (Section 6.1) and Stew-

ardship theory (Section 6.2)).

Figure 2.1 assembles these building blocks into one picture, using a signaling example developed fully in Chapter 8. The upper row is the theory proper. The lower row is its empirical shadow.

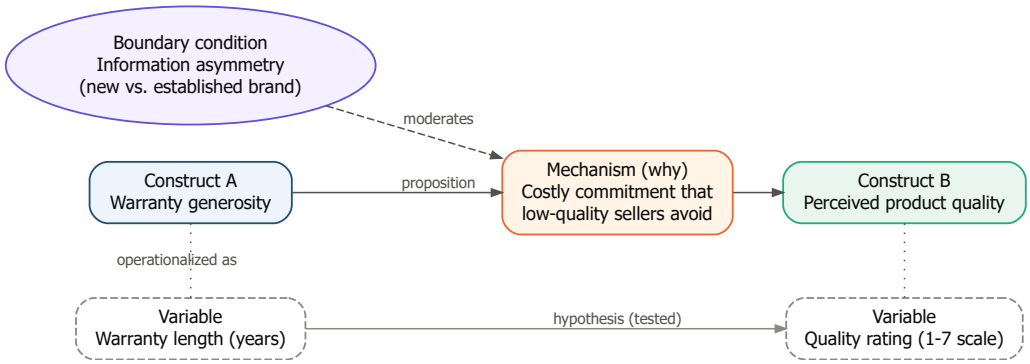


Figure 2.1. The anatomy of a theory. Constructs (upper row) are linked by propositions and a mechanism. Boundary conditions moderate the mechanism. Each construct is operationalized by a measured variable (lower row), and variables are linked by testable hypotheses. Example drawn from signaling theory (Spence, 1973, Connelly et al., 2011).

2.3.4 Mechanisms: opening the black box

A theory that says *X leads to Y* without saying how is a **black-box** theory. A **mechanism** is the set of entities and activities, organized in a particular way, that brings about the effect (Hedström and Swedberg 1998). Mechanism-based explanation became a programme in analytical sociology because correlations between macro variables, religion and economic development, inequality and crime, firm size and innovation, say little about the processes that generate them. Specifying the mechanism makes a theory more credible (because the process is plausible), more testable (because intermediate steps can be measured), and more useful (because managers can intervene on the process itself).

THEORY CARD

Core claim. Macro-level associations should be explained by showing how macro conditions shape individual situations, how individuals act in those situations, and how individual actions aggregate into macro outcomes.
Origins. Sociology. Coleman (1990), reconstructing Max Weber's argument linking Protestant religious doctrine to the rise of capitalism. Elaborated as a typology of social mechanisms by Hedström and Swedberg (1998).
Level of analysis. Explicitly multi-level: macro (society, market, organization) and micro (individual).
Key constructs. Situational mechanism (macro to micro), action-formation mechanism (micro to micro), transformational mechanism (micro to macro).
Boundary conditions. Most useful when the outcome of interest is a collective property produced by many individual actions. Less informative for purely individual-level phenomena.
Related theories. Social exchange theory (Section 9.1), Institutional theory (Section 7.2), Wisdom of the crowd (Section 30.5).

Coleman's diagram, often called **Coleman's boat** or the "bathtub", is redrawn in Figure 2.2. The dashed line across the top is the observed macro association. Coleman argued that this line is not itself an explanation. The explanation runs down, across, and back up. Hedström and Swedberg label the three arrows the **situational**, **action-formation**, and **transformational** mechanisms (Hedström and Swedberg 1998).

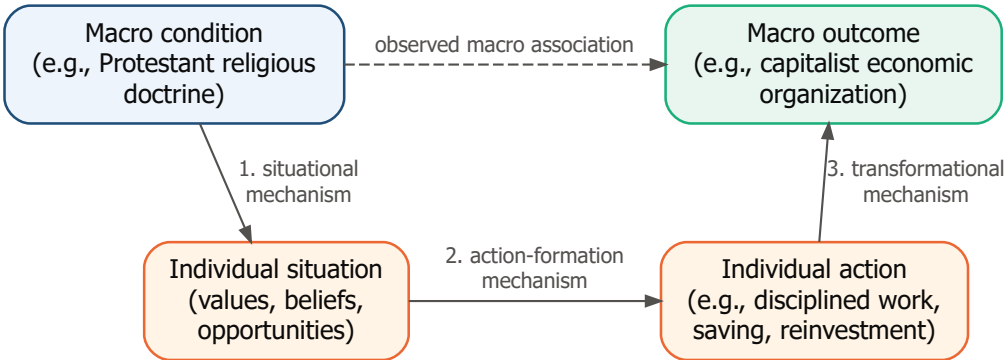


Figure 2.2. Coleman's boat: a macro association (dashed) is explained by a situational mechanism (macro to micro), an action-formation mechanism (micro to micro), and a transformational mechanism (micro to macro). Labels follow Weber's Protestant-ethic example as reconstructed by Coleman (1990) and the typology of Hedstrom and Swedberg (1998).

BUSINESS CASE

Illustrative scenario: forced ranking and collaboration. A firm adopts a forced-distribution appraisal system (a macro, organization-level policy). The macro association the board observes a year later is lower cross-unit collaboration. Coleman's boat structures the explanation. *Situational mechanism:* the policy turns colleagues into competitors for a fixed number of top ratings, changing how each employee perceives the payoff to helping others. *Action-formation mechanism:* employees who believe that helping a peer may cost them a rating rationally help less and share knowledge more selectively. *Transformational mechanism:* many such individual choices aggregate into thin cross-unit networks and slower knowledge flows. Each arrow suggests a different intervention: redesign the policy, change the individual payoff (e.g., reward helping explicitly), or build structures that make collaboration less dependent on discretionary effort.

2.3.5 Boundary conditions

No theory holds everywhere. **Boundary conditions** specify the actors, settings, and times for which a theory's relationships are expected to hold: Whetten's *who, where,* and *when* (Whetten 1989). In empirical work, boundary conditions are often represented as **moderators**: variables that change the strength or direction of a relationship. Stating boundary conditions is not an admission of weakness. A theory that claims to hold everywhere is either trivially true or unfalsifiable. A theory that states where it stops is both more credible and more useful, because it tells a manager when *not* to rely on it.

Figure 2.3 shows the logic for the signaling example. Warranty generosity should matter a great deal when buyers know little about a seller (a new entrant) and much less when buyers already

have abundant information (an established brand), because the warranty's signal value depends on the information gap it closes.

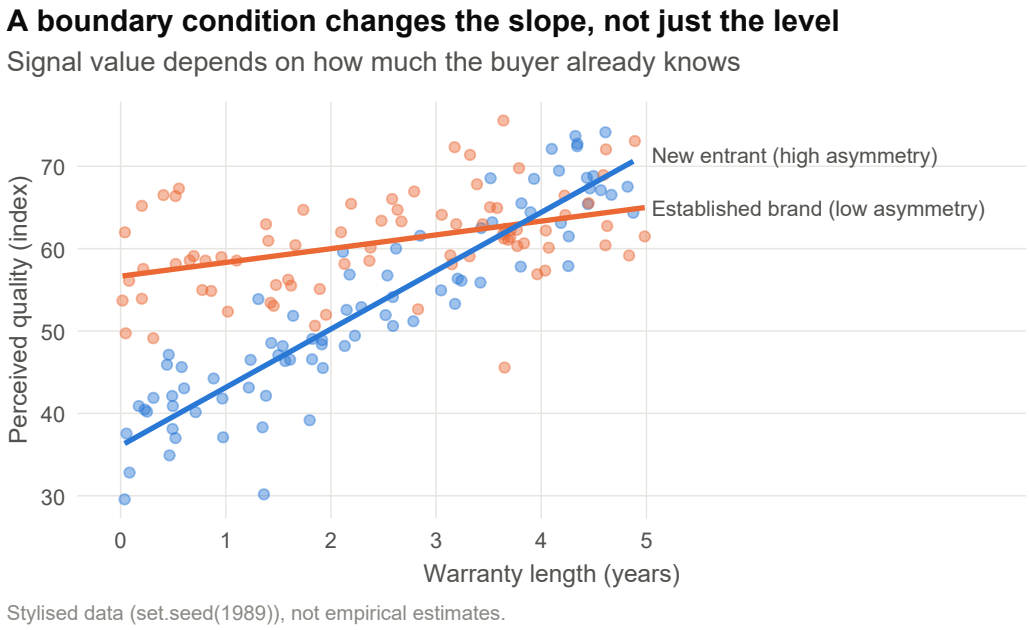


Figure 2.3. Illustrative boundary condition. The effect of warranty length on perceived quality is steep when information asymmetry is high (new entrant) and nearly flat when it is low (established brand). Stylised simulated data, lines are least-squares fits.

2.3.6 Variance and process theories

Most of the theories discussed so far take the form of **variance theories**: they explain variation in an outcome by variation in its antecedents ("the more X, the more Y, under condition Z"). Mohr (1982) contrasted variance theories with **process theories**, which explain an outcome as the result of a sequence of events unfolding over time, where the order and timing of events matter and the necessary conditions need not be sufficient. A variance theory of mergers might relate acquirer experience to post-merger performance. A process theory would describe how integration teams, early decisions about systems and leadership, and emerging conflicts sequence into success or failure.

The two forms answer different questions and call for different evidence. Variance theories are naturally tested with cross-sectional or panel regressions and experiments. Process theories are built and tested with longitudinal case studies, event-sequence data, and narrative analysis. Langley (1999) reviews strategies for theorizing from process data, from narrative and visual mapping to temporal bracketing and quantification of events. For managers, the distinction is practical: a variance theory says *which* levers matter. A process theory says *when* and *in what order* to pull them. Many business failures, a sound strategy launched before the organization could absorb it, a pricing change introduced before customers understood the value proposition, are failures of sequencing that a variance theory cannot see.

COMMON PITFALL

Boxes and arrows without a why. The most common weakness in theory sections is a model diagram whose arrows are justified only by citations ("prior research has found that X affects Y"). A citation shows that a relationship has been observed. It does not explain it. For every arrow, write one or two sentences stating the mechanism and one sentence stating when the mechanism should weaken. If you cannot, the arrow is an empirical conjecture, not theory.

TRY IT YOURSELF

Pick a management idea you have encountered recently: "remote work lowers innovation", "subscriptions increase customer lifetime value (Section 28.4)", "diverse boards improve decisions". Write it out using Whetten's four elements: list the constructs (*what*), state the relationship (*how*), specify the mechanism (*why*), and name at least two boundary conditions (*who, where, when*). Then sketch it as a Coleman boat: which arrows are situational, action-formation, and transformational?

2.4 Types of Theory

Theories differ not only in content but in ambition and purpose. Two distinctions are especially useful: *scope* (grand versus middle-range) and *goal* (explanation versus prediction, among others).

2.4.1 Grand theory and theories of the middle range

Merton (1968a) argued that sociology had been caught between two unproductive poles: the minor working hypotheses of everyday research, which accumulate without integration, and all-inclusive systematic theories that attempt to explain all social behaviour and are too abstract to be tested. Between them he advocated **theories of the middle range**: theories limited to a delimited class of phenomena, abstract enough to generalize across settings but concrete enough to yield testable propositions. His own examples included reference group theory and the theory of role-sets.

Most of the theories in this book are middle-range in Merton's sense. Signaling theory does not try to explain all of economic life, only how informed parties reveal hidden quality. Optimal distinctiveness theory (Section 22.3) does not explain all of identity, only the tension between assimilation and differentiation (Brewer 1991). The middle range is where theory is most easily tested, accumulated, and applied, and where most business research operates.

2.4.2 Gregor's taxonomy of theory types

Writing for information systems, Gregor (2006) proposed a taxonomy that applies well beyond that field. It distinguishes theories by their *primary goal* (Table 2.5).

Table 2.5. Five types of theory distinguished by their primary goal, following Gregor (2006).

Type	Primary goal	What it offers	Business illustration
I. Analysis	Describe and classify	Taxonomies, typologies, frameworks	Porter's generic strategies, Hofstede's cultural dimensions
II. Explanation	Explain how and why, without precise prediction	Causal accounts and mechanisms	Case-based accounts of why incumbents miss disruptive technologies

Table 2.5. Five types of theory distinguished by their primary goal, following Gregor (2006). *(continued)*

Type	Primary goal	What it offers	Business illustration
III. Prediction	Forecast outcomes without necessarily explaining them	Accurate forecasts, black-box models	Machine-learning churn or demand forecasts
IV. Explanation and prediction	Explain and predict	Testable causal propositions	Signaling theory, prospect theory, goal-setting theory
V. Design and action	Prescribe how to do something	Methods, design principles, interventions	A/B testing protocols, incentive-design principles

The distinction between types II and III deserves emphasis. A model can predict well without explaining anything: a demand-forecasting algorithm may exploit weather, search trends, and calendar effects with no account of why consumers behave as they do. Conversely, a theory can explain well and predict poorly, as in evolutionary accounts of industries. Shmueli (2010) shows that explanatory and predictive modelling call for different choices about variables, model complexity, and evaluation. Business analytics increasingly mixes the two, which makes it important to be explicit about which one a given analysis is attempting.

Figure 2.4 crosses scope with goal to place some familiar theories. Placement is a matter of judgement, not measurement. The point is that "theory" spans a wide range of ambitions.

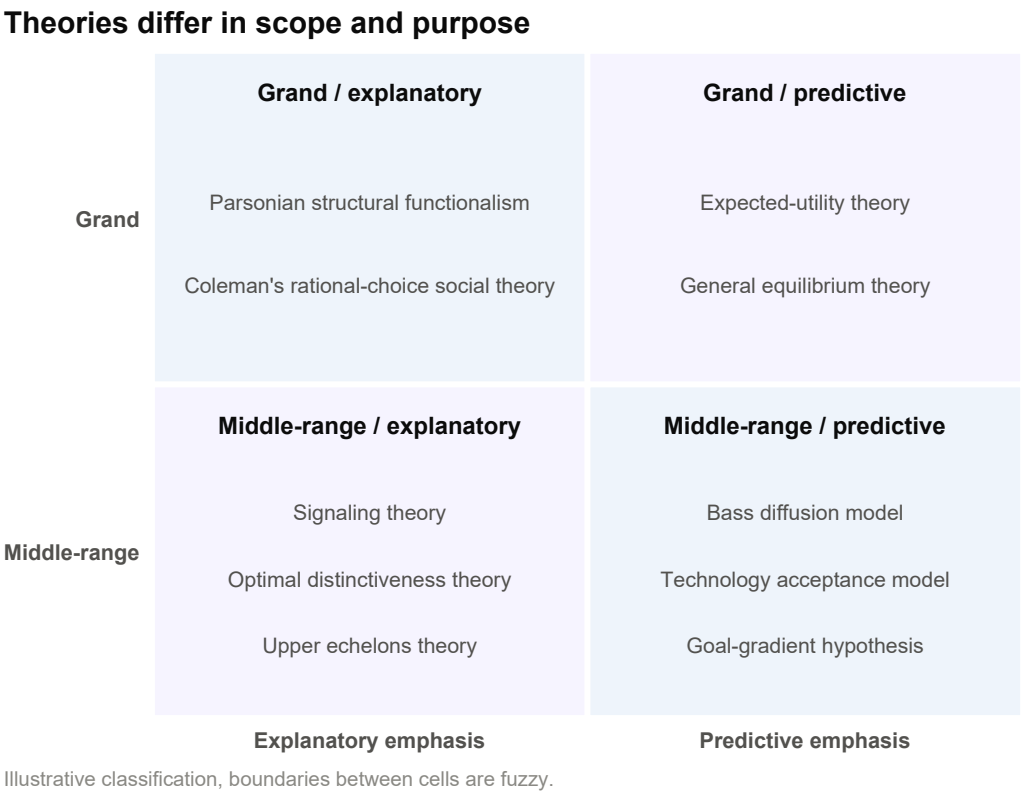


Figure 2.4. A stylised map of theories by scope (grand vs. middle-range. Merton, 1968) and primary goal (explanatory vs. predictive emphasis. Gregor, 2006). Placements are the author's illustrative judgements, not measurements.

BUSINESS CASE

Prediction without explanation in customer analytics. A subscription business can build a churn model that identifies at-risk customers with impressive accuracy using hundreds of behavioural features. The model is a type III theory: it predicts but does not explain. Problems arise when managers treat its predictors as levers. If "number of support tickets" predicts churn, it does not follow that suppressing tickets reduces churn: tickets may simply reveal dissatisfaction that has other causes. An explanatory theory (for example, that churn results from unmet expectations formed at sign-up) suggests different interventions and different tests. Mature analytics teams therefore pair predictive models, which target *whom* to act on, with explanatory theory and experiments, which determine *what* action works.

2.5 What Makes a Theory Good?

If theories differ in type and scope, by what standards should they be judged? Philosophers of science and management scholars have converged on a set of criteria that pull in somewhat different directions.

2.5.1 Falsifiability

Popper (1959) proposed that the line between science and non-science is **falsifiability**: a scientific theory must prohibit some observable states of affairs, so that observing them would count against it. A theory compatible with every possible outcome explains nothing. Bacharach (1989) applies the criterion to organizational theory: constructs must be defined clearly enough, and relationships stated precisely enough, that empirical refutation is possible. Vague constructs, tautological propositions ("successful firms have the capabilities needed for success"), and post hoc boundary conditions all undermine falsifiability.

2.5.2 Utility: explanation and prediction

Bacharach's second criterion is **utility**, which he splits into *explanatory potential* (does the theory explain the relationships it addresses, through clearly specified constructs and mechanisms?) and *predictive adequacy* (does it anticipate outcomes beyond the observations from which it was built?) (Bacharach 1989). A theory that merely redescribes its own data has low utility even if it is technically falsifiable.

2.5.3 Parsimony, generality, and accuracy

A good theory explains much with little. **Parsimony** favours theories with fewer constructs and assumptions, other things equal. **Generality** (or scope) favours theories that apply across many settings. **Accuracy** favours theories whose predictions match observations closely. Weick (1979) popularized Thorngate's argument that a theory of social behaviour cannot simultaneously maximize generality, accuracy, and simplicity: gaining on any two typically costs the third. The trade-off explains why the theory landscape contains both broad, abstract frameworks and narrow, precise models, and why neither is simply "better".

2.5.4 Novelty and interestingness

Theories must also be *worth* attention. In a classic essay, M. S. Davis (1971) argued that theories are judged interesting when they deny some assumption their audience holds, when what seems to be the case turns out not to be, and boring when they affirm what the audience already believes. Prospect theory (Chapter 24) was interesting because it contradicted the assumption that decision-makers evaluate final wealth states consistently (Kahneman and Tversky 1979). The finding that open offices reduce face-to-face interaction is interesting because it contradicts the design intuition behind them (Bernstein and Turban 2018). Interestingness is audience-relative: a result that surprises economists may be obvious to psychologists.

2.5.5 Usefulness

Corley and Gioia (2011) review how management scholars define a theoretical contribution and identify two dimensions: **originality**, ranging from incremental to revelatory, and **utility**, ranging from scientific usefulness (advancing scholarly conversation) to practical usefulness (informing what practitioners do). They argue that the field has overweighted scientific utility and call for more **prescience**: theorizing that anticipates problems managers and societies will face, rather than only explaining the past.

BUSINESS CASE

Illustrative scenario: an unfalsifiable strategy claim. A consultancy publishes a study concluding that "customer-obsessed" companies outperform their peers. On inspection, firms were classified as customer-obsessed partly on the basis of strong growth and high customer ratings: the very outcomes the theory claims to explain. The claim cannot fail: any high performer will be coded as customer-obsessed after the fact. A falsifiable version would define the construct independently of the outcome (for instance, by the share of executive time spent with customers, or by whether product decisions require customer evidence), measure it *before* performance is observed, and specify what pattern , such as customer-obsessed firms underperforming in markets where customer preferences shift rapidly, would count against it. The scenario illustrates Bacharach's point that falsifiability depends as much on construct definition as on the hypothesis itself (Bacharach 1989).

TRY IT YOURSELF

Choose a theory from later in this book and evaluate it against Table 2.6. On which criterion does it score highest? Which criterion does it sacrifice, and is that sacrifice consistent with Thorngate's trade-off between generality, accuracy, and simplicity?

Table 2.6. Criteria for evaluating a theory.

Criterion	Diagnostic question	Typical failure	Key source
Falsifiability	What observation would show the theory wrong?	Tautology, constructs too vague to measure	Popper (1959), Bacharach (1989)
Explanatory potential	Does it say why, through a clear mechanism?	Black-box correlations	Bacharach (1989), Hedstrom & Swedberg (1998)
Predictive adequacy	Does it anticipate outcomes beyond its founding data?	Post hoc fitting, overfitting	Bacharach (1989), Shmueli (2010)

Table 2.6. Criteria for evaluating a theory. *(continued)*

Criterion	Diagnostic question	Typical failure	Key source
Parsimony	Could fewer constructs explain as much?	Kitchen-sink models	Weick (1979)
Generality / scope	How many settings does it cover?	Context-bound results presented as general	Merton (1968), Whetten (1989)
Internal consistency	Do the assumptions and propositions cohere?	Mutually inconsistent assumptions	Bacharach (1989)
Interestingness	Does it challenge an assumption the audience holds?	Confirming the obvious	Davis (1971)
Usefulness	Does it inform scholars, practitioners, or both?	Theory that changes no one's action	Corley & Gioia (2011), Christensen & Raynor (2003)

DEBATE AND EVIDENCE

Can a field be too devoted to theory? Hambrick (2007a) argued that management's insistence that every published paper make a novel theoretical contribution has costs. It discourages the publication of important facts that lack a ready theoretical explanation, encourages authors to dress up findings in elaborate but weakly connected theory, and slows the accumulation of replicated empirical regularities from which better theory could later be built. Others respond that without theory, facts do not cumulate and cannot be transported to new contexts. The debate is a useful reminder that the criteria in Table 2.6 are ideals in tension, not a checklist to be maximized simultaneously, a theme that returns in the discussion of replication in Chapter 3.

2.6 Levels of Analysis

Every theory is a theory *of something*: individuals, dyads, teams, organizations, networks, industries, or societies. The **level of analysis** is the unit to which a theory's constructs and generalizations apply. Business phenomena are almost always nested, employees within teams, teams within firms, firms within industries and countries, which makes level mistakes both common and costly.

2.6.1 Three levels that must align

K. J. Klein, Dansereau, and Hall (1994) distinguish the **level of theory** (the unit the theory seeks to explain), the **level of measurement** (the unit to which data are directly attached), and the **level of statistical analysis** (the unit treated as the observation in the analysis). Problems arise when these are misaligned: for example, when a theory about *team* climate is tested with individual survey responses analysed as independent observations. Klein and colleagues also show that a level-specific theory carries an implicit assumption about how members of a higher-level unit relate to it: members may be **homogeneous** (alike within the unit, so the unit is the relevant level), **independent** of the unit (so the individual is the relevant level), or **heterogeneous** within the unit (so position relative to the group matters, as in a frog-pond model).

2.6.2 Emergence and cross-level models

Rousseau (1985) distinguished **composition models**, which specify how a construct at one level relates to a functionally similar construct at another (e.g., individual perceptions of climate and organizational climate), from **cross-level models**, which relate constructs at one level to outcomes at another, and **multilevel models**, which claim that a relationship holds at several levels. Kozlowski and Klein (2000) add the concept of **emergence**: higher-level properties that arise from the interactions of lower-level elements. Emergence can be *compositional*, when the higher-level construct is essentially the shared version of the lower-level one (team efficacy as shared individual efficacy), or *compilational*, when the higher-level construct is a configuration different in kind from its parts (a team's coordination pattern, which no individual possesses).

2.6.3 Cross-level fallacies

Two inferential errors recur. The **ecological fallacy** infers individual-level relationships from aggregate data. W. S. Robinson (1950) provided the classic demonstration: across US states, the proportion of foreign-born residents was *negatively* correlated with illiteracy, while among individuals foreign birth was *positively* associated with illiteracy, immigrants tended to settle in states whose native-born residents were more literate. The opposite error, sometimes called the **atomistic** or **individualistic fallacy**, infers group-level relationships from individual data. Figure 2.5 shows how the two levels can point in opposite directions.

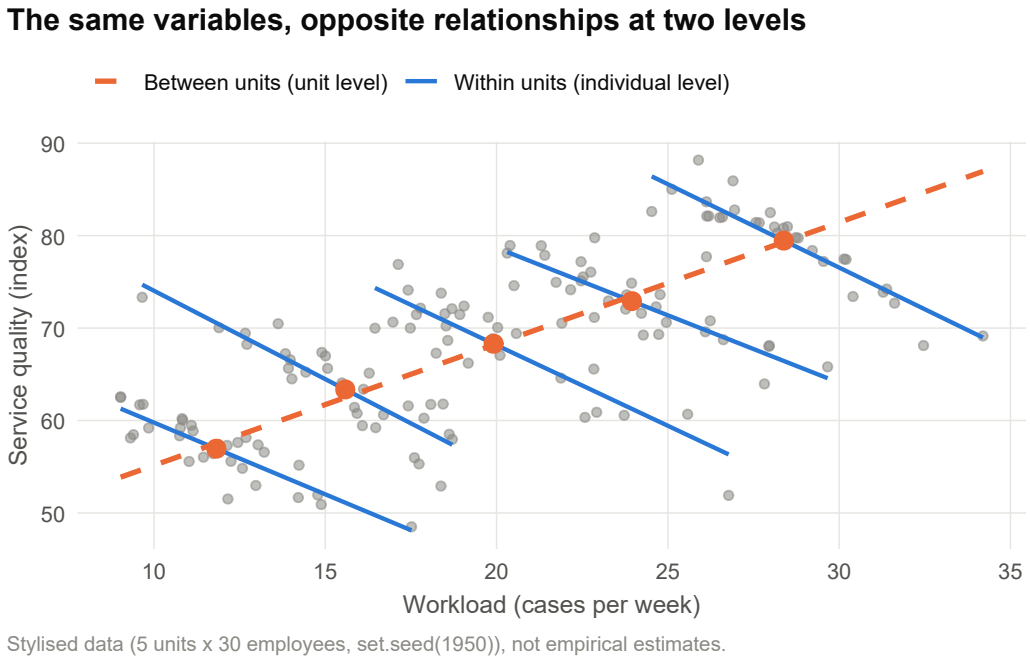


Figure 2.5. Illustrative cross-level reversal. Across business units, units with higher average workload show higher average service quality (orange), perhaps because high-performing units attract more demand. Within each unit, employees carrying more cases than their colleagues deliver lower quality (blue). Stylised simulated data.

BUSINESS CASE

Engagement surveys and the level of inference. Harter, Schmidt, and Hayes (2002) meta-analysed the relationship between employee satisfaction and engagement and business outcomes , customer satisfaction, productivity, profit, turnover, and safety, at the level of **business units**, finding generalizable positive relationships. The level of the evidence matters for how managers use it. The findings support unit-level claims (units with more engaged workforces tend to perform better) and unit-level interventions (improving local management practices). They do not, by themselves, establish that any given employee would perform better if made more engaged, nor do they settle the direction of causality, since well-performing units may also make work more engaging. Reading the evidence at the right level is the difference between a sound engagement programme and an overreaching one.

RESEARCH TOOLKIT

Checklist for levels in a research design.

- *State the focal level* of each construct in the theory section, and justify it.
- *Specify the composition model* when aggregating (e.g., direct consensus, referent shift, dispersion), and report within-group agreement before aggregating.
- *Align* the levels of theory, measurement, and analysis, or explain and model the misalignment.
- *Model nesting* explicitly (multilevel or hierarchical models, clustered standard errors, fixed effects for higher-level units).
- *Guard against both fallacies*: do not draw individual conclusions from unit data or unit conclusions from individual data without a cross-level argument.

2.6.4 A map of the book by level

The theories in this book can be organized by the level at which they primarily operate. Table 2.7 serves as a navigational map. Many theories span levels, signaling theory (Section 8.2) has been applied to individuals, firms, and nations, so the "dominant level" column indicates centre of gravity, not exclusive domain.

Table 2.7. A map of the book: theory families by part, chapter, and dominant level of analysis.

Part	Chapter	Theory families	Dominant level(s)
Foundations	1-3. Introduction, What is a theory?, Using theory	Meta-theory, theory building and testing	All levels
Organizations & Strategy	4. Resources	Resource-based theory, VRIO, knowledge-based view, resource-advantage theory	Firm
Organizations & Strategy	5. Dynamic capabilities	Dynamic capabilities, absorptive capacity, exploration-exploitation, ambidexterity	Firm, business unit
Organizations & Strategy	6. Governance	Agency, stewardship, transaction cost economics, upper echelons	Dyad, top management team, firm
Organizations & Strategy	7. Environment	Resource dependence, institutional, legitimacy, stakeholder, ecology, contingency, CSR	Organization, field, population

Table 2.7. A map of the book: theory families by part, chapter, and dominant level of analysis. *(continued)*

Part	Chapter	Theory families	Dominant level(s)
Organizations & Strategy	8. Information	Information asymmetry, signaling, screening, costly signaling	Dyad, market
Organizations & Strategy	9. Networks & exchange	Social exchange, network exchange, weak ties, structural holes, social capital	Dyad, network
Organizations & Strategy	10. Innovation & adoption	Diffusion of innovations, Bass model, TAM, UTAUT, IS success	Individual adopter, market
Organizations & Strategy	11. Strategy & competition	Structure-conduct-performance, competitive dynamics, game theory, real options, behavioral theory of the firm	Firm, industry
Organizations & Strategy	12. Innovation & entrepreneurship	Creative destruction, disruptive innovation, profiting from innovation, effectuation	Firm, venture, industry
Organizations & Strategy	13. International business	Uppsala model, OLI paradigm, internalization, institutional voids	Multinational firm, country
Organizations & Strategy	14. Organizations & decisions	Garbage can, escalation, organizational learning, culture, sensemaking	Organization, team
Organizations & Strategy	15. Work & leadership	Job design, JD-R, LMX, transformational leadership, justice, tournaments	Individual, dyad, team
Markets, Finance & Economics	16. Finance	Portfolio theory, CAPM, efficient markets, capital structure, behavioral finance	Investor, firm, market
Markets, Finance & Economics	17. Economics & institutions	Coase theorem, collective action, network effects, auctions, social preferences	Market, institution
Individuals & Groups	18. Culture	Hofstede, power distance, GLOBE, Schwartz values, tight-loose	Society, individual
Individuals & Groups	19. Motivation	Expectancy, equity, goal setting, goal gradient, self-determination	Individual
Individuals & Groups	20. Self-regulation	Regulatory focus and fit, construal level, self-control	Individual
Individuals & Groups	21. Attitudes & attribution	Cognitive dissonance, attribution, self-perception, social cognitive theory, reactance	Individual
Individuals & Groups	22. Social influence	Social comparison, social identity, optimal distinctiveness, conformity, norms	Individual within groups
Individuals & Groups	23. Judgment & decision	Bounded rationality, heuristics and biases, dual process, choice architecture	Individual
Individuals & Groups	24. Prospect theory	Expected utility, prospect theory, loss aversion, reference dependence	Individual, market
Individuals & Groups	25. Emotion & well-being	Broaden-and-build, hedonic adaptation, appraisal theory	Individual
Individuals & Groups	26. Persuasion & communication	Elaboration likelihood, persuasion knowledge, agenda setting	Individual, audience
Consumers & Marketing	27. Consumer behavior	Satisfaction, brand equity, brand relationships, self-congruity, reference price	Consumer
Consumers & Marketing	28. Marketing strategy	Market orientation, service-dominant logic, service quality, customer equity	Firm, customer base
Effects, Phenomena & Frameworks	29. Effects	Mere exposure, decoy, anchoring, IKEA, snob, glass cliff	Individual (mostly)

Table 2.7. A map of the book: theory families by part, chapter, and dominant level of analysis. *(continued)*

Part	Chapter	Theory families	Dominant level(s)
Effects, Phenomena & Frameworks	30. Phenomena	Similarity asymmetry, wisdom of the crowd, groupthink, bullwhip effect	Individual, dyad, collective
Effects, Phenomena & Frameworks	31. Frameworks	Information processing, five forces, creativity models, mechanism toolkit	Individual to industry

2.7 How Theories Change

Theories are not discovered once and then preserved. They are proposed, tested, amended, combined, and sometimes abandoned. Understanding how this happens helps readers place any single theory in its intellectual trajectory.

THEORY CARD

Core claim. Science alternates between long periods of **normal science**, conducted within a shared paradigm, and occasional **revolutions** in which an accumulation of anomalies leads a community to adopt a new paradigm.
Origins. History and philosophy of science, Kuhn (1962).
Level of analysis. Scientific community and discipline.
Key constructs. Paradigm, normal science, puzzle-solving, anomaly, crisis, revolution, incommensurability.
Boundary conditions. Developed from the history of the natural sciences. Its fit to the social sciences, where several paradigms often coexist, is debated.
Related theories. Lakatos's methodology of research programmes (Lakatos 1970). Popper's falsificationism (Popper 1959).

Kuhn (1962) argued that mature sciences are organized around **paradigms**: shared exemplars, assumptions, and methods that define which problems are legitimate and what counts as a solution. Within a paradigm, scientists solve puzzles rather than test foundations. Anomalies, observations the paradigm cannot accommodate, are initially set aside or explained away. When they accumulate, a crisis may follow, and a new paradigm can replace the old one. Kuhn controversially argued that competing paradigms can be partly **incommensurable**, so that the choice between them is not settled by data alone. Figure 2.6 summarizes the cycle.

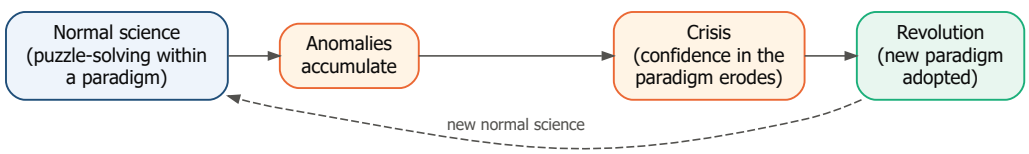


Figure 2.6. The Kuhnian cycle of theory change (after Kuhn, 1962). Normal science proceeds within a paradigm until anomalies accumulate. A crisis can lead to a revolution and a new period of normal science. Lakatos (1970) offers a gradualist alternative in which research programmes progress or degenerate.

Lakatos (1970) proposed a less dramatic account. Scientific **research programmes** consist of a

hard core of assumptions protected by a *belt* of auxiliary hypotheses that absorb anomalies. A programme is *progressive* when its modifications predict novel facts and *degenerating* when they merely patch failures after the fact. The Lakatosian lens fits management and economics well, where theories are rarely overthrown outright but are amended, bounded, and gradually displaced.

The history of strategy research illustrates the pattern. Industrial-organization reasoning, popularized for strategy by Porter, located the sources of profitability primarily in industry structure (Michael E. Porter 1980a). Persistent performance differences among firms *within* the same industry were an anomaly for that view, and the resource-based view relocated the explanation inside the firm (Wernerfelt 1984, Jay Barney 1991). When critics noted that static resource endowments could not explain how firms adapt to rapid change, the dynamic-capabilities framework extended the programme (Teece, Pisano, and Shuen 1997). Each shift retained much of its predecessor. None was a clean Kuhnian revolution. In behavioural decision research, by contrast, the accumulation of anomalies for expected-utility theory produced something closer to a paradigm contest, with prospect theory offering a descriptive alternative built on reference dependence and loss aversion (Section 24.4) (Kahneman and Tversky 1979).

DEBATE AND EVIDENCE

Does management have, or need, a paradigm? Pfeffer (1993) argued that organizational science suffers from low **paradigm development**: weak consensus on theories and methods, compared with fields such as economics. In his view, this fragmentation makes it harder to accumulate knowledge, allocate resources, and defend the field's standing. Critics countered that theoretical pluralism is appropriate to a complex subject matter and that enforced consensus could suppress valuable alternatives. The debate maps onto a practical question for readers of this book: when theories disagree, should we seek a winner, a synthesis, or a clear statement of the conditions under which each applies? Chapter 3 takes up that question under the heading of competing-theory tests and strong inference.

COMMON PITFALL

Mistaking a fashion for a revolution. Management ideas often rise and fall with the business press, and new labels are sometimes presented as paradigm shifts. Before treating a new framework as a replacement for an established theory, ask Lakatos's question: does it predict novel facts that the older theory cannot, or does it re-describe familiar facts in new vocabulary?

2.8 Chapter Summary

KEY TAKEAWAYS

- A **theory** is a set of related propositions about constructs, together with an explanation of *why* they are related and *where* the relationships hold (Whetten 1989, Bacharach 1989).
- References, data, variable lists, diagrams, and hypotheses are ingredients of theory, not substitutes for it. The missing element is usually the *why* (R. I. Sutton and Staw 1995), although theorizing is a process whose interim products matter (Weick 1995b).
- **Constructs** are linked by **propositions**, **variables** by **hypotheses**. Every empirical test

is also a test of measurement.

- **Mechanisms** open the black box. Coleman's boat decomposes macro associations into situational, action-formation, and transformational mechanisms (Coleman 1990, Hedström and Swedberg 1998).
- Theories vary in **scope** (grand vs. middle-range) and **goal** (analysis, explanation, prediction, design) (Merton 1968a, Gregor 2006).
- Good theories are falsifiable, useful, parsimonious, appropriately general, and interesting, but these criteria trade off against one another (Popper 1959, M. S. Davis 1971, Corley and Gioia 2011).
- Levels of theory, measurement, and analysis must align. Relationships can reverse across levels (K. J. Klein, Dansereau, and Hall 1994, W. S. Robinson 1950).
- Theories change through revolutions and, more often in management, through the progressive or degenerating evolution of research programmes (Kuhn 1962, Lakatos 1970).
- Managers always act on theories. The choice is between explicit, tested theories and implicit, untested ones (Christensen and Raynor 2003, Rousseau 2006).

Table 2.8 presents the summary of the chapter's core ideas.

Table 2.8. Summary of the chapter's core ideas.

Idea	Core claim	Practical test	Key source
What theory is not	Ingredients are not explanations	Can you state the why for each arrow?	Sutton & Staw (1995)
Four building blocks	What, how, why, who/where/when	Are all four explicit?	Whetten (1989)
Constructs vs. variables	Propositions link constructs, hypotheses link variables	Is the construct-to-measure link defensible?	Bacharach (1989)
Mechanisms	Explain macro patterns through micro processes	Can the intermediate steps be observed?	Coleman (1990), Hedstrom & Swedberg (1998)
Middle-range theory	Bounded, testable theories accumulate best	Is the scope stated and testable?	Merton (1968)
Theory types	Analysis, explanation, prediction, design	Is the goal explicit?	Gregor (2006)
Evaluation criteria	Falsifiable, useful, parsimonious, interesting	Which criterion is being traded off?	Popper (1959), Davis (1971)
Levels of analysis	Align theory, measurement, analysis	Is a cross-level inference being made?	Klein et al. (1994)
Theory change	Paradigms and research programmes	Does the new idea predict novel facts?	Kuhn (1962), Lakatos (1970)

2.9 Discussion Questions

1. Take a recent empirical article in your field. Using Whetten's four elements, identify which element is weakest. What would it take to strengthen it?
2. Sutton and Staw argue that diagrams are not theory, while Weick argues that they may be valuable steps in theorizing. Under what circumstances is a diagram-heavy paper a genuine contribution?
3. Choose a managerial "best practice" (e.g., open-plan offices, forced ranking, loyalty programmes). Reconstruct the implicit theory behind it as a Coleman boat. Which arrow is least

supported by evidence?

4. A consultancy reports that firms with more diverse boards have higher returns on equity. What level-of-analysis and causal questions would you ask before recommending board diversification to a specific client?
5. Thorngate's postulate suggests that no theory can be simultaneously general, accurate, and simple. Pick one theory from this book and argue which of the three it sacrifices.
6. Is the shift from industry-structure explanations to the resource-based view better described as a Kuhnian revolution or a Lakatosian progression? Defend your answer with evidence.

2.10 Further Reading

- Whetten (1989): a five-page statement of what a theoretical contribution requires. The most practical starting point.
- R. I. Sutton and Staw (1995): a sharp diagnosis of what theory is not, with the replies by Weick (1995b) and DiMaggio (1995) in the same issue.
- Bacharach (1989): rigorous criteria for evaluating organizational theories, including the construct-variable distinction.
- Hedström and Swedberg (1998): the core statement of mechanism-based explanation in the social sciences.
- K. J. Klein, Dansereau, and Hall (1994): the standard reference on levels issues in theory, measurement, and analysis.
- M. S. Davis (1971): why some theories capture attention and others do not. Still provocative.
- Christensen and Raynor (2003): a practitioner-oriented argument for why managers need explicit causal theory.
- Kuhn (1962): the classic account of paradigms and scientific revolutions.

Chapter 3

Using Theory in Research and Practice

A subscription software company discovers that customers who use its collaboration feature in their first month are far less likely to cancel. The growth team proposes a simple plan: push every new customer into the feature. A data scientist asks a different question: *why* would the feature reduce churn? If it creates switching costs (shared files are painful to move), forcing adoption should help. If it merely reveals customers who were already committed (teams that collaborate were always going to stay), forcing adoption will do little. If it works by connecting users to colleagues who keep them engaged, it should help only in accounts with several active users. Three theories, three predictions, three different experiments. The company's choice of what to test, whom to test it on, and what to measure is a theoretical choice, whether or not anyone calls it one.

Chapter 2 asked what theory *is*. This chapter asks what researchers and managers *do* with it. We trace the path from a theory to a testable hypothesis, explain how constructs earn their meaning in a nomological network, set out the logic of mediation and moderation, discuss how to choose among competing theories, distinguish types of theoretical contribution and modes of reasoning, show how causal diagrams make a theory's assumptions explicit, and confront the replication crisis and what it implies for credible research. We close with theory in managerial practice, from A/B tests to evidence-based management (Table 3.1).

Table 3.1. Chapter at a glance: putting theory to work.

Task	Core question	Key tools	Key sources
From theory to hypotheses	How does an abstract argument become a falsifiable prediction?	Hypothesis-development paragraph, propositions vs. hypotheses	Bacharach (1989), Whetten (1989)
Construct validity	Does a measure capture the construct the theory names?	Nomological network, convergent and discriminant validity	Cronbach & Meehl (1955), Campbell & Fiske (1959)
Mediation and moderation	Through what process, and under what conditions?	Indirect effects, interactions, conditional process models	Baron & Kenny (1986), Zhao et al. (2010), Hayes (2018)
Choosing among theories	Which explanation survives a head-to-head test?	Strong inference, competing predictions	Platt (1964)
Theoretical contribution	Does the work build theory, test it, or both?	Contribution taxonomy, deduction, induction, abduction	Colquitt & Zapata-Phelan (2007)
Causal identification	Which variables must be controlled, and which must not?	Directed acyclic graphs, back-door criterion	Pearl (2009)

Table 3.1. Chapter at a glance: putting theory to work. *(continued)*

Task	Core question	Key tools	Key sources
Research integrity	Why do findings fail to replicate, and what helps?	Preregistration, power, transparency	Open Science Collaboration (2015), Simmons et al. (2011)

3.1 From Theory to Hypotheses

Theory and evidence meet at the hypothesis. As Chapter 2 explained, a **proposition** relates constructs and a **hypothesis** relates measured variables (Bacharach 1989). Moving from one to the other is a chain of translations, from the general theory to a proposition about the focal context, from constructs to variables, from variables to a research design, and each link can fail. Figure 3.1 shows the chain and the feedback loop by which evidence returns to revise the theory.

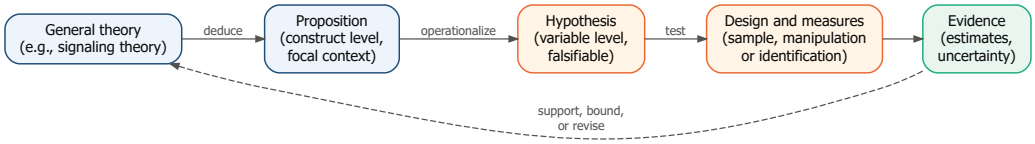


Figure 3.1. The ladder from theory to evidence. Each step translates the previous one into a more concrete claim. Evidence feeds back to the theory, strengthening, bounding, or revising it.

DEFINITION

Definition 3.1. Hypothesis. A conjectural statement of the relationship between two or more measured variables, derived from theory, stated precisely enough that a specified pattern of data would contradict it. A good hypothesis names the direction (and ideally the form) of the relationship, the population to which it applies, and, where the theory implies one, the condition under which it should weaken or reverse.

3.1.1 Anatomy of a hypothesis-development paragraph

In most empirical articles, each hypothesis is preceded by a paragraph (or several) that justifies it. Reviewers read these paragraphs for one thing above all: the *why*. A strong paragraph contains five moves, summarized in Table 3.2.

Table 3.2. Five moves of a hypothesis-development paragraph.

Move	Purpose	Typical phrasing	Common failure
1. Context and problem	Locate the hypothesis in a concrete setting with a real tension	In enterprise software markets, buyers cannot directly observe ...	Starting from the literature rather than the phenomenon
2. Theoretical premise	State the general principle being applied	Signaling theory holds that ...	Name-checking a theory without stating its logic
3. Mechanism	Explain step by step why X should affect Y here	Because ..., buyers infer ..., which reduces ...	Replacing mechanism with citations of prior findings
4. Boundary or alternative	Acknowledge when the mechanism should weaken or what else could explain the link	This logic should be strongest when ...	Presenting the link as universal

Table 3.2. Five moves of a hypothesis-development paragraph. *(continued)*

Move	Purpose	Typical phrasing	Common failure
5. Formal hypothesis	State a falsifiable, directional prediction about variables	H1: Start-ups with an independent security certification will ...	Vague verbs (affects, is related to) with no direction or population

3.1.2 A worked example: signaling in enterprise software

Consider a study of whether young business-to-business software firms win more enterprise contracts when they obtain an independent security audit report (for example, a SOC 2 report). The relevant theory is signaling theory (Section 8.2): an informed party can credibly reveal hidden quality through a costly action that low-quality types find too expensive to imitate (Spence 1973, Connelly et al. 2011).

Before (weak).

> Signaling theory suggests that firms send signals to stakeholders (Connelly et al. 2011). Prior research has found that certifications have positive effects on firm outcomes. Security is important to enterprise customers. Therefore, certification should be related to sales.
>
> *H1: Security certification is positively related to firm performance.*

The paragraph cites a theory but never uses it. It does not say what is hidden, from whom, why the certification is credible, or when its effect should be largest. "Firm performance" is not the construct the argument implies, and the hypothesis cannot fail in an informative way: almost any positive correlation would be read as support, even one produced by larger firms both certifying and selling more.

After (strong).

> Enterprise buyers evaluating a young software vendor face acute information asymmetry about its security practices: weaknesses become visible only after a breach, when the buyer bears much of the cost. Signaling theory implies that vendors can reduce this asymmetry through actions whose cost is lower for high-quality types (Spence 1973, Connelly et al. 2011). An independent security audit is such an action. Passing it requires controls that a vendor with weak practices would have to build at substantial expense, and an auditor's reputation is at stake if it attests falsely. Buyers can therefore infer that certified vendors are more likely to have sound security, lowering perceived risk and shortening procurement reviews. Because the signal's value lies in closing an information gap, its effect should be strongest when buyers have little other information about the vendor, that is, for vendors with short track records, and weaker for established vendors whose reputation already conveys similar information (Bergh et al. 2014).
>
> *H1: Among business-to-business software start-ups, obtaining an independent security audit report will increase the rate at which enterprise sales opportunities convert into signed contracts.*
>
> *H2: The positive effect in H1 will be weaker for vendors with longer operating track records.*

The revised version names the hidden attribute, explains the differential cost that makes the signal credible, specifies the receiver's inference, links the mechanism to a measurable outcome (opportunity-to-contract conversion), states a theoretically derived boundary condition, and produces a hypothesis pair that could fail in a specific way. It also suggests the right test: the theory implies a causal effect, so the design must address the possibility that better-run start-ups both certify and sell more, for example, by exploiting variation in the timing of certification within firms or buyer-side requirements that change exogenously. H2 is a **moderation** hypothesis. The remainder of this chapter shows how such hypotheses are modelled.

3.1.3 Direction, form, and magnitude

Most hypotheses in business research predict only a *direction* ("positively related"). Theories often imply more. Some imply a **functional form**: diminishing returns to advertising, a threshold beyond which a signal stops separating types, or an inverted U in which moderate levels of a variable are optimal. Testing such claims requires more than a significant squared term. Haans, Pieters, and He (2016) show that an inverted-U hypothesis should be supported by evidence that the slope is significantly positive at the low end of the observed range and significantly negative at the high end, and that the turning point lies within that range. Otherwise the data may simply reflect a monotonic relationship with diminishing returns. Theories can also imply **magnitudes**: for instance, that an effect should be larger in one condition than another by a meaningful amount, or that an effect too small to matter economically would count against the theory even if statistically significant.

Specifying form and magnitude makes a hypothesis riskier and therefore more informative. A directional prediction that any positive coefficient confirms is a weak test, because many rival theories predict the same sign. A prediction about shape, threshold, or relative size exposes the theory to genuine refutation, in the spirit of the falsifiability criterion discussed in Chapter 2.

COMMON PITFALL

Theory as decoration. A theory that is cited in the introduction but plays no role in deriving hypotheses, choosing measures, or interpreting results is decoration. A simple diagnostic: delete the theory's name from the paper. If no hypothesis, design choice, or interpretation changes, the theory was not doing any work. The remedy is not more citations but a sharper statement of the mechanism and of the observations that would contradict it.

3.2 Constructs, Measures, and Nomological Networks

A hypothesis is only as good as the measures that test it. If "brand trust" is measured with items that really capture brand familiarity, a test of trust theory is a test of something else. The idea that measures acquire meaning through their place in a theory is one of the most important contributions of psychometrics to social science.

THEORY CARD

Core claim. A construct's meaning is given by the network of lawful relationships, its **nomological network**, that link it to other constructs and to observable measures. A measure is valid to the extent that it behaves as that network predicts.

Origins. Psychometrics. Cronbach and Meehl (1955), extended by the multitrait-multimethod logic of D. T. Campbell and Fiske (1959).
Level of analysis. Constructs and measures at any level (individual, team, firm, market).
Key constructs. Construct validity, nomological network, convergent validity, discriminant validity, content validity, criterion validity.
Boundary conditions. Requires a theory rich enough to generate predictions about a construct's relationships. Weakly theorized constructs cannot be validated this way.
Related theories. Construct clarity (Suddaby 2010), measurement and validation procedures (MacKenzie, Podsakoff, and Podsakoff 2011).

Cronbach and Meehl (1955) argued that when a construct cannot be defined by a single criterion, validation is necessarily theoretical: the researcher predicts how the measure should relate to other measures and tests those predictions. Construct validation and theory testing are therefore intertwined: a failed prediction may indicate a poor measure, a wrong theory, or both. D. T. Campbell and Fiske (1959) operationalized part of this logic in the **multitrait-multimethod matrix**: measures of the same construct obtained by different methods should converge (**convergent validity**), and measures of different constructs should be distinguishable even when obtained by the same method (**discriminant validity**). Contemporary procedures for developing and validating scales, including the choice between reflective and formative measurement models, are reviewed by MacKenzie, Podsakoff, and Podsakoff (2011).

BUSINESS CASE

The Net Promoter Score as a construct-validity problem. Reichheld (2003) proposed that a single question, how likely a customer is to recommend a company to a friend or colleague, captures customer loyalty well enough to predict company growth. The claim is, in effect, a nomological-network prediction: if the measure captures loyalty, it should predict growth at least as well as alternative loyalty and satisfaction measures. Keiningham et al. (2007) examined this prediction with longitudinal data from multiple industries and did not find that the Net Promoter metric outperformed the American Customer Satisfaction Index in predicting revenue growth. The lesson is not that a recommendation question is useless, but that a metric's managerial authority should rest on evidence about how it behaves in its nomological network, not on its simplicity.

3.3 Mediation and Moderation

Mediation and moderation are the two most common ways of putting Whetten's *why* and *who/where/when* into a statistical model (Whetten 1989). A **mediator** answers *how* or *why* X affects Y: it is the mechanism through which the effect travels. A **moderator** answers *when* or *for whom*: it changes the strength or direction of the X-Y relationship. R. M. Baron and Kenny (1986) drew the distinction sharply in a paper that became one of the most cited in the social sciences (Table 3.3).

Table 3.3. Mediation and moderation compared.

Feature	Mediation	Moderation
Theoretical question	How or why does X affect Y?	When, for whom, or under what conditions does X affect Y?
Whetten element	Why (mechanism)	Who / where / when (boundary condition)
Causal structure	X -> M -> Y	W changes the X -> Y relationship
Typical statistical test	Indirect effect a x b with bootstrap confidence interval	Interaction term X x W, simple slopes, Johnson-Neyman region
Signaling example	Certification -> perceived risk -> contract signed	Certification effect weaker for vendors with long track records
Key threat	Confounding of the M -> Y path, even when X is randomized	Low power to detect interactions, arbitrary spotlight values
Key sources	Baron & Kenny (1986), Zhao et al. (2010)	Aiken & West (1991), Spiller et al. (2013)

3.3.1 Moderation: boundary conditions in data

Moderation is tested by including the product of the focal predictor and the moderator in a regression (Aiken and West 1991). Because the coefficient on the product term is rarely meaningful on its own, researchers probe the interaction by estimating **simple effects**: the effect of X at particular values of the moderator. Spiller et al. (2013) distinguish **spotlight** analysis, which tests simple effects at chosen values (such as one standard deviation above and below the mean), from **floodlight** analysis, which reports the full range of moderator values over which the simple effect is statistically significant (the Johnson-Neyman region). Floodlight analysis avoids arbitrary spotlight choices and is especially useful when the moderator has meaningful units, such as years of operation.

Figure 3.2 applies this logic to hypothesis H2 from the worked example, using simulated data.

3.3.2 Mediation: testing the mechanism

R. M. Baron and Kenny (1986) proposed a sequence of conditions for mediation: X must predict Y, X must predict M, M must predict Y controlling for X, and the X-Y relationship must shrink when M is controlled. Later work showed that several of these steps are neither necessary nor sufficient. X. Zhao, Lynch, and Chen (2010) argue that the single requirement for establishing mediation is a significant **indirect effect**, the product of the X $\rightarrow$ M path (a) and the M $\rightarrow$ Y path (b), best tested with a bootstrap confidence interval rather than with the causal-steps procedure or the Sobel test. A significant total effect of X on Y is not required: an indirect effect can exist even when opposing mechanisms cancel out in the total effect. Hayes (2018) provides the standard regression-based treatment, including **conditional process models** that combine mediation and moderation (Table 3.4).

Table 3.4. A typology of mediation outcomes, following Zhao, Lynch and Chen (2010).

Pattern	Indirect effect (a x b)	Direct effect (c')	Interpretation
Complementary mediation	Significant	Significant, same sign as a x b	The hypothesized mediator is one of several mechanisms. Others may be omitted
Competitive mediation	Significant	Significant, opposite sign	An omitted mediator likely works in the opposite direction
Indirect-only mediation	Significant	Not significant	Consistent with the hypothesized mediator, omitted mediators less likely

Table 3.4. A typology of mediation outcomes, following Zhao, Lynch and Chen (2010). *(continued)*

Pattern	Indirect effect (a x b)	Direct effect (c')	Interpretation
Direct-only non-mediation	Not significant	Significant	The hypothesized mediator is not the mechanism. Look for another
No-effect non-mediation	Not significant	Not significant	Neither effect detected, theory or design may be flawed

Figure 3.3 shows how a bootstrap test works in a simulated loyalty-programme study, where enrolment is hypothesized to raise spending by increasing customers' perceived switching costs.

Mediation and moderation are often combined. In a **moderated mediation** model, a moderator changes the strength of one path in the causal chain, so the size of the indirect effect depends on the moderator (Preacher, Rucker, and Hayes 2007, Hayes 2018). Figure 3.4 shows the full theoretical model for the worked example: certification works by reducing buyers' perceived risk, and this first stage of the mechanism is weaker for vendors with long track records.

DEBATE AND EVIDENCE

Can statistical mediation reveal causal mechanisms? Even when X is randomly assigned, the mediator is not, so the M → Y path is vulnerable to confounding by unmeasured variables that affect both M and Y. Bullock, Green, and Ha (2010) argue that standard mediation analyses therefore rest on strong and often implausible assumptions, and that convincing evidence about mechanisms usually requires experiments that manipulate the mediator as well. Spencer, Zanna, and Fong (2005) propose an **experimental-causal-chain** design: one experiment shows that X changes M, and a second manipulates M directly to show that it changes Y. Defenders of regression mediation respond that it remains valuable for probing theoretically specified processes, provided its causal assumptions are stated and sensitivity analyses are reported. The practical consensus is to treat cross-sectional mediation as consistent with a mechanism, not as proof of one.

BUSINESS CASE

Illustrative scenario: why does the loyalty programme work? A retailer finds that its points programme raises spending among enrolled members. Two mechanisms are plausible. Under an *economic* account, points create switching costs that lock customers in. Under a *psychological* account, tier status (silver, gold, platinum) provides recognition that customers value for its own sake. The accounts imply different designs: if switching costs drive the effect, raising the cash value of points should matter. If status drives it, making tiers more visible should matter even with unchanged point values. A sequence of randomized tests, manipulating point value in some markets and tier visibility in others, would identify which mechanism dominates far more convincingly than a survey-based mediation model, and would tell the retailer where to invest.

RESEARCH TOOLKIT

Specifying mediation and moderation hypotheses.

- *State the mechanism in words first*, then map each step to a path (a, b, c').

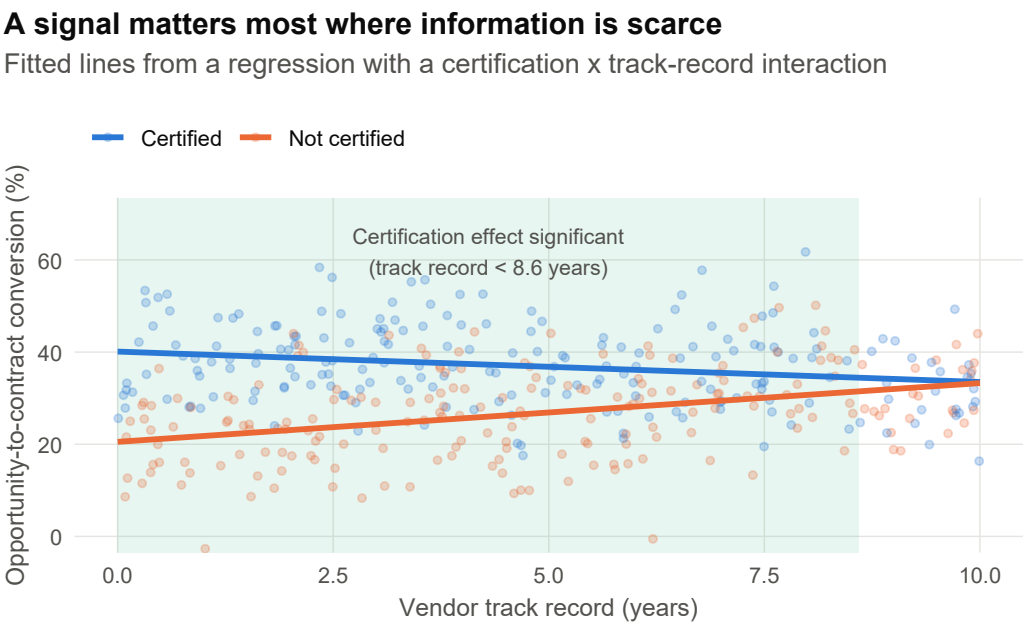


Figure 3.2. Illustrative moderation (floodlight) analysis of the worked example. The effect of a security certification on contract conversion is large for young vendors and shrinks with track record. The shaded region marks track-record values for which the simple effect is significant at the 5 percent level (Johnson-Neyman region). Simulated data.

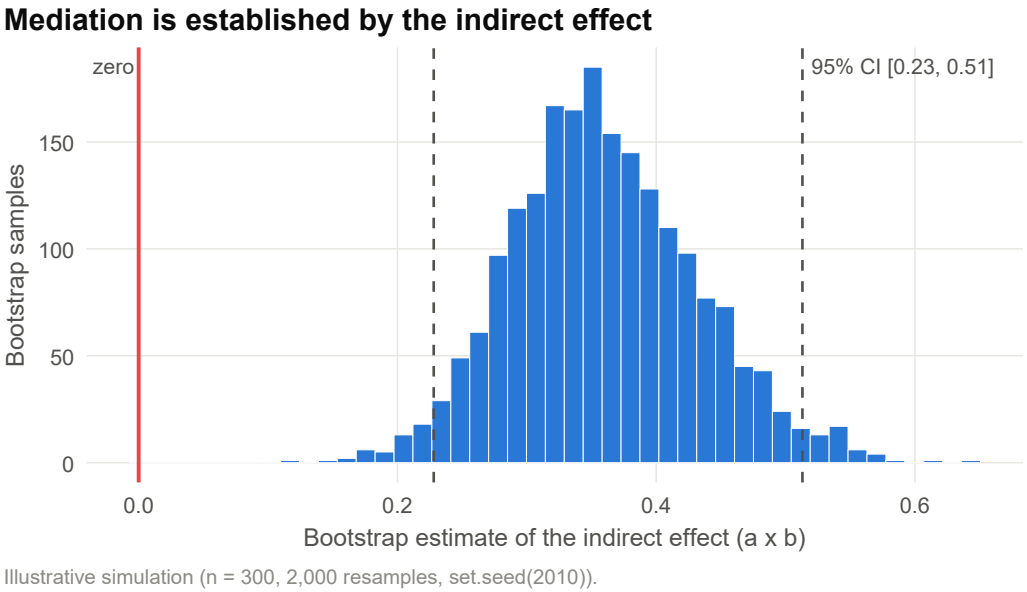


Figure 3.3. Illustrative bootstrap test of an indirect effect. Enrolment in a loyalty programme (randomized) raises perceived switching costs, which raise annual spending. The histogram shows 2,000 bootstrap estimates of the indirect effect a x b. Dashed lines mark the 95 percent percentile interval, which excludes zero. Simulated data.

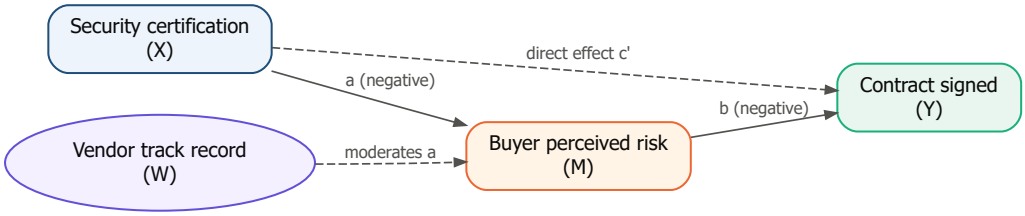


Figure 3.4. A first-stage moderated mediation model for the worked example. Track record moderates the effect of certification on perceived risk, so the indirect effect of certification on contract signing is conditional on track record (Preacher et al., 2007, Hayes, 2018).

- *Derive moderators from the mechanism*: ask which conditions would strengthen or disable each step, and which path (first or second stage) the moderator acts on.
- *Choose designs that match the claim*: randomize X where possible. Consider manipulating M in a follow-up study. Use longitudinal data so that M is measured between X and Y.
- *Report* indirect effects with bootstrap intervals, simple effects across the moderator's range (floodlight), and the index of moderated mediation where relevant (Hayes 2018).
- *Plan power for interactions*: detecting an attenuation typically requires much larger samples than detecting a main effect.

3.4 Choosing Among Theories

Most business phenomena can be approached through more than one theory. Executive compensation can be studied through agency theory (Section 6.1), tournament theory (Section 15.13), or social comparison. A firm's decision to outsource through transaction cost economics (Section 6.3) or capability-based reasoning. Choosing well requires matching the theory to the research question, its phenomenon, level of analysis, and the type of explanation sought, and, where theories compete, designing tests that can tell them apart.

3.4.1 Matching theory to question

Table 3.5 offers a starting point for common business research questions, organized by the chapters of this book. It is a heuristic, not an algorithm: the best study often combines two theories, using one to explain the main effect and another to explain its boundary conditions.

Table 3.5. Which theory family for which research question? A heuristic guide.

Research question	Candidate theory families	Level	Where in this book
Why do some firms earn persistently superior returns?	Resource-based theory, VRIO, dynamic capabilities	Firm	Ch. 4-5
Should the firm make, buy, or ally?	Transaction cost economics, resource-based and knowledge-based views	Transaction, firm	Ch. 4, 6
How should owners monitor and pay executives?	Agency theory, stewardship theory, upper echelons	Dyad, top team	Ch. 6

Table 3.5. Which theory family for which research question? A heuristic guide. *(continued)*

Research question	Candidate theory families	Level	Where in this book
Why do organizations adopt similar practices?	Institutional theory, legitimacy, resource dependence	Organization, field	Ch. 7
How can a firm credibly convey hidden quality?	Signaling, screening, costly signaling	Dyad, market	Ch. 8
How do relationships and network positions create value?	Social exchange, structural holes, social capital	Dyad, network	Ch. 9
Why, and how fast, do new products spread?	Diffusion of innovations, Bass model, technology acceptance	Individual, market	Ch. 10
How does national culture shape management practice?	Hofstede's dimensions, GLOBE, tight-loose	Society	Ch. 11
What drives effort and persistence at work?	Expectancy, goal setting, self-determination	Individual	Ch. 12
Why does framing as gains or losses change choices?	Prospect theory, regulatory focus	Individual	Ch. 13, 15
Why do people follow, or avoid, the choices of others?	Social comparison, social identity, optimal distinctiveness	Individual in groups	Ch. 14
How do messages change attitudes and behaviour?	Elaboration likelihood, persuasion knowledge	Individual, audience	Ch. 17

For quick navigation, the core sections are: Resource-based theory (Section 4.1), Dynamic capabilities (Section 5.1), Transaction cost economics (Section 6.3), Agency theory (Section 6.1), Stewardship theory (Section 6.2), Upper echelons theory (Section 6.4), Institutional theory (Section 7.2), Resource dependence theory (Section 7.1), Signaling theory (Section 8.2), Screening theory (Section 8.3), Social exchange theory (Section 9.1), Structural holes (Section 9.4), Diffusion of innovations (Section 10.1), Technology acceptance model (Section 10.4), Hofstede's dimensions (Section 18.2), Expectancy theory (Section 19.2), Goal-setting theory (Section 19.4), Self-determination theory (Section 19.6), Regulatory focus theory (Section 20.1), Prospect theory (Chapter 24), Social identity theory (Section 22.2), Optimal distinctiveness theory (Section 22.3), and the Elaboration likelihood model (Section 26.1).

Three further questions help narrow the choice. First, **does the theory's level match the outcome?** A theory of individual motivation will not, on its own, explain differences in firm profitability (see the discussion of levels in Chapter 2). Second, **do the theory's assumptions fit the setting?** Agency theory's assumption of goal conflict may fit a widely held public firm better than a founder-led family business. Third, **what kind of contribution is sought?** If the aim is to test a theory's boundary, the chosen theory should be one whose boundary is genuinely in doubt.

3.4.2 Strong inference and competing-theory tests

In an influential essay, J. R. Platt (1964) argued that some fields progress faster than others because they systematically practise **strong inference**: (1) devise alternative hypotheses. (2) Devise a crucial experiment (or several) whose possible outcomes will exclude one or more of the hypotheses. (3) Carry out the experiment so as to get a clean result, and (4) recycle the procedure, refining the hypotheses that remain. Platt's examples came from molecular biology and high-energy physics, but the logic applies wherever theories make divergent predictions.

The key is to find conditions under which theories *disagree*. Two theories that both predict a positive main effect cannot be distinguished by observing a positive main effect. They can be distinguished if they predict different moderators, mediators, or functional forms. Table 3.6 gives examples from business research.

Table 3.6. Competing-theory tests: finding the conditions under which theories disagree.

Question	Theory A	Theory B	Discriminating test
Does CEO duality help or hurt shareholders?	Agency: combining CEO and chair weakens monitoring (negative)	Stewardship: unified leadership empowers a pro-organizational CEO (positive)	Direction of the duality effect. Moderation by ownership structure (Donaldson & Davis, 1991)
Why do firms vertically integrate?	Transaction costs: integrate when assets are specific and opportunism is likely	Capabilities: integrate when the firm is better at the activity	Do asset specificity and capability gaps predict integration independently?
Why do degrees raise wages?	Human capital: education raises productivity	Signaling: education reveals pre-existing ability	Sheepskin effects: wage jumps at degree completion beyond years of schooling
Why does the loyalty programme raise spending?	Switching costs (economic)	Status and recognition (psychological)	Manipulate point value vs. tier visibility, examine behaviour near tier thresholds

COMMON PITFALL

Confirming a theory that had no rival. Many studies "support" a theory by finding a relationship that several other theories also predict. Before collecting data, list at least two alternative explanations for the expected result and ask what observation would favour your theory over each. If no observation would, the study can describe a relationship but cannot adjudicate among explanations, and should not claim to.

3.5 Theoretical Contribution and Modes of Reasoning

3.5.1 Building versus testing theory

Empirical work can contribute to theory in two broad ways: by **testing** existing theory, deriving hypotheses from established arguments and confronting them with data, and by **building** theory, introducing new constructs, relationships, or mechanisms. Colquitt and Zapata-Phelan (2007) rated five decades of empirical articles in the *Academy of Management Journal* on both dimensions and derived a five-category taxonomy, sketched in Figure 3.5. **Reporters** do little of either, **testers** rigorously test existing theory. **Builders** introduce new constructs or relationships with limited grounding in existing theory. **Expanders** do both, testing existing theory while introducing new elements, and **qualifiers** occupy the middle ground, testing and extending theory moderately.

Neither building nor testing is intrinsically superior. A field with only builders accumulates untested constructs. A field with only testers stops generating new ideas. Whetten's checklist of questions reviewers ask about a contribution remains a useful self-assessment tool (Table 3.7).

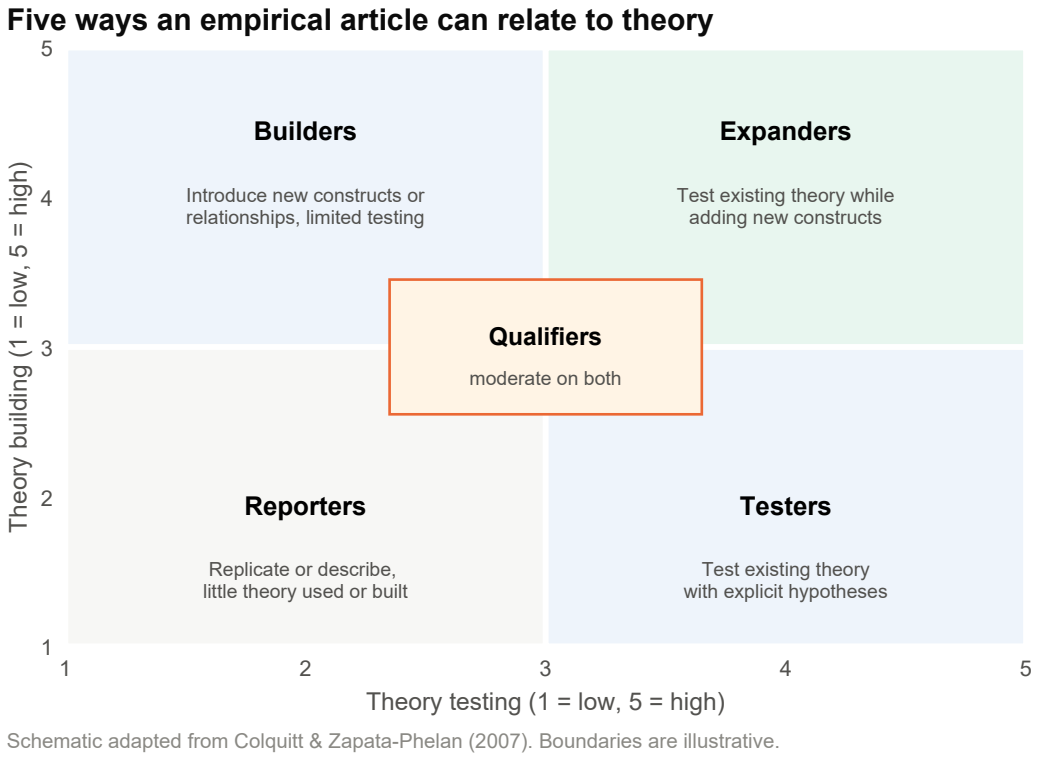


Figure 3.5. A schematic of the theory-building by theory-testing taxonomy of empirical articles proposed by Colquitt and Zapata-Phelan (2007). Each dimension is rated on a five-point scale. Region boundaries here are drawn schematically for illustration.

Table 3.7. Seven questions reviewers ask about a theoretical contribution, following Whetten (1989).

Question	What the reviewer is asking	How to answer it
What's new?	Does the work add something significant to current thinking?	State precisely which construct, relationship, mechanism, or boundary changes
So what?	Will the contribution change how scholars (or managers) think or act?	Show which prior conclusions or practices must be revised
Why so?	Is the underlying logic sound and convincing?	Make the mechanism explicit and address rival explanations
Well done?	Is the work thorough and does it reflect careful thinking?	Match design to claim, report robustness and limitations
Done well?	Is it clearly written and logically organized?	One idea per paragraph, consistent construct labels
Why now?	Is the topic of contemporary interest?	Link to a current phenomenon, debate, or gap
Who cares?	What share of the audience will find it relevant?	Identify the scholarly conversation and practitioners affected

3.5.2 Deduction, induction, and abduction

Behind the build-test distinction lie three modes of reasoning. **Deduction** derives specific predictions from general premises. It is the logic of hypothesis testing. **Induction** generalizes from observed cases to broader patterns. It is the logic of grounded theory (Glaser and Strauss 1967) and of

case-based theory building (Eisenhardt 1989b). **Abduction** reasons from a surprising observation to the explanation that would, if true, make it unsurprising. It is the logic of anomaly-driven discovery. Timmermans and Tavory (2012) argue that abductive analysis, the iterative confrontation of surprising data with multiple existing theories, is a more accurate description of how qualitative researchers actually build theory than pure induction (Table 3.8).

Table 3.8. Three modes of reasoning in theory development.

Mode	Starting point	Output	Typical designs	Main risk
Deduction	General theory	Testable hypotheses, support or refutation	Experiments, surveys, archival regressions	Testing only what the theory already says. Confirmation bias
Induction	Systematic observation of cases	New concepts and generalizations	Grounded theory, multiple-case studies	Generalizing from idiosyncratic cases
Abduction	A surprising fact or anomaly	The most plausible new explanation	Iterative qualitative analysis, exploratory data analysis followed by tests	Mistaking a plausible story for a tested one

The modes are complementary stages of a cycle rather than rival philosophies: an anomaly prompts an abductive conjecture, case evidence refines it inductively, and a deductive test confronts it with new data. A. C. Edmondson and McManus (2007) add that the appropriate method depends on the state of prior theory. Where theory is **nascent**, open-ended qualitative data help identify constructs. Where it is **intermediate**, hybrid designs combine qualitative insight with preliminary measures, and where it is **mature**, precise quantitative tests of well-specified hypotheses are appropriate. Mismatches, such as a hypothesis-testing survey on a phenomenon nobody yet understands, weaken contribution.

TRY IT YOURSELF

Take a surprising business fact you have encountered (for example, a product that sells *more* after a price increase). Write one abductive explanation, identify two existing theories that could account for it (see Table 3.5), and specify one observation that would distinguish between them. You have just designed a strong-inference study.

3.6 Causal Diagrams as Explicit Theory

Most theories in business research make causal claims, yet most data are observational. Bridging the gap requires assumptions about which variables cause which. **Directed acyclic graphs** (DAGs) make those assumptions explicit (Pearl 2009). A DAG is a diagram in which nodes are variables and arrows are direct causal effects, with no feedback loops. Drawing one forces the researcher to state the theory in a form from which the correct analysis can be derived.

Three structures matter most. A **confounder** causes both treatment and outcome and must be controlled to block a non-causal "back-door" path. A **mediator** lies on the causal path from treatment to outcome. Controlling for it removes part of the effect one may be trying to estimate. A **collider** is caused by both treatment and outcome (or their causes). Controlling for it, or selecting the sample on it, *opens* a spurious path and creates bias where none existed (Elwert and Winship 2014). Pearl's **back-door criterion** formalizes which sets of control variables suffice to identify a causal

effect given the diagram (Pearl 2009) (Figure 3.6).

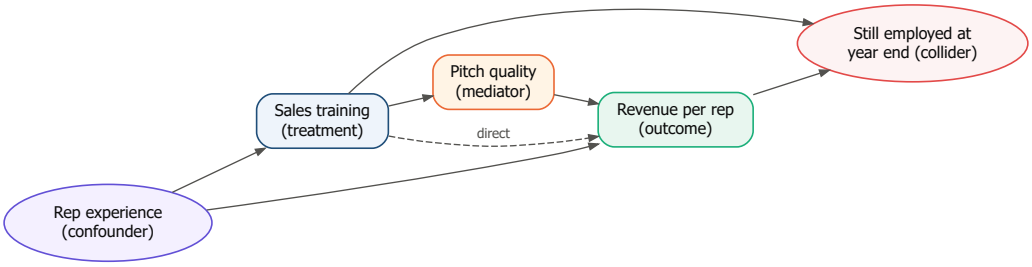


Figure 3.6. A causal diagram (DAG) for evaluating a sales-training programme. Representative experience is a confounder (control for it). Pitch quality is a mediator (do not control if the total effect is of interest). Being retained in the sample is a collider (do not condition on it). Illustrative example.

Table 3.9 presents the role of a variable in the causal diagram determines whether to control for it.

Table 3.9. The role of a variable in the causal diagram determines whether to control for it.

Variable role	Causal structure	Control for it?	Business example
Confounder	C -> X and C -> Y	Yes (blocks a back-door path)	Experienced reps both choose training and sell more
Mediator	X -> M -> Y	No, if the total effect is the target. Yes for a direct effect, with care	Training improves pitches, which raise revenue
Collider	X -> K <- Y	No (conditioning opens a spurious path)	Analysing only reps still employed, when retention depends on both training and sales
Outcome proxy / descendant of Y	Y -> Z	No	Controlling for commission earned when estimating revenue effects
Instrument	Z -> X only	Not as a control, can be used for IV estimation	Random assignment of training slots by calendar availability

BUSINESS CASE

Survivorship as collider bias. Studies of "what successful start-ups have in common" sample on success, a collider influenced by many causes at once. Among surviving firms, two causes of survival can appear negatively related even if they are independent in the full population: for instance, founder experience and market timing, if either one alone is often enough to survive. Strategy advice drawn from such samples can therefore recommend the wrong trade-offs. The remedy is theoretical before it is statistical: draw the diagram, identify selection as a node, and either study the full population at risk (including failures) or model selection explicitly (Elwert and Winship 2014).

COMMON PITFALL

Kitchen-sink controls. Adding every available variable as a control is often defended as conservative. It is not. Controlling for mediators removes part of the effect of interest. Controlling for colliders introduces bias. Every control variable should be justified by its role in a causal diagram, and the diagram should be reported. Antonakis et al. (2010) review the most

common threats to causal claims in management research and the designs that address them. Shadish, Cook, and Campbell (2002) remains the standard reference on experimental and quasi-experimental design.

3.7 The Replication Crisis and Research Integrity

Theories can only be as credible as the evidence used to test them. During the 2010s, a series of large-scale replication projects and methodological critiques forced the social sciences to confront the possibility that much published evidence was weaker than it appeared.

3.7.1 What the replication projects found

The most influential project, the Open Science Collaboration's attempt to reproduce 100 experimental and correlational studies published in three psychology journals, found that 97 percent of the original studies had reported statistically significant results but only 36 percent of the replications did. On average, replication effect sizes were about half as large as the original estimates (Open Science Collaboration 2015). Depending on the criterion used, statistical significance, whether the original effect fell within the replication's confidence interval, or subjective assessments by the replication teams, estimates of how many findings replicated ranged from roughly one-third to just under one-half. A later project replicating 21 social-science experiments published in *Nature* and *Science* between 2010 and 2015 likewise found that replication effects were systematically smaller than original effects (C. F. Camerer et al. 2018).

DEBATE AND EVIDENCE

Is there a replication crisis? Gilbert et al. (2016) argued that the Open Science Collaboration's design underestimated reproducibility: some replications departed from the original procedures or populations, and the statistical benchmark did not account for sampling error in the replications themselves. The original authors replied that these objections did not overturn their main conclusions. Most observers now agree on a narrower point that matters for theory: a single statistically significant result, especially from a small study with flexible analyses, is weak evidence, and effect sizes reported in the literature are likely inflated by selective publication. The debate has shifted from *whether* there is a problem to *which* reforms address it best (Munafò et al. 2017).

3.7.2 Why false positives proliferate

Ioannidis (2005) argued on probabilistic grounds that in fields where true effects are rare, studies are small, effects are modest, designs and analyses are flexible, and competition is intense, most claimed findings are more likely to be false than true. The argument turns on the **pre-study odds** that a tested hypothesis is true: when most hypotheses tested are false, even a nominal 5 percent false-positive rate can produce a literature in which a large share of significant findings are errors.

Simmons, Nelson, and Simonsohn (2011) showed how undisclosed **researcher degrees of freedom**, choices about which dependent variables to report, when to stop collecting data, which covariates to include, and which conditions to compare, can push the actual false-positive rate far above 5 percent even when each choice seems reasonable in isolation. Gelman and Loken (2014)

emphasize that the problem does not require deliberate fishing: a **garden of forking paths** arises whenever analytic choices are contingent on the data, even if only one analysis is ever run. Figure 3.7 reproduces the logic with a simple simulation in which there is *no* true effect.

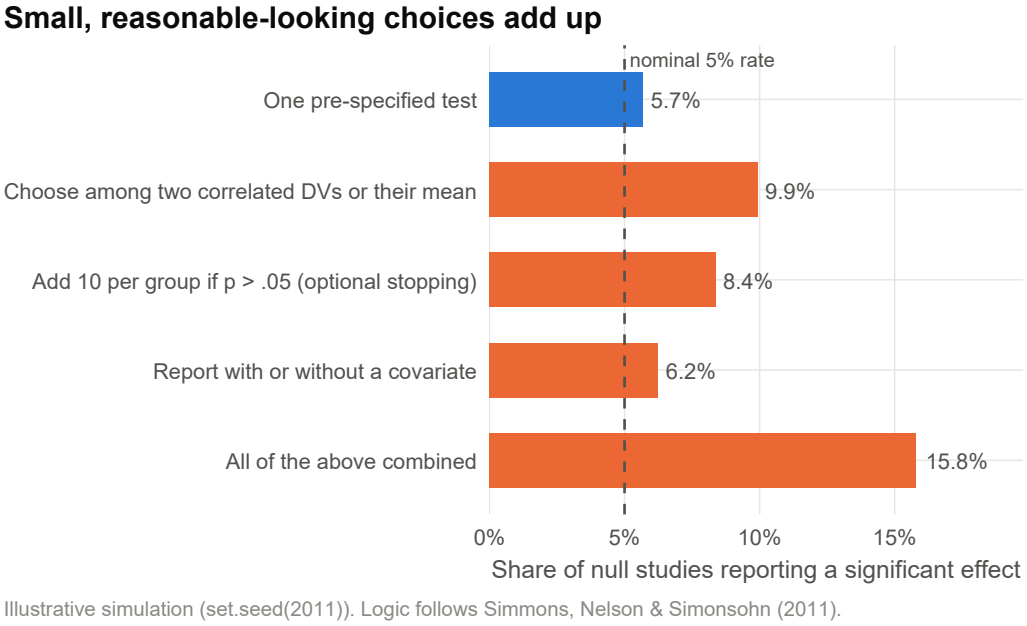


Figure 3.7. Illustrative simulation of how analytic flexibility inflates false positives. Every simulated study compares two groups (20 per group) with no true difference. Each bar shows the share of studies that report p < .05 when the researcher may choose among the listed options and reports any significant result. Simulated data. 4,000 Studies per strategy.

Two further practices compound the problem. **HARKing**, hypothesizing after the results are known, presents post hoc hypotheses as if they had been specified in advance, converting exploratory findings into apparently confirmatory tests (N. L. Kerr 1998). Survey evidence suggests that such **questionable research practices** were not rare: in a large survey of academic psychologists, substantial proportions admitted to practices such as failing to report all dependent measures or deciding whether to collect more data after looking at results (John, Loewenstein, and Prelec 2012) (Table 3.10).

Table 3.10. Questionable research practices and their remedies.

Practice	Why it inflates false positives	Remedy
Selective reporting of outcomes	Multiple chances to find p < .05. Only winners reported	Preregister primary outcomes, report all measures
Optional stopping	Repeated testing as data accrue raises the error rate	Fix sample size in advance or use sequential designs with corrected thresholds
Flexible covariates and exclusions	Choices contingent on results create forking paths	Prespecify. Report results with and without covariates and exclusions
HARKing	Exploratory findings presented as confirmatory tests	Label exploratory analyses, preregister confirmatory hypotheses
Underpowered studies	Significant results overestimate effects, low pre-study odds	Power analysis, larger samples, multi-site collaboration
Publication bias	Null results remain unpublished, distorting the literature	Registered reports, publishing replications and null results

Table 3.10. Questionable research practices and their remedies. *(continued)*

Practice	Why it inflates false positives	Remedy
Misreading p-values	Treating $p < .05$ as proof or $p > .05$ as proof of no effect	Report effect sizes and intervals, interpret in context

3.7.3 Reforms: preregistration and transparency

The most widely adopted reform is **preregistration**: recording hypotheses, designs, and analysis plans in a time-stamped public registry before observing the data. Nosek et al. (2018) argue that its value lies in distinguishing **prediction** (testing hypotheses fixed in advance) from **postdiction** (generating hypotheses from data). Both are legitimate, but only the former supports the inferential claims that p-values imply. **Registered reports**, in which journals review and accept a study's design before results are known, address publication bias directly. Complementary reforms include data and code sharing, larger and multi-site samples, and more careful statistical reporting. The American Statistical Association's statement on p-values emphasizes that a p-value does not measure the probability that a hypothesis is true or the size of an effect, and that scientific conclusions should not rest on whether a p-value crosses a threshold (Wasserstein and Lazar 2016).

For theory, the implication is direct. A theory "supported" by a body of small, flexible, selectively published studies is less well supported than it appears. Several effects discussed later in this book have faced replication challenges. Where they have, the relevant chapters say so. Readers should weigh evidence by its design and transparency, not by the number of citations a finding has accumulated.

RESEARCH TOOLKIT

A credibility checklist for theory-testing studies.

- *Prespecify* confirmatory hypotheses, primary outcomes, sample size, exclusions, and the analysis model. Label everything else exploratory (Nosek et al. 2018).
- *Power the study* for the smallest effect that would matter theoretically or practically, remembering that published effects are likely inflated.
- *Test the theory where it could fail*: include conditions under which rival theories make different predictions (J. R. Platt 1964).
- *Report fully*: all conditions, measures, and exclusions, and results with and without covariates (Simmons, Nelson, and Simonsohn 2011).
- *Share* materials, data, and code where possible, and treat a successful replication as part of the contribution rather than an afterthought.

3.8 Theory in Practice

The tools in this chapter are not confined to academic journals. Every serious attempt to improve a business decision with evidence uses them, often under different names.

3.8.1 A/B testing as theory testing

Online controlled experiments, A/B tests, are the most widespread application of the experimental method in business. Kohavi, Tang, and Xu (2020) document how large technology firms run

thousands of experiments a year to evaluate product changes. At their best, these experiments are theory tests: a product team holds a hypothesis about why a change should alter user behaviour, randomizes exposure, and measures the predicted outcome. At their worst, they are atheoretical searches in which many variants and metrics are tried and the "winners" are shipped: a corporate version of the garden of forking paths. Practitioners have developed defences that mirror academic reforms: pre-specified primary metrics, fixed experiment durations, corrections for multiple comparisons, and scepticism toward surprising results. Kohavi and colleagues call the last of these Twyman's law: a figure that looks unusually interesting or different is usually the product of an error.

BUSINESS CASE

Testing an idea nobody prioritized. Kohavi and Thomke (2017) describe an experiment at Microsoft's Bing search engine in which a modest change to how advertisement headlines were displayed, an idea that had sat in the development backlog for months because it seemed unimportant, produced one of the largest revenue improvements in the search engine's history when it was finally tested. The case is usually told as an argument for experimentation, but it is also a lesson about theory. Intuitions about which changes matter are themselves implicit theories, and they are often wrong. Experiments discipline those theories. Explicit theory, in turn, tells teams which experiments are worth running and why a result should generalize to the next product decision.

3.8.2 From theory to scale: the social-norms case

The path from a theory to a business at scale is illustrated by home energy reports. Schultz et al. (2007) showed in a field experiment that telling households how their energy use compared with their neighbours' reduced consumption among above-average users but *increased* it among below-average users, a "boomerang" consistent with social-norms theory, and that adding a simple signal of approval for low users eliminated the boomerang. The company Opower built a business on this logic, sending comparative reports to millions of utility customers. Large randomized evaluations found average reductions in energy use of about 2 percent (Allcott 2011). Allcott (2015) then showed that effects in the earliest sites were larger than in later ones, because the utilities that adopted the programme first differed systematically from those that adopted later: a reminder that **boundary conditions** shape what a theory delivers when it moves from one setting to many (see Social norms (Section 22.5)).

3.8.3 Evidence-based management

The broader practice of using theory and evidence in decisions is **evidence-based management** (Rousseau 2006, Pfeffer and Sutton 2006). It asks managers to treat their beliefs about cause and effect as hypotheses, to seek the best available evidence from research, organizational data, stakeholders, and professional experience, and to evaluate the outcomes of their decisions. Theory plays three roles in this process. It tells managers *what to look for* (which variables a mechanism implies), *where a finding should travel* (which boundary conditions limit generalization), and *how to interpret anomalies* (as signals that a theory needs revision rather than as noise to be ignored).

RESEARCH TOOLKIT

From managerial question to theory-driven test.

1. *State the decision and the implicit theory behind each option* ("if we bundle onboarding with the premium tier, churn will fall because …").
2. *Name the mechanism and at least one rival* (switching costs vs. selection of committed customers).
3. *Derive divergent predictions* (who should respond most under each account?).
4. *Choose the design* (randomized test, staggered roll-out, or quasi-experiment) and pre-specify the primary metric and sample size.
5. *Interpret against the theory*, including boundary conditions, before generalizing to other markets or products.

COMMON PITFALL

Treating a single test as a law. A successful experiment shows that a change worked in one population, at one time, for one metric. Without a theory of *why* it worked, a firm cannot know whether it will work in a new market, for a new customer segment, or after competitors respond. Replicate important results, and use the mechanism to decide where they should, and should not, travel.

3.9 Chapter Summary

KEY TAKEAWAYS

- Hypotheses translate construct-level propositions into falsifiable claims about variables. A strong hypothesis-development paragraph states the mechanism and a boundary condition, not just prior findings.
- Constructs earn meaning through their **nomological network**. Measures must show convergent and discriminant validity (Cronbach and Meehl 1955, D. T. Campbell and Fiske 1959).
- **Mediation** tests the *why* (indirect effect a x b, bootstrapped). **Moderation** tests the *when* (interactions, floodlight analysis) (X. Zhao, Lynch, and Chen 2010, Spiller et al. 2013, Hayes 2018). Mechanism claims are strongest when mediators are manipulated (Spencer, Zanna, and Fong 2005, Bullock, Green, and Ha 2010).
- Choose theories by fit to question, level, and assumptions. Adjudicate between rivals with **strong inference**: find where they disagree (J. R. Platt 1964).
- Empirical work can build theory, test it, or both (Colquitt and Zapata-Phelan 2007). Deduction, induction, and abduction are complementary stages of theory development.
- **Causal diagrams** make a theory's assumptions explicit and determine which variables to control, and which not to (Pearl 2009, Elwert and Winship 2014).
- Replication projects found that many published findings are weaker than reported. In one large project only 36 percent of replications were significant (Open Science Collaboration 2015). Flexibility, HARKing, and publication bias inflate false positives (Simmons, Nelson, and Simonsohn 2011, N. L. Kerr 1998, Ioannidis 2005). Preregistration and transparency help (Nosek et al. 2018).
- In practice, A/B tests, scaled interventions, and evidence-based management are all

forms of theory testing (Kohavi, Tang, and Xu 2020, Allcott 2015, Rousseau 2006).

Table 3.11 presents the comparison of the chapter's tools for using theory.

Table 3.11. Comparison of the chapter's tools for using theory.

Tool	Question answered	Signature practice	Common misuse	Key source
Hypothesis development	What exactly does the theory predict here?	Mechanism plus boundary plus directional prediction	Theory as decoration	Bacharach (1989)
Nomological network	Does the measure capture the construct?	Convergent and discriminant validation	Single-item metrics adopted on faith	Cronbach & Meehl (1955)
Mediation	How or why does X affect Y?	Bootstrapped indirect effect, manipulating M	Causal claims from cross-sectional mediation	Zhao et al. (2010)
Moderation	When or for whom?	Interaction with floodlight analysis	Arbitrary spotlights, underpowered tests	Spiller et al. (2013)
Strong inference	Which rival explanation survives?	Crucial tests where theories diverge	Confirming a theory without rivals	Platt (1964)
Causal diagrams	Which controls identify the effect?	Back-door criterion	Kitchen-sink controls, collider bias	Pearl (2009)
Preregistration	Is this a prediction or a postdiction?	Time-stamped plans, registered reports	HARKing, forking paths	Nosek et al. (2018)

3.10 Discussion Questions

1. Rewrite a hypothesis-development paragraph from a published article you know using the five moves in Table 3.2. Which move was missing in the original?
2. A firm finds that employees who attend voluntary leadership training are promoted faster. Draw a causal diagram that includes at least one confounder, one mediator, and one collider. What analysis does your diagram imply?
3. Why might an indirect effect be significant even when the total effect of X on Y is not? Give a business example in which two mechanisms work in opposite directions.
4. Choose a pair of competing theories from Table 3.6 and design a strong-inference study that could favour one over the other.
5. Your company's experimentation platform reports that 30 percent of last quarter's A/B tests "won". What questions would you ask before concluding that the product team's ideas are unusually good?
6. Is preregistration appropriate for theory-*building* research? Discuss how abductive and inductive studies can be made transparent without being preregistered in the confirmatory sense.

3.11 Further Reading

- Cronbach and Meehl (1955): the foundational argument that measurement validity is inseparable from theory.
- X. Zhao, Lynch, and Chen (2010): a concise and influential correction to the causal-steps approach to mediation. Pair with Hayes (2018) for implementation.

- Spiller et al. (2013): a clear guide to probing interactions with spotlight and floodlight analyses.
- J. R. Platt (1964): a short classic on how fields progress by testing rival hypotheses.
- Colquitt and Zapata-Phelan (2007): a map of how empirical articles build and test theory.
- Pearl (2009): the formal foundation of causal diagrams and identification. Elwert and Winship (2014) is a more accessible entry point on collider bias.
- Open Science Collaboration (2015) and Simmons, Nelson, and Simonsohn (2011): the empirical and methodological core of the replication debate.
- Kohavi, Tang, and Xu (2020): the standard practitioner guide to trustworthy online experiments.

Part II

Organizations and Strategy

Chapter 4

Resources and Competitive Advantage

In the 1980s Toyota did something that puzzled Western competitors: it let them in. Executives from rival carmakers toured Toyota's plants, and in 1984 General Motors and Toyota opened a joint venture in Fremont, California (NUMMI) that put the Toyota Production System inside a GM-staffed factory. The methods were visible, documented, and widely taught. Yet for decades few rivals matched Toyota's combination of quality and productivity (Womack, Jones, and Roos 1990, Spear and Bowen 1999). Why would a firm give away its playbook, and why did the playbook not travel?

The same puzzle appears across industries. Walmart's distribution system was described in business school cases for years before rivals approached its cost position. Disney's characters are known to every child, yet no competitor can simply build a comparable library. Zara's quick-response supply chain has been analysed in countless articles, yet it remains distinctive. Industry structure cannot explain these differences, because the firms face the same suppliers, customers, and substitutes as their rivals. The explanation must lie *inside* the firm: in what it owns, what it knows, and how it combines the two.

This chapter develops the family of theories that locate competitive advantage in firm-specific **resources**. We begin with the **resource-based view (RBV)**, its origins, assumptions, and central propositions, and its best-known diagnostic, the **VRIO framework**. We then examine the **knowledge-based view**, which treats knowledge as the most strategically important resource, and the related idea of **core competences**. Two marketing-born extensions follow: **resource-advantage theory**, which embeds resources in a dynamic theory of competition, and the **market-based assets** framework, which links customer and channel relationships to shareholder value. Chapter 5 then asks how firms *renew* their resources when environments change (Table 4.1).

Table 4.1. Chapter at a glance: resource-based theories of competitive advantage.

Theory	Core question	Level of analysis	Key source
Resource-based view	Why do some firms persistently outperform rivals in the same industry?	Firm	Wernerfelt (1984), Barney (1991)
VRIO framework	Which resources can sustain an advantage, and what must the firm do to exploit them?	Resource / firm	Barney (1991), Barney & Hesterly (2012)
Knowledge-based view	Why do firms exist, and how do they create value by integrating knowledge?	Firm / team	Kogut & Zander (1992), Grant (1996)

Table 4.1. Chapter at a glance: resource-based theories of competitive advantage. *(continued)*

Theory	Core question	Level of analysis	Key source
Core competences	Which collective skills underpin a firm's portfolio of businesses?	Corporation	Prahalad & Hamel (1990)
Resource-advantage theory	How does competition among heterogeneous firms unfold over time?	Firm / market	Hunt & Morgan (1995), Hunt (2000)
Market-based assets	How do customer and channel relationships create shareholder value?	Firm / market	Srivastava, Shervani, & Fahey (1998)

4.1 The Resource-Based View

THEORY CARD

Core claim. Firms differ in the bundles of resources and capabilities they control. When a resource is valuable, rare, and costly for rivals to imitate or substitute, and the firm is organized to exploit it, it can yield a **sustained competitive advantage** and above-normal returns.
Origins. Strategic management, with roots in economics: Penrose (1959) on the firm as a bundle of productive resources, Wernerfelt (1984) on resource position barriers, J. B. Barney (1986) on strategic factor markets, Dierickx and Cool (1989) on asset stock accumulation, Jay Barney (1991) on the conditions for sustained advantage, M. A. Peteraf (1993) on the cornerstones of advantage. J. B. Barney, Ketchen, and Wright (2011) argue that the perspective has matured into a *theory*, often called **resource-based theory (RBT)**.
Level of analysis. Primarily the firm (business unit). Also the resource, process, and exchange-relationship levels (Kozlenkova, Samaha, and Palmatier 2013, Ray, Barney, and Muhanna 2004).
Key constructs. Resources, capabilities, resource heterogeneity, resource immobility, value, rarity, imitability, substitutability, isolating mechanisms, economic rent, sustained competitive advantage.
Boundary conditions. Strongest where resources are firm-specific and traded in imperfect factor markets. Weaker in markets with rapid, discontinuous change, where the value of today's resources can evaporate (see Dynamic capabilities (Section 5.1)).
Related theories. VRIO, Knowledge-based view, Resource-advantage theory, Market-based assets, Dynamic capabilities (Section 5.1), Transaction cost economics (Section 6.3), Five forces (Section 31.4).

4.1.1 Origins and intellectual history

The resource-based view emerged as a corrective to the dominant strategy paradigm of the early 1980s. Industrial-organization economics, translated into strategy by Michael E. Porter (1980a), explained profitability through **industry structure**: firms earn high returns when entry barriers, buyer and supplier power, and rivalry are favourable. That perspective treats firms within an industry as largely similar, or as differing only in the strategies they choose. Yet a recurring empirical observation is that performance differences *within* industries are large and persistent. Something firm-specific was missing.

The intellectual seed was planted much earlier. Penrose (1959) described the firm as a collection of

productive resources organized within an administrative framework, whose growth is driven by the *services* those resources can render and limited by the managerial capacity to plan and coordinate expansion. Because resources are lumpy and managers learn on the job, firms always carry unused productive services, which create an internal incentive to grow into new activities: an idea that later shaped theories of diversification and of capabilities.

Wernerfelt (1984) coined the phrase "resource-based view of the firm" and proposed analysing firms from the resource side rather than the product side. He introduced **resource position barriers**, the resource-side analogue of entry barriers: a firm that holds a resource first can make it harder for later holders to catch up. J. B. Barney (1986) added a crucial economic argument. If resources could be bought in competitive **strategic factor markets**, their price would reflect their expected value and no buyer could earn above-normal returns from them. Superior returns must therefore come from resources that are acquired with better expectations or luck, or, more importantly, that are developed *inside* the firm rather than bought.

Dierickx and Cool (1989) sharpened that point. Strategic assets are **stocks** that are accumulated through a stream of **flows** (investment in R&D, advertising, training) over time. They cannot be bought ready-made. What sustains an advantage is the difficulty rivals face in replicating the accumulation path. The capstone statement came with Jay Barney (1991), who specified the conditions under which resources yield **sustained competitive advantage**, the VRIN criteria (valuable, rare, imperfectly imitable, non-substitutable), and M. A. Peteraf (1993), who translated the argument into the language of economic rents. Reviews soon followed (Mahoney and Pandian 1992, Amit and Schoemaker 1993, J. Barney and A. Arikan 2001), and by the 2000s the RBV had become the most widely used theory in strategic management and a major lens in marketing (Kozlenkova, Samaha, and Palmatier 2013, J. B. Barney, Ketchen, and Wright 2011).

4.1.2 Resources and capabilities

DEFINITION

Definition 4.1. Resources are "tangible and intangible assets firms use to conceive of and implement its strategies" (J. Barney and A. Arikan 2001, 138).
A **capability** is a special type of resource: an "organizationally embedded non-transferable firm-specific resource whose purpose is to improve the productivity of the other resources possessed by the firm" (Makadok 2001, 389).

The distinction matters. A resource is something the firm *has*: a patent, a brand, a plant location, a pool of cash. A capability is something the firm *can do* by deploying and combining resources: designing cars quickly, replenishing stores accurately, launching films globally. Amit and Schoemaker (1993) make the same distinction: resources are stocks of available factors owned or controlled by the firm, and capabilities are the firm's capacity to deploy them through organizational processes. Capabilities are typically more firm-specific than resources because they are embedded in routines, relationships, and tacit knowledge, and therefore harder to trade or copy.

Makadok (2001) uses the distinction to identify two distinct mechanisms of rent creation. **Resource picking** generates rents by being better than rivals at identifying and acquiring undervalued resources in factor markets (an argument that follows J. B. Barney (1986)). **Capability building** generates rents by making the resources the firm already owns more productive. The two mechanisms can be complements or substitutes, and they place different demands on managers: picking

rewards superior information and judgement, whereas building rewards organizational design and patience.

Scholars have proposed several classifications of resources. Jay Barney (1991) distinguishes physical, human, and organizational capital. Later textbook treatments add financial capital (J. B. Barney and Hesterly 2012), and Hunt (2000) proposes a seven-category typology that is widely used in marketing. Table 4.2 adopts Hunt's categories and maps them to typical examples and to their usual degree of imitability.

Table 4.2. A typology of firm resources (categories follow Hunt, 2000. Examples are illustrative).

Category	What it includes	Business example	Typical ease of imitation
Financial	Cash reserves, access to capital markets, borrowing capacity	A cash-rich firm able to fund acquisitions in a downturn	Easy: money is fungible
Physical	Plants, equipment, locations, raw-material access	Distribution centres positioned near clusters of stores	Moderate: can be built, but location may be pre-empted
Legal	Patents, trademarks, copyrights, licences, contracts	Copyrighted characters and registered trademarks	Hard while protection lasts
Human	Skills, judgement, and experience of individual employees	Experienced animators, engineers, or buyers	Moderate: people can be hired away
Organizational	Routines, structures, controls, culture, policies	Shop-floor problem-solving routines	Hard: embedded and tacit
Informational	Knowledge about customers, markets, technologies, competitors	Store-level demand data feeding design decisions	Moderate to hard
Relational	Relationships with customers, suppliers, partners, regulators	Long-term, high-trust supplier networks	Hard: socially complex and history-dependent

In marketing, the resources most often studied are intangible: **brands**, **relational assets** such as customer and channel relationships, and **knowledge** about markets (Kozlenkova, Samaha, and Palmatier 2013). Marketing scholars use the RBV to justify investment in marketing activities whose pay-off is long-term rather than immediate, a theme developed further in the section on market-based assets.

4.1.3 Assumptions and core logic

The RBV rests on two assumptions (Jay Barney 1991, M. A. Peteraf and Barney 2003):

1. **Resource heterogeneity.** Firms in the same industry control different bundles of resources and capabilities. As a result, firms are skilled at different activities, and some are more efficient or effective than others.
2. **Resource immobility.** Some resources cannot be traded freely across firms, or can be traded only at high cost. Differences in resource endowments can therefore persist over time.

If both assumptions hold, some firms will possess resources that others find too costly or too difficult to obtain, and these firms can earn a **sustained competitive advantage** (J. B. Barney and Hesterly 2012). Jay Barney (1991) defines a competitive advantage as implementing a value-creating strategy that is not simultaneously being implemented by current or potential competitors. It is *sustained* when those competitors are unable to duplicate its benefits. Note that "sustained" refers

to resistance to imitation, not to a specific calendar period. An advantage can be sustained and still be destroyed by an unanticipated change in technology or demand.

M. A. Peteraf (1993) translated these ideas into four **cornerstones** of competitive advantage, each of which is necessary:

- **Heterogeneity.** Superior resources yield **Ricardian rents**, returns to scarce, more efficient factors, or monopoly rents.
- **Ex post limits to competition.** After a firm gains superior position, forces such as imperfect imitability and imperfect substitutability must prevent rivals from competing the rents away.
- **Imperfect mobility.** Resources that are firm-specific, or whose value depends on co-specialized assets, remain bound to the firm, so the rents stay with it rather than being bid away by the resource owners.
- **Ex ante limits to competition.** Before the superior position is established, there must have been limited competition for it. Otherwise, the cost of acquiring the position would have offset its rents (J. B. Barney 1986).

The theory's central propositions can be stated as follows:

- **P1 (Value).** Resources that enable a firm to conceive or implement strategies that improve efficiency or effectiveness are a source of competitive advantage only if they lower net costs or raise net revenues relative to what would otherwise be possible (J. B. Barney and Arikan 2017).
- **P2 (Rarity).** Valuable resources held by many competitors produce only **competitive parity**.
- **P3 (Inimitability).** Valuable, rare resources produce a *sustained* advantage only if competitors face a cost disadvantage in obtaining or substituting for them.
- **P4 (Organization).** The performance consequences of valuable, rare, and inimitable resources depend on whether the firm is organized, through structures, controls, and compensation, to exploit them (J. B. Barney and Hesterly 2012).
- **P5 (Accumulation).** The most durable advantages arise from resources that must be accumulated over time through internal investment, rather than bought in factor markets (Dierickx and Cool 1989).

4.1.4 Mechanism: asset stock accumulation

Dierickx and Cool (1989) offer a vivid metaphor. A strategic asset is like water in a bathtub: the level (the stock) rises with inflows (investment) and falls with outflows (depreciation, forgetting, obsolescence). Because the stock can be changed only gradually, an incumbent's lead can persist even when rivals are willing to spend heavily. Several properties of the accumulation process make imitation slow or costly:

- **Time compression diseconomies.** Doubling the rate of investment does not halve the time needed to build the stock. A crash R&D programme, an intensive brand-building blitz, or a rushed training effort produces less than the same spending spread over time.
- **Asset mass efficiencies. Existing stocks make further accumulation easier** "success breeds success". A firm with a large base of customers learns more from each new customer. A firm with a deep patent portfolio finds it easier to generate the next invention.

- **Interconnectedness of asset stocks.** Building one stock may require another (a dealer service network makes it easier to build a reputation for reliability).
- **Asset erosion.** Stocks decay without maintenance. Imitators must outrun the incumbent's own replenishment.
- **Causal ambiguity.** It may not be clear which flows built the stock, so rivals do not know what to copy.

Figure 4.1 illustrates the first two properties. An incumbent invests steadily from year one. An imitator enters in year 15 and spends two and a half times as much each year. A naive projection that ignores accumulation dynamics suggests a quick catch-up. With time compression diseconomies and the incumbent's mass efficiencies, catch-up takes much longer.

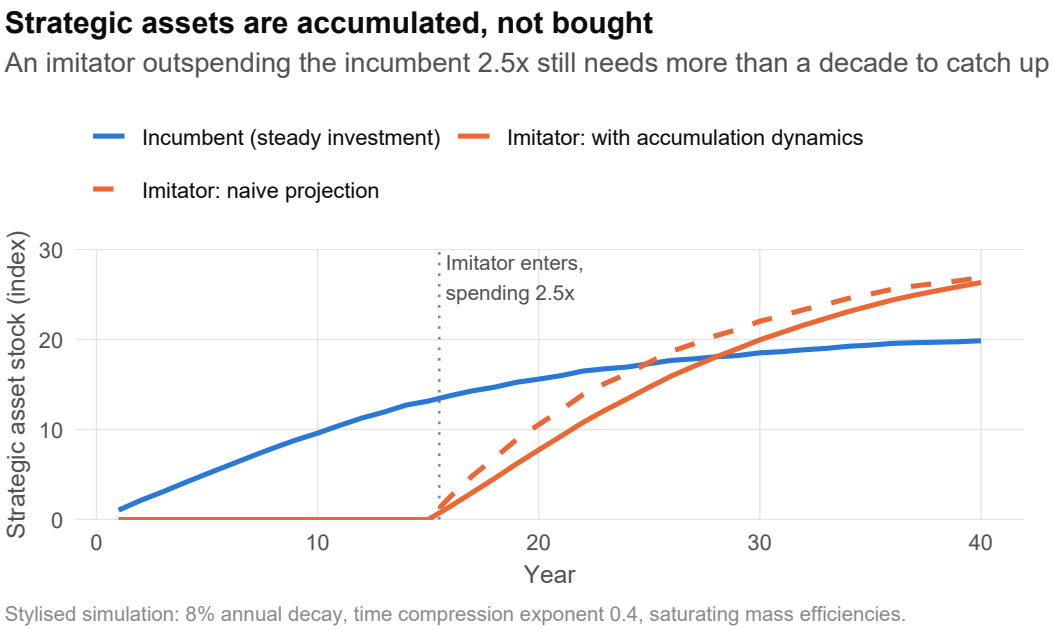


Figure 4.1. Illustrative simulation of asset stock accumulation (Dierickx & Cool, 1989). An imitator that enters late and outspends the incumbent catches up more slowly than a naive projection implies, because accelerated investment runs into time compression diseconomies while the incumbent enjoys asset mass efficiencies.

4.1.5 Isolating mechanisms: why imitation fails

The term **isolating mechanisms** (Rumelt 1984) describes the forces that protect a firm's rents from being competed away, much as mobility barriers protect groups of firms within an industry. Lippman and Rumelt (1982) showed formally that **uncertain imitability**, uncertainty about the causes of a rival's efficiency, can sustain performance differences in equilibrium even when entry is free: would-be imitators cannot be sure that copying will succeed, so fewer of them try. Jay Barney (1991) identified three reasons why resources may be imperfectly imitable: **unique historical conditions** (path dependence), **causal ambiguity**, and **social complexity**. Reed and DeFillippi (1990) elaborated causal ambiguity, arguing that it arises from the **tacitness**, **complexity**, and **specificity** of the skills and resources underlying an advantage. Table 4.3 summarises the main mechanisms.

Table 4.3. Isolating mechanisms that make resources costly to imitate.

Mechanism	Logic	Key sources	Business illustration
Unique historical conditions (path dependence)	The resource was built along a path that cannot be re-run. Being early or lucky mattered	Barney (1991), Dierickx & Cool (1989)	A character library accumulated over many decades
Causal ambiguity	Neither rivals (nor sometimes the firm) know exactly which factors produce the advantage	Lippman & Rumelt (1982), Reed & DeFillippi (1990)	Tacit rules behind a production system that is open to visitors
Social complexity	The resource rests on culture, trust, reputation, or relationships that cannot be engineered quickly	Barney (1991)	High-trust, long-term supplier relationships
Time compression diseconomies	Accelerated investment yields less than the same investment spread over time	Dierickx & Cool (1989)	Reputation for reliability, scientific know-how
Asset mass efficiencies	Existing stocks make further accumulation cheaper	Dierickx & Cool (1989)	A large installed base attracts complementary developers
Complementarity and system fit	Value depends on many interlocking activities, copying one element is not enough	Porter (1996), Milgrom & Roberts (1995)	A vertically integrated quick-response supply chain
Legal protection	Property rights block direct copying for a period	Teece (1986)	Patents, copyrights, trademarks

Causal ambiguity has a double edge. The ambiguity that blocks imitators may also prevent the firm from replicating its own success in a new plant, region, or business, and from understanding which resources it should protect. The tension between protecting knowledge from rivals and transferring it internally is a central theme of the knowledge-based view.

4.1.6 Resource orchestration

A common criticism of early RBV work was that it said much about *which* resources matter but little about what managers *do* with them. Sirmon, Hitt, and Ireland (2007) responded with a process model of **resource management**, later broadened under the label **resource orchestration**. Value is created when managers **structure** the resource portfolio (acquiring, accumulating, and divesting resources), **bundle** resources into capabilities (stabilizing, enriching, and pioneering), and **leverage** those capabilities in the market (mobilizing, coordinating, and deploying them). The model emphasises that the same resources can produce very different outcomes depending on how they are managed, and that the right orchestration depends on environmental uncertainty (Figure 4.2).

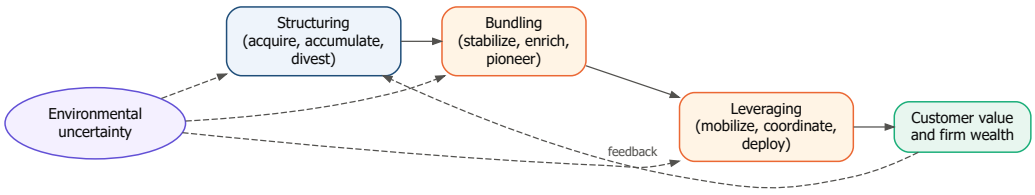


Figure 4.2. Resource orchestration: managers structure the resource portfolio, bundle resources into capabilities, and leverage capabilities to create value. Environmental uncertainty conditions each step. Adapted from the process model in Sirmon, Hitt, and Ireland (2007).

4.1.7 Evidence and extensions

Empirical testing of the RBV has been both voluminous and contested. Three contributions anchor the evidence base:

- **Systematic review.** Newbert (2007) reviewed the empirical RBV literature and found only modest overall support. Support was stronger for studies that examined *capabilities* and *core competences* than for those that measured static resource endowments, and he called for designs that test the full causal chain from resource attributes to advantage to performance.
- **Meta-analysis.** Crook et al. (2008) meta-analysed the resource-performance relationship and found it to be positive and meaningful. Importantly, the relationship was stronger for resources that met the theory's criteria for strategic resources than for resources in general: evidence that the VRIN attributes matter, not just resource quantity.
- **Choice of dependent variable.** Ray, Barney, and Muhanna (2004) argued that firm-level performance is too distal: a valuable resource may create advantage in one business process while being offset elsewhere. In a study of customer service in insurance firms, they showed that intangible, socially complex resources predicted process-level performance: a design now widely recommended.

DEBATE AND EVIDENCE

Is the RBV a useful theory? In an influential critique, Priem and Butler (2001) argued that the RBV is close to **tautological**: if a valuable resource is defined as one that produces advantage, then the statement "valuable resources produce advantage" cannot be false. They further argued that resource *value* is determined outside the theory, by product-market demand, so the RBV is silent on what makes a resource valuable. That it is largely **static**, and that it offers few actionable prescriptions. In reply, J. B. Barney (2001) contended that the theory's propositions are empirically testable because value, rarity, and imitability can be measured independently of performance. He accepted that value is determined by the environment and called for integrating resource-based logic with models of product-market demand. The exchange spurred work on **demand-side** perspectives, process-level tests (Ray, Barney, and Muhanna 2004), resource management (Sirmon, Hitt, and Ireland 2007), and the dynamic extensions discussed in Chapter 5. A related debate concerns **measurement**: resources that are causally ambiguous to rivals are, by construction, hard for researchers to observe, which pushes empirical work toward proxies (R&D stocks, advertising stocks, employee tenure) that may not capture the theory's constructs.

4.1.8 Applications in business

BUSINESS CASE

Toyota Production System. Toyota's production system, just-in-time flow, built-in quality (*jidoka*), the authority of any worker to stop the line, and continuous improvement (*kaizen*), has been documented in detail since the 1980s (Ohno 1988, Womack, Jones, and Roos 1990). That it remained hard to imitate is a textbook case of **causal ambiguity** and **social complexity**. Spear and Bowen (1999) argued that observers copied the visible tools (kanban cards, andon cords) while missing the tacit rules that govern how work is specified, how people connect, how flow paths are designed, and how improvements are made through structured experimentation. The NUMMI joint venture is a natural illustration: Toyota's methods worked

in a plant staffed largely by former GM workers, yet GM found it difficult to transfer what it learned across its wider organization. A resource that is *visible* is not necessarily *imitable*.

BUSINESS CASE

Walmart's logistics system. Walmart grew by locating stores in clusters around its own distribution centres, by moving goods through **cross-docking** (goods arriving at a distribution centre are transferred to outbound trucks with minimal storage), and by investing early in information systems that linked stores, distribution centres, and suppliers. Stalk, Evans, and Shulman (1992) used the company to argue that advantage comes from *business processes* rather than individual products. In RBV terms, the advantage rested on complementary physical, informational, and relational resources accumulated over decades: a system that could not be bought in a factor market. The case also illustrates Priem and Butler's point that value is set by the environment: the rise of e-commerce changed which logistics capabilities matter, forcing even strong incumbents to build new ones.

4.1.9 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms with larger accumulated stocks of R&D and advertising will earn more persistent abnormal returns than firms with similar current spending but smaller accumulated stocks (asset stock accumulation).
- *H2.* The positive effect of a resource on process-level performance will be stronger when the resource is socially complex (e.g., cross-functional relationships) than when it is purchasable (e.g., off-the-shelf software).
- *H3.* The performance benefit of a valuable, rare resource will be greater when the firm's compensation and control systems are aligned with its exploitation (the "O" in VRIO).

Designs. Panel data with resource stocks computed from flows (perpetual inventory method). Process-level outcomes as dependent variables (Ray, Barney, and Muhanna 2004). Natural experiments that shock resource value (regulatory change, technology shifts). Multi-level designs that separate resource effects from industry and corporate effects. Qualitative process studies of how resources are built.

Measures. Resource stocks (R&D, advertising, patents, brand equity (Section 27.3)), human-capital indicators (tenure, education, turnover), relational measures (supplier tenure, alliance count), and survey scales of value, rarity, and imitability. (Kozlenkova, Samaha, and Palmatier 2013) tabulate how marketing studies have operationalized these attributes.

Data sources. Compustat and CRSP, patent databases, brand-equity indices, employer-review platforms for culture proxies, industry-specific regulatory filings.

COMMON PITFALL

Inferring resource value from performance. The most common misuse of the RBV is to observe a high-performing firm, identify something distinctive about it, and conclude that the distinctive attribute caused the performance. This is the tautology Priem and Butler warned about, compounded by survivor bias: firms that invested in the same attribute and failed are

invisible. Measure resource attributes independently of the outcome, include unsuccessful firms, and specify the mechanism linking the resource to customer value or cost.

4.2 The VRIO Framework

THEORY CARD

Core claim. A resource or capability yields a sustained competitive advantage only if it is **V**aluable, **R**are, costly to **I**mitate, and the firm is **O**rganized to exploit it.
Origins. Jay Barney (1991) proposed the VRIN criteria (with non-substitutability as the fourth condition). Barney's later formulation, popularised in his strategy textbook, replaced N with O and absorbed substitutability into imitability (J. B. Barney and Hesterly 2012).
Level of analysis. Individual resource or capability within a firm.
Key constructs. Value, rarity, imitability (including substitution), organization, competitive implication (disadvantage, parity, temporary advantage, sustained advantage).
Boundary conditions. A diagnostic, not a predictive model: assessments are relative to a set of competitors and a period of time, and judgements of value depend on the product-market environment.
Related theories. Resource-based view, Resource-advantage theory, Dynamic capabilities (Section 5.1).

4.2.1 The four questions

Value. Firm resources are valuable if they enable a firm to develop and implement strategies that lower its net costs or raise its net revenues beyond what would have been possible without them (J. B. Barney and Arikan 2017). Value is always relative to a market: the same resource can be valuable in one environment and worthless in another. A large fleet of physical stores was valuable in pre-internet retailing. In some categories it became a cost burden.

Rarity. A resource is rare when it is held by few current or potential competitors. Valuable but common resources, modern accounting systems, standard manufacturing equipment, are necessary to compete but produce only parity. Rarity need not mean uniqueness: a resource held by fewer firms than are needed to generate perfect competition may suffice (Jay Barney 1991).

Imitability. A valuable, rare resource yields only a *temporary* advantage if rivals can obtain it at no cost disadvantage, either by **duplication** (building the same resource) or **substitution** (building a different resource that produces equivalent value). Imitability is reduced by the isolating mechanisms in Table 4.3.

Organization. Finally, firms must be "organized to exploit the full competitive potential of its resources and capabilities" (J. B. Barney and Hesterly 2012, 94). Organization includes formal reporting structures, management control systems, and compensation policies. These are **complementary resources**: they have limited ability to create advantage on their own but allow the firm to realise the potential of its other resources. A research laboratory that invents a breakthrough technology but whose parent lacks the structures to commercialise it illustrates a resource that fails the O test.

Figure 4.3 presents the framework as a sequential decision flow, and Table 4.4 applies it to firms

discussed in this chapter.

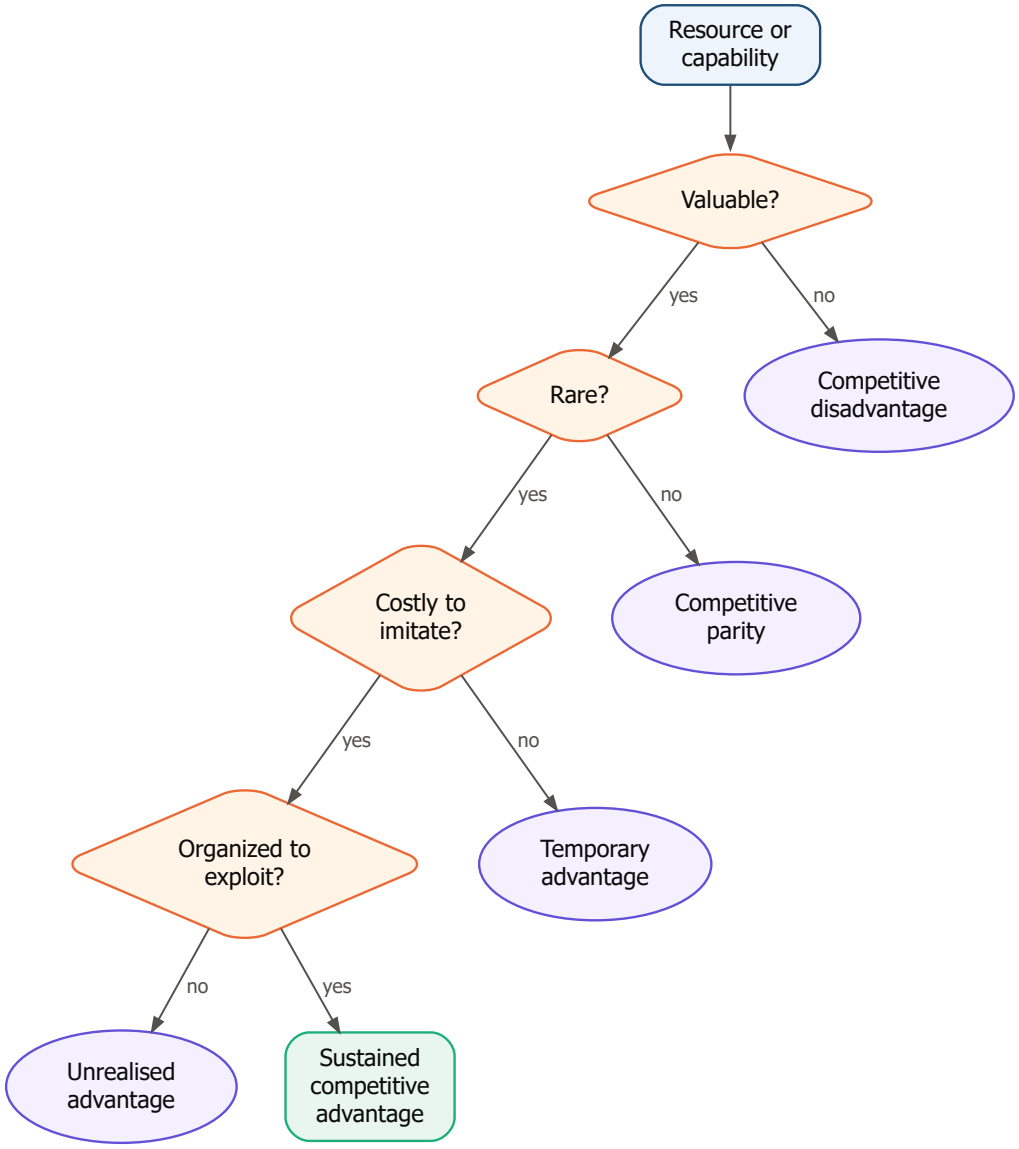


Figure 4.3. The VRIO framework as a decision flow. Each 'no' answer determines the competitive implication of a resource. Only resources that pass all four tests support a sustained competitive advantage. Based on Barney (1991) and Barney and Hesterly (2012).

Table 4.4. Illustrative VRIO assessments. Entries for named firms are interpretive judgements based on public accounts, intended for teaching rather than as empirical measurements.

Firm	Resource or capability	V	R	I	O	Competitive implication
Toyota	Production system: problem-solving routines and supplier relations	Yes	Yes	Yes (tacit, socially complex)	Yes	Sustained advantage (historically long-lived)

Table 4.4. Illustrative VRIO assessments. Entries for named firms are interpretive judgements based on public accounts, intended for teaching rather than as empirical measurements. *(continued)*

Firm	Resource or capability	V	R	I	O	Competitive implication
Walmart	Store-cluster logistics and supplier information systems	Yes	Yes	Partly (eroded by new logistics models)	Yes	Long-lived advantage, now contested
Disney	Character IP and family-entertainment brand	Yes	Yes	Yes (legal protection, history)	Yes	Sustained advantage
Zara (Inditex)	Quick-response design, production, and distribution system	Yes	Yes	Yes (system complementarity)	Yes	Sustained advantage
Apple	Integrated hardware, software, and services ecosystem	Yes	Yes	Yes (complementarity, installed base)	Yes	Sustained advantage
Illustrative firm A	Off-the-shelf enterprise software	Yes	No	-	-	Competitive parity
Illustrative firm B	Patented technology without commercialisation units	Yes	Yes	Yes	No	Unrealised advantage
Illustrative firm C	Legacy plant network in a shrinking channel	No	-	-	-	Competitive disadvantage

4.2.2 Complementarity and resource bundles

VRIO is usually applied one resource at a time, but advantage often resides in *combinations*. Two resources are **complements** when the return to one is higher in the presence of the other (P. Milgrom and Roberts 1995). In marketing, N. A. Morgan, Slotegraaf, and Vorhies (2009) show that the profit-growth effects of marketing capabilities depend on their interplay with other firm capabilities, and Kozlenkova, Samaha, and Palmatier (2013) highlight **complementarity** as one of two distinctive features of marketing resources (the other being **intangibility**). Michael E. Porter (1996) makes the strategic point: when a strategy rests on a system of mutually reinforcing activities, a rival must imitate the entire system to match the outcome, and the probability of copying every element correctly is far lower than the probability of copying any one. Teece (1986) adds that innovators often fail to profit from their innovations when **complementary assets**, manufacturing, distribution, service, or complementary technologies, are controlled by others.

BUSINESS CASE

Apple's ecosystem complementarity. Apple's products are rarely the first in their categories, yet the company has captured a large share of industry profits in several of them. A resource-based explanation emphasises complementarity among hardware design, proprietary operating systems, services, and retail. The App Store (opened in 2008) turned the installed base of iPhones into an asset that attracted developers, whose applications in turn raised the value of the device: an asset mass efficiency in Dierickx and Cool's terms. Apple's own retail stores (from 2001) and its later in-house processors (Apple silicon, introduced in Macs in 2020) extended control over complementary assets. Each component could, in principle, be imitated. The integrated system is much harder to replicate, which is precisely the logic of ecosystem theories (Jacobides, Cennamo, and Gawer 2018, Teece 1986).

TRY IT YOURSELF

Choose a firm you know well. List five resources or capabilities you believe contribute to its performance. For each, answer the four VRIO questions relative to its two closest rivals and record your confidence in each answer. Then ask: which resources are *complements*, and would your VRIO verdict change if you assessed the bundle rather than the individual resource?

4.3 The Knowledge-Based View

THEORY CARD

Core claim. Knowledge is the most strategically significant resource of the firm. Firms exist because they integrate specialized knowledge held by individuals more efficiently than markets can, and their advantage depends on how well they create, integrate, and transfer knowledge.
Origins. Strategy and organization theory, Kogut and Zander (1992) on combinative capabilities, R. M. Grant (1996) on knowledge integration, Nonaka (1994) on knowledge creation, drawing on Polanyi (1966) on tacit knowledge and Penrose (1959).
Level of analysis. Firm, team, individual knowledge holders.
Key constructs. Tacit vs. explicit knowledge, knowledge integration, combinative capabilities, codification, replication, stickiness.
Boundary conditions. Most applicable where production requires integrating specialized, partly tacit knowledge. Less distinctive where knowledge is fully codified and traded in markets.
Related theories. Resource-based view, Absorptive capacity (Section 5.2), Exploration and exploitation (Section 5.3), Transaction cost economics (Section 6.3).

4.3.1 Core logic

The knowledge-based view (KBV) begins with a distinction drawn by Polanyi (1966): much of what people know is **tacit**, it can be demonstrated but not fully articulated, while some knowledge is **explicit** and can be codified in documents, formulas, and software. Explicit knowledge is easy to transfer but also easy to leak. Tacit knowledge is hard to transfer within the firm and correspondingly hard for rivals to acquire.

Kogut and Zander (1992) argue that firms are social communities that specialize in the creation and internal transfer of knowledge. Firms do better than markets not because of opportunism (the transaction-cost explanation) but because shared language, identity, and routines allow members to communicate and recombine knowledge efficiently. They call the ability to synthesize and apply current and acquired knowledge a **combinative capability**. They also identify a fundamental trade-off: codifying knowledge makes it easier to **replicate** internally (to open a new plant or enter a new country) but also easier for competitors to **imitate**.

R. M. Grant (1996) builds a theory of the firm on two premises: knowledge is acquired and held by *individuals*, and efficient production requires *integrating* many specialists' knowledge. Because transferring specialized knowledge between people is costly, the firm's role is to integrate it without requiring everyone to learn what everyone else knows. Grant identifies four integration mech-

anisms (Table 4.5). Their efficiency depends on the common knowledge members share, the frequency and variability of the task, and the structure that minimises the need for communication.

Table 4.5. Knowledge integration mechanisms in the knowledge-based view (after Grant, 1996).

Integration mechanism	How it works	Communication burden	Business example
Rules and directives	Specialists' knowledge is converted into standards, procedures, and instructions others can follow	Low	Clinical protocols, safety standards in a plant
Sequencing	Tasks are ordered so each specialist contributes in turn, with minimal interaction	Low to moderate	Stage-gate product development
Routines	Recurring, partly tacit patterns of interaction coordinate many people without explicit communication	Moderate (built through repetition)	A pit crew, a restaurant kitchen at peak hours
Group problem solving and decision making	Specialists interact face to face to address novel, complex, or high-stakes problems	High	A cross-functional launch team, an incident response war room

Nonaka (1994) complements this account with a theory of knowledge *creation* based on conversion between tacit and explicit forms: socialization (tacit to tacit), externalization (tacit to explicit), combination (explicit to explicit), and internalization (explicit to tacit). Organizational knowledge grows as this cycle spirals from individuals to groups to the organization.

Knowledge transfer within firms is harder than it looks. Szulanski (1996) studied the transfer of best practices between units of the same company and found that the main barriers were not motivational but knowledge-related: the recipient's lack of **absorptive capacity**, **causal ambiguity** about why the practice works, and an arduous relationship between source and recipient. He called this **internal stickiness**. The same features that protect knowledge from imitators therefore impede its replication inside the firm: the KBV restatement of the double edge of causal ambiguity.

4.3.2 Core competences

Writing for managers, Prahalad and Hamel (1990) argued that diversified corporations should be understood not as portfolios of businesses but as portfolios of **core competences**: the collective learning of the organization, especially how to coordinate diverse production skills and integrate multiple streams of technology. They used the image of a tree: core competences are the roots, **core products** are the trunk, business units are the branches, and end products are the fruit. Honda's competence in engines and power trains, they noted, supported businesses from cars to motorcycles to lawn mowers. Canon's competences in optics, imaging, and microprocessor controls supported cameras, copiers, and printers. A competence is *core* if it (a) provides potential access to a wide variety of markets, (b) makes a significant contribution to the benefits customers perceive in the end product, and (c) is difficult for competitors to imitate. The concept anticipated the KBV's emphasis on integration and warned that outsourcing could hollow out the competences on which future products depend. It also has a dark side: competences can harden into **core rigidities** when markets change (Leonard-Barton 1992), a theme taken up in Chapter 5.

BUSINESS CASE

Zara's knowledge loop. Inditex's Zara is often analysed as a supply-chain success, but it is equally a case of knowledge integration. Store managers observe what customers ask for, try on, and reject. That information flows to designers and commercial teams at headquarters, who translate it into new designs. A large share of production is sourced close to Spain, which allows small batches to be made and shipped quickly through a centralized distribution system. The result is a fashion offer that is replenished frequently and adapts within a season. In KBV terms, Zara integrates tacit market knowledge (what is selling and why) with design and operations knowledge through routines and co-location rather than through long forecasting cycles. The advantage is hard to imitate because it depends on the *combination* of proximity sourcing, vertical integration, information flow, and decision rights: a rival that copies any one element without the others gains little.

COMMON PITFALL

Equating knowledge with information systems. Many applications of the KBV measure "knowledge management" by investment in databases or intranets. The theory's argument is the opposite: the knowledge that sustains advantage is largely tacit, socially embedded, and integrated through routines and interaction. Codified repositories are valuable but are also the most imitable form of knowledge. Research designs should capture integration mechanisms and tacitness, not just IT spending.

4.4 Resource-Advantage Theory

THEORY CARD

Core claim. Competition is a constant, disequilibrium-provoking struggle among firms for **comparative advantages in resources** that yield **marketplace positions of competitive advantage** for some market segments and thereby superior financial performance.
Origins. Marketing. Hunt and Morgan (1995), Hunt and Morgan (1996), and the book-length treatment in Hunt (2000, see also Hunt 2007). The theory integrates resource-based strategy with evolutionary and Austrian economics and with heterogeneous-demand theory in marketing.
Level of analysis. Firm and market segment, the economy as an evolving system.
Key constructs. Comparative advantage in resources, relative resource-produced value, relative resource costs, marketplace position, superior financial performance, efficiency and effectiveness advantage.
Boundary conditions. Assumes heterogeneous demand, imperfect information for firms and consumers, and resources that are heterogeneous and imperfectly mobile. Less informative in markets that approximate perfect competition.
Related theories. Resource-based view, Market-based assets, Dynamic capabilities (Section 5.1).

4.4.1 Core logic

Hunt and Morgan (1995) introduced the theory (initially as the "comparative advantage theory of competition") as an alternative to neoclassical perfect competition, which assumes homogeneous demand, perfect information, and homogeneous, perfectly mobile resources. Resource-advantage (R-A) theory replaces each of these with its opposite. Demand is heterogeneous both across and within industries. Information is imperfect and costly. Firm resources, financial, physical, legal, human, organizational, informational, and relational, are heterogeneous and imperfectly mobile, and the firm's objective is **superior** financial performance relative to referents such as rivals or its own past, rather than "maximum" performance that no firm could identify (Hunt 2000).

The causal chain runs from resources to market position to performance. A firm has a **comparative advantage in resources** when its resources allow it to produce a market offering that some segment perceives as having superior value, or to produce it at lower cost. This translates into a **marketplace position**, which in turn produces financial performance. Superior performance provides feedback and the means to invest in further resources. Inferior performance signals rivals' advantages and motivates them to imitate, substitute, or **innovate**, which is why R-A competition is inherently dynamic and why innovation is endogenous to the theory (Hunt and Morgan 1996).

4.4.2 The marketplace position matrix

The theory's signature tool is a three-by-three matrix that crosses **relative resource-produced value** (lower, parity, superior) with **relative resource costs** (lower, parity, higher). Figure 4.4 redraws it. Three cells yield competitive advantage, corresponding to the three forms of advantage that Hunt (2007) distinguishes: an **efficiency advantage** (parity value at lower cost), an **effectiveness advantage** (superior value at parity cost), and an **efficiency-effectiveness advantage** (superior value at lower cost). Two cells are **indeterminate**: superior value at higher cost, or lower value at lower cost, may produce advantage or disadvantage depending on how the segment trades off value against price.

R-A theory makes several distinctive predictions. Because demand is heterogeneous, a firm can hold an advantage in one segment and a disadvantage in another with the same resources. The unit of analysis for advantage is therefore the *firm-segment* pair, not the firm. Because firms seek superior rather than maximum performance and learn through competition, the theory expects **proactive innovation** (firms seek new resources before being forced to) and **reactive innovation** (laggards innovate to neutralize rivals' advantages). And because competition continually upsets positions, sustained advantage is the exception rather than the rule: a view closer to Schumpeter's creative destruction than to equilibrium economics.

DEBATE AND EVIDENCE

Theory of competition or relabelled RBV? Supporters see R-A theory as a general theory of competition that explains why market economies are productive and dynamic, and that integrates marketing's concern with segments and value with strategy's concern with resources (Hunt 2000). Critics have questioned whether its foundational premises add predictive content beyond the RBV and whether the matrix, which classifies positions *ex post*, can be operationalized *ex ante*. The theory's strongest contribution is arguably conceptual: it makes explicit that advantage is defined relative to a segment and to competitors on **both** value and cost, which the VRIO framework leaves implicit.

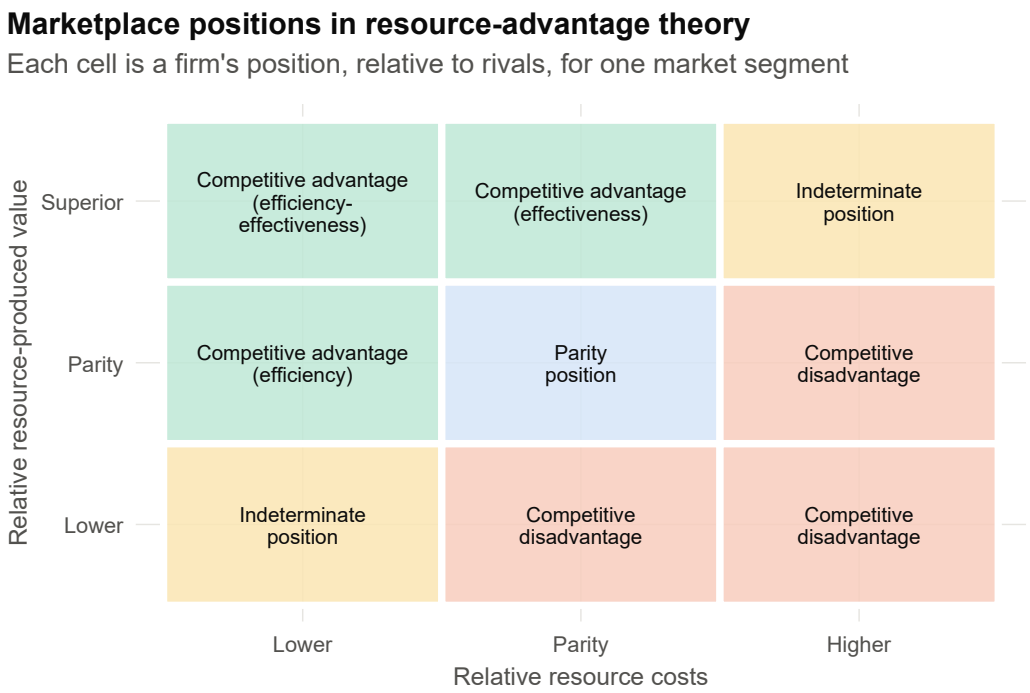


Figure 4.4. The resource-advantage marketplace position matrix. A firm's position for a given market segment depends on the value its resources produce and the cost of producing it, both relative to rivals. Adapted from Hunt and Morgan (1995). Cell labels for the three advantage types follow Hunt (2007).

BUSINESS CASE

Illustrative scenario: one firm, three positions. A regional grocery chain has superior fresh-food sourcing relationships and a cost structure slightly above national discounters. For *quality-focused urban shoppers* it delivers superior value at higher cost: an indeterminate cell that becomes an advantage because this segment pays a premium. For *price-focused family shoppers* the same resources produce parity value at higher cost: a disadvantage. For *convenience shoppers* near its compact stores it offers superior value (location) at parity cost: an effectiveness advantage. The R-A lens tells managers to evaluate resources segment by segment and to direct investment toward segments where the matrix position is favourable.

4.5 Market-Based Assets

THEORY CARD

Core claim. Customer, channel, and partner relationships and market knowledge are **market-based assets** that create shareholder value by accelerating, enhancing, and stabilizing cash flows and by raising their residual value.

Origins. Marketing strategy and finance, Srivastava, Shervani, and Fahey (1998), extended to market-based processes and the RBV in Srivastava, Fahey, and Christensen (2001), related to marketing capabilities in Day (1994).

Level of analysis. Firm (and the firm's relationships with customers, channels, and partners).

Key constructs. Relational market-based assets, intellectual market-based assets, market-based processes, cash-flow acceleration, enhancement, volatility and vulnerability reduction, residual value, shareholder value.
Boundary conditions. Most relevant where customer relationships are durable and repeat purchase or network effects (Section 17.6) matter. Harder to apply where relationships are transactional.
Related theories. Resource-based view, Resource-advantage theory, Signaling theory (Section 8.2), Commitment-trust theory (Section 9.6).

4.5.1 Core logic

Marketing has long struggled to show its value in the language of finance. Srivastava, Shervani, and Fahey (1998) addressed the problem by treating marketing outcomes as **assets** that sit outside the balance sheet but drive the cash flows investors value. They distinguish two types. **Relational market-based assets** are the outcomes of the relationships between a firm and its external stakeholders, customers, channels, and strategic partners, including brand equity, customer loyalty, and installed bases. **Intellectual market-based assets** are the knowledge a firm possesses about its environment: customers' needs, competitors' moves, channel economics. Both are typically intangible, are accumulated through investment over time, and are hard to imitate because they are embedded in relationships, which places them squarely within the RBV (Kozlenkova, Samaha, and Palmatier 2013, Srivastava, Fahey, and Christensen 2001).

Market-based assets create value through four routes that map onto the drivers of discounted cash flow (Figure 4.5): they **accelerate** cash flows (for example, faster adoption of new products among loyal customers), **enhance** cash flows (price premiums, lower selling and service costs, higher share of wallet), **reduce the volatility and vulnerability** of cash flows (loyal customers are less sensitive to competitive moves), and increase the **residual value** of the business (a strong customer base persists beyond the forecast horizon).

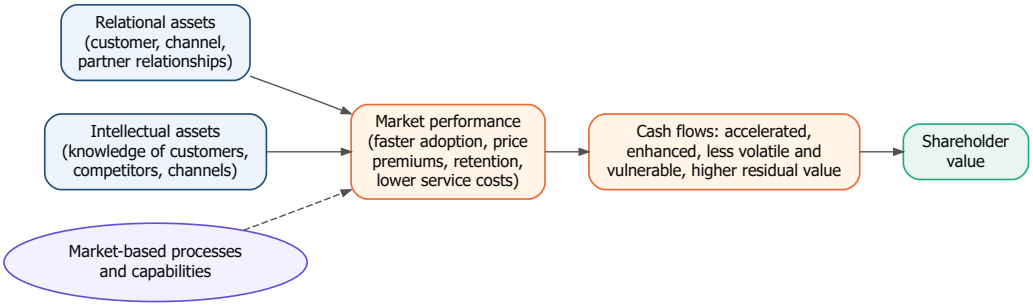


Figure 4.5. How market-based assets create shareholder value: relational and intellectual assets improve market performance, which accelerates and enhances cash flows, reduces their volatility and vulnerability, and raises residual value. Adapted from the framework in Srivastava, Shervani, and Fahey (1998).

Day (1994) provides the capability counterpart. Market-driven organizations excel at **outside-in** capabilities, **market sensing** and **customer linking**, that connect internal processes to the market, complemented by **inside-out** capabilities (such as manufacturing and logistics) and **spanning** capabilities (such as pricing and new product development) that integrate the two. These capabilities are the processes through which market-based assets are built and leveraged. Table 4.6

summarises the main asset types.

Table 4.6. Market-based assets and their links to cash flows (classification follows Srivastava, Shervani, & Fahey, 1998. Indicators are illustrative).

Type	Asset	Illustrative indicators	Primary cash-flow effect
Relational	Customer relationships and brand equity	Retention, share of wallet, price premium, brand-equity indices	Enhance cash flows, reduce volatility
Relational	Installed base and user network	Active users, attach rates of complements	Accelerate adoption of new offerings
Relational	Channel relationships	Shelf space, distributor tenure, co-marketing commitments	Accelerate and enhance cash flows
Relational	Partner and ecosystem relationships	Alliances, co-branding, developer participation	Enhance cash flows, raise residual value
Intellectual	Customer and market knowledge	Segment-level insight, data assets, research capability	Accelerate cash flows via better-targeted launches
Intellectual	Competitor and channel knowledge	Intelligence systems, channel economics models	Reduce vulnerability to competitive moves

BUSINESS CASE

Disney's brand and character franchises. Disney is a canonical example of market-based assets working as a system. Its characters and stories are protected by intellectual property law, but their economic value comes from relationships with audiences built over generations. Those relationships let Disney **accelerate** cash flows, a new film reaches a large, pre-committed audience, and **enhance** them by extending each franchise across theme parks, consumer products, and media. Its acquisitions of Pixar (2006), Marvel (2009), and Lucasfilm (2012) added further franchises that could be exploited through the same cross-business system, and its direct-to-consumer streaming service, launched in 2019, turned audience relationships into a subscription relationship. In VRIO terms, the brand-and-IP portfolio is valuable, rare, and protected by both law and history. In market-based-asset terms, its value lies in the durability and leverage of audience relationships.

RESEARCH TOOLKIT

Sample hypotheses for market-based assets.

- *H1.* Increases in customer satisfaction or brand equity will be associated with lower volatility of future cash flows, controlling for current profitability.
- *H2.* Firms with larger installed bases will achieve faster adoption of new product generations than rivals with comparable advertising spending.
- *H3.* The effect of marketing capabilities on profit growth will be stronger when they are combined with complementary R&D or operations capabilities (N. A. Morgan, Slotegraaf, and Vorhies 2009).

Designs. Event studies and long-horizon stock-return models. Panel models of cash-flow level and volatility. Customer-lifetime-value models linking relational assets to firm value.

4.6 Chapter Summary

KEY TAKEAWAYS

- The **resource-based view** explains persistent performance differences within industries by resource **heterogeneity** and **immobility**. Advantage is sustained when valuable, rare resources are costly to imitate or substitute (Jay Barney 1991, M. A. Peteraf 1993).
- Strategic assets are **accumulated** rather than bought. Time compression diseconomies, asset mass efficiencies, and causal ambiguity protect incumbents (Dierickx and Cool 1989, J. B. Barney 1986).
- **VRIO** turns the theory into a diagnostic. The "O" reminds managers that resources create value only when structures, controls, and incentives allow them to be exploited (J. B. Barney and Hesterly 2012).
- Advantage often lies in **bundles** of complementary resources and in how they are **orchestrated**, not in single assets (P. Milgrom and Roberts 1995, Sirmon, Hitt, and Ireland 2007).
- The **knowledge-based view** treats knowledge integration as the reason firms exist. Tacitness protects knowledge from rivals but also makes it sticky inside the firm (R. M. Grant 1996, Kogut and Zander 1992, Szulanski 1996).
- **Resource-advantage theory** embeds resources in a dynamic, segment-specific theory of competition with efficiency and effectiveness advantages (Hunt and Morgan 1995, Hunt 2007).
- **Market-based assets** connect customer, channel, and partner relationships to shareholder value through the level, timing, risk, and residual value of cash flows (Srivastava, Shervani, and Fahey 1998).
- The empirical record supports the resource-performance link, especially for resources that meet the theory's criteria, but measurement and tautology remain live concerns (Crook et al. 2008, Newbert 2007, Priem and Butler 2001).

Table 4.7 presents the comparison of the chapter's theories.

Table 4.7. Comparison of the chapter's theories.

Theory	Unit of analysis	Source of advantage	Key mechanism	Signature prediction	Typical business use
Resource-based view	Firm	Heterogeneous, immobile resources	Isolating mechanisms protect rents	Within-industry performance differences persist	Strategy formulation, M&A screening
VRIO	Resource	Valuable, rare, inimitable, organized resources	Sequential tests of value and protection	Only VRIO resources sustain advantage	Capability audits, investment prioritization

Table 4.7. Comparison of the chapter's theories. *(continued)*

Theory	Unit of analysis	Source of advantage	Key mechanism	Signature prediction	Typical business use
Knowledge-based view	Firm / team	Integration of specialized, tacit knowledge	Integration mechanisms, combinative capabilities	Codification eases replication but invites imitation	Organization design, knowledge transfer
Core competences	Corporation	Collective, cross-business skills	Competences feed core products across businesses	Competence-based diversification outperforms	Portfolio and outsourcing decisions
Resource-advantage theory	Firm-segment	Comparative advantage in resources	Relative value and relative cost	Positions shift through innovation	Segment-level positioning
Market-based assets	Firm-market	Customer, channel, partner relationships, market knowledge	Cash-flow acceleration, enhancement, de-risking	Relational assets raise firm value	Justifying marketing investment to finance

4.7 Discussion Questions

1. Toyota allowed competitors to tour its plants. Using the concepts of causal ambiguity and social complexity, explain why this was not as risky as it appears. Under what conditions would such openness be a strategic mistake?
2. Priem and Butler argue that the RBV is tautological. Design a study that tests an RBV hypothesis while measuring resource value *independently* of performance. What would falsify your hypothesis?
3. Choose an industry undergoing technological change (for example, retail banking or automotive). Which resources have shifted from valuable to non-valuable, and which have become newly valuable? What does this imply for the "sustained" in sustained competitive advantage?
4. Kogut and Zander describe a trade-off between replication and imitation. How should a multinational franchise business decide how much of its operating know-how to codify?
5. Using the resource-advantage matrix, analyse a firm that competes in at least two segments. Is its position the same in each? How should that affect resource investment?
6. A chief marketing officer must justify a multi-year investment in customer relationships to a sceptical chief financial officer. Using the market-based assets framework, which cash-flow effects would you measure, and how?

4.8 Further Reading

- Jay Barney (1991): the foundational statement of the conditions for sustained competitive advantage. Still the most cited entry point.
- Dierickx and Cool (1989): a short, elegant paper on why strategic assets must be accumulated, introducing time compression diseconomies and asset mass efficiencies.
- M. A. Peteraf (1993): translates the RBV into the economics of rents through four cornerstones.
- Priem and Butler (2001) and J. B. Barney (2001): read together as the defining debate about the theory's usefulness and testability.
- Crook et al. (2008): the meta-analytic evidence on strategic resources and performance.
- R. M. Grant (1996): the clearest statement of the knowledge-based theory of the firm.
- Kozlenkova, Samaha, and Palmatier (2013): a review of resource-based theory in marketing, with guidance on operationalizing VRIO attributes.
- Srivastava, Shervani, and Fahey (1998): the market-based assets framework that links marketing to shareholder value.

Chapter 5

Dynamic Capabilities, Learning, and Renewal

In 1975 an engineer at Eastman Kodak, Steven Sasson, built one of the first digital cameras. Kodak had world-class chemistry, an iconic brand, global distribution, and deep research budgets: by any resource-based standard, a formidable collection of valuable, rare, and hard-to-imitate assets. In January 2012 the company filed for Chapter 11 bankruptcy protection. Its Japanese rival Fujifilm faced the same collapse in demand for photographic film, yet emerged as a diversified company with businesses in medical systems, pharmaceuticals, cosmetics, and advanced materials. Netflix, meanwhile, cannibalized its own DVD-by-mail business to build a streaming service, then reinvented itself again as a producer of original content.

These cases expose a limitation of the theories in Chapter 4. Resource-based logic explains why a firm with superior resources outperforms its rivals *at a point in time*. It says much less about how firms **renew** their resources when technologies, customers, and competitors change, and why the very resources that made a firm successful can become the reason it fails. This chapter develops the theories that address renewal. We begin with the **dynamic capabilities framework** and its main variants, then examine **absorptive capacity** (how firms learn from outside), the tension between **exploration and exploitation** (how firms allocate attention between the new and the known), and **organizational ambidexterity** (how firms try to do both). We close each section with business cases drawn from the public record: Netflix, Kodak and Fujifilm, Microsoft, and IBM (Table 5.1).

Table 5.1. Chapter at a glance: theories of capability renewal and organizational learning.

Theory	Core question	Level of analysis	Key source
Dynamic capabilities	How do firms integrate, build, and reconfigure competences as environments change?	Firm	Teece, Pisano, & Shuen (1997), Teece (2007)
Dynamic capabilities as best practices	Are dynamic capabilities idiosyncratic or common across firms?	Firm / process	Eisenhardt & Martin (2000)
Capability hierarchy and learning	Where do dynamic capabilities come from?	Firm / routine	Winter (2003), Zollo & Winter (2002)
Core rigidities	Why do strengths become weaknesses?	Firm / project	Leonard-Barton (1992)
Absorptive capacity	How do firms recognize, assimilate, and apply external knowledge?	Firm / unit	Cohen & Levinthal (1990), Zahra & George (2002)

Table 5.1. Chapter at a glance: theories of capability renewal and organizational learning. *(continued)*

Theory	Core question	Level of analysis	Key source
Exploration and exploitation	How should organizations allocate effort between new possibilities and known certainties?	Organization / system	March (1991), Levinthal & March (1993)
Ambidexterity	How can a firm explore and exploit at the same time?	Firm / business unit	Tushman & O'Reilly (1996), Gibson & Birkinshaw (2004)

5.1 Dynamic Capabilities Framework

THEORY CARD

Core claim. In environments of rapid technological and market change, sustained advantage depends less on the resources a firm owns than on its **dynamic capabilities**: its ability to integrate, build, and reconfigure internal and external competences to address changing environments.

Origins. Strategic management, building on evolutionary economics (R. R. Nelson and Winter 1982) and the resource-based view (Section 4.1), Teece, Pisano, and Shuen (1997), elaborated in Teece (2007) and Helfat et al. (2007), contrasting interpretation in Eisenhardt and Martin (2000).

Level of analysis. Firm. Also specific organizational processes (product development, acquisition integration, alliance management) and top management.

Key constructs. Ordinary (operational) capabilities, dynamic capabilities, processes, positions, and paths, sensing, seizing, and transforming, replication, imitation, path dependence, core rigidities.

Boundary conditions. Most consequential when environments change rapidly or discontinuously. In stable environments, investment in dynamic capabilities may cost more than it returns (Winter 2003). The form of effective dynamic capabilities differs between moderately dynamic and high-velocity markets (Eisenhardt and Martin 2000).

Related theories. Resource-based view (Section 4.1), Knowledge-based view (Section 4.3), Absorptive capacity, Exploration and exploitation, Ambidexterity, Diffusion of innovations (Section 10.1).

5.1.1 Origins and intellectual history

The framework's deepest root is the evolutionary theory of R. R. Nelson and Winter (1982), which replaced the neoclassical firm, a production function that maximizes profit, with a firm that operates through **routines**: regular, predictable patterns of activity that function as organizational memory and are the organizational analogue of genes. Routines are efficient but sticky. Firms change mainly by searching locally around what they already do, and markets select among firms with different routines. From this perspective, a firm's capabilities are not chosen at will. They are inherited from its history and modified incrementally.

Teece, Pisano, and Shuen (1997) combined this evolutionary view with resource-based logic. They observed that the "winners" in global markets were firms that could demonstrate "timely responsiveness and rapid and flexible product innovation, coupled with the management capability to

effectively coordinate and redeploy internal and external competences" (Teece, Pisano, and Shuen 1997, 515). Accumulating valuable assets is not enough. Firms must also be able to change them. The term **dynamic** refers to the capacity to renew competences so as to achieve congruence with a changing business environment. **Capabilities** emphasises the role of strategic management in adapting, integrating, and reconfiguring skills, resources, and functional competences. Their formal definition has become the standard one:

> **DEFINITION**
>
> **Definition 5.1. Dynamic capabilities** are "the firm's ability to integrate, build, and reconfigure internal and external competences to address rapidly changing environments" (Teece, Pisano, and Shuen 1997, 516).

Teece and colleagues distinguished several layers of firm-specific assets, from undifferentiated factors of production to the dynamic capabilities that change them (Table 5.2). The hierarchy foreshadows later work on the distinction between ordinary and dynamic capabilities (Winter 2003, Teece 2014).

Table 5.2. Layers of firm-specific assets in the dynamic capabilities framework (definitions paraphrased from Teece, Pisano, & Shuen, 1997, p. 516).

Term	Meaning in Teece, Pisano, & Shuen (1997)	Business example
Factors of production	Undifferentiated inputs available in disaggregated form in factor markets	Commodity land, generic labour, standard machinery
Resources	Firm-specific assets that are difficult if not impossible to imitate	Trade secrets, specialized production facilities, engineering experience
Organizational routines / competences	Firm-specific assets assembled in integrated clusters spanning individuals and groups so that they enable distinctive activities to be performed	Quality management, miniaturization, systems integration
Core competences	The competences that define a firm's fundamental business	Optics and imaging for a camera and copier maker
Dynamic capabilities	The ability to integrate, build, and reconfigure internal and external competences in rapidly changing environments	Repeatedly moving into new product generations or business models
Products	Final goods and services produced by deploying the firm's competences	Cameras, copiers, printers

5.1.2 Paradigms of competitive advantage

Teece, Pisano, and Shuen (1997) located their framework among four paradigms of strategy. Two emphasise the exploitation of **market power**. The **competitive forces** approach (Michael E. Porter 1980a) holds that profitability is determined by industry structure, and that firms should position themselves to defend against the five forces (Section 31.4). The **strategic conflict** approach (Shapiro 1989) uses game theory to analyse how firms influence rivals' behaviour through investment, pricing, signalling, and commitment. Two emphasise **efficiency**. The **resource-based perspective** holds that firm-specific assets matter more than industry position, because resources are heterogeneous and "sticky". The **dynamic capabilities** approach adds that, when environments change, the advantage lies in the capacity to renew those assets. Table 5.3 summarises the comparison.

Table 5.3. Four paradigms of competitive advantage (organized following Teece, Pisano, & Shuen, 1997).

Dimension	Competitive forces	Strategic conflict	Resource-based view	Dynamic capabilities
Key source	Porter (1980)	Shapiro (1989)	Wernerfelt (1984), Barney (1991)	Teece, Pisano, & Shuen (1997)
Source of rents	Market power from favourable industry structure	Market power from strategic moves against rivals	Ricardian rents from scarce, firm-specific resources	Schumpeterian rents from renewing resources and competences
Unit of analysis	Industry and firm position	Firm and rivals (game)	Resources and capabilities	Processes, positions, and paths
View of the firm	Largely homogeneous within an industry	Strategic actor with commitments	Bundle of heterogeneous resources	Evolving bundle of routines and competences
Role of management	Choose industry and position	Out-manoeuvre rivals	Acquire and protect valuable resources	Sense, seize, and transform
Time horizon	Static (structure)	Short-run interaction	Comparative static	Dynamic, path-dependent
Typical prescription	Enter attractive industries, build barriers	Use commitment and signalling to deter rivals	Exploit VRIO resources	Invest in learning and reconfiguration

The paradigms are not mutually exclusive. Industry structure constrains what any firm can earn. Game-theoretic interactions shape short-run outcomes. Resources determine which firms can occupy attractive positions, and dynamic capabilities determine whether those firms can move when the structure itself shifts. The dynamic capabilities view is most distinctive when the main threat to profits is not rivalry within a stable industry but **Schumpeterian** change: new technologies and business models that make existing resources obsolete.

5.1.3 Processes, positions, and paths

Teece, Pisano, and Shuen (1997) argued that a firm's distinctive competence and dynamic capabilities are determined by three classes of factors (Table 5.4).

Processes are the way things are done in the firm: its routines and patterns of current practice and learning. They serve three roles. **Coordination and integration** is a static concept: how efficiently the firm coordinates activities inside and outside its boundaries. **Learning** is a dynamic concept: how repetition and experimentation allow tasks to be performed better and faster, including through collaborations and partnerships. **Reconfiguration and transformation** is the capacity to sense the need to change the asset structure and to accomplish the internal and external transformation. Because change is costly, firms need processes that reduce the cost of low-return change.

Positions are the firm's current endowments: technological assets, complementary assets, financial assets, reputational assets, structural assets (formal and informal organization and external linkages), institutional assets (the regulatory and intellectual-property environment), market-structure assets, and organizational boundaries (the degree of vertical, lateral, and horizontal integration). Positions shape what the firm can do next. Complementary assets in particular determine whether an innovator can profit from its innovation (Teece 1986).

Paths are the strategic alternatives available to the firm, which are shaped by **path dependence**, a firm's previous investments and its repertoire of routines constrain its future behaviour, and by **technological opportunities**, which depend on the pace and direction of scientific discovery in-

side and outside the industry. Path dependence also complicates learning: when many parameters change at once, firms find it hard to identify causal relationships, so learning tends to be local.

Table 5.4. Processes, positions, and paths (after Teece, Pisano, & Shuen, 1997).

Factor	Elements	Strategic question	Business illustration
Processes	Coordination and integration, learning, reconfiguration and transformation	How well do we coordinate, learn, and change?	Repeatable product-launch and post-merger integration routines
Positions	Technological, complementary, financial, reputational, structural, institutional, and market assets. Organizational boundaries	What do we hold now, and what does it enable?	An installed customer base and global distribution that a new product can ride on
Paths	Path dependence, technological opportunities	Which futures are open to us given our history?	A film maker's chemistry know-how opening paths into cosmetics and pharmaceuticals

5.1.4 Replicability and imitability

If competences are valuable, why do firms struggle to copy them, including from their own units? Teece, Pisano, and Shuen (1997) distinguish **replication**, transferring or redeploying competences from one economic setting to another within the firm, from **imitation**, replication by competitors. Replication is more than information transfer because much of what a firm knows is tacit. The organizational context in which a routine works is part of the routine. Imitation is harder still, because imitators lack access to that context. The ease of imitation is governed by the **appropriability regime**: the strength of intellectual-property protection and the inherent replicability of the technology. Process technologies hidden inside the factory are usually easier to protect than product technologies that can be reverse-engineered once sold. These arguments mirror the replication-imitation trade-off in the knowledge-based view (Section 4.3) (Kogut and Zander 1992).

5.1.5 Microfoundations: sensing, seizing, and transforming

A decade later, Teece (2007) disaggregated dynamic capabilities into three clusters of activities, each with identifiable **microfoundations**, skills, processes, procedures, organizational structures, decision rules, and disciplines:

- **Sensing** opportunities and threats: scanning, searching, and exploring technologies and markets. Directing internal R&D, tapping developments in science, suppliers, and complementors. Identifying changing customer needs and latent demand.
- **Seizing** opportunities: designing the business model and customer solution. Selecting enterprise boundaries and managing complements. Making timely, sometimes irreversible, investment decisions while avoiding decision biases, and building loyalty and commitment among employees.
- **Transforming** (managing threats and reconfiguring): continually realigning and recombining tangible and intangible assets, decentralizing and maintaining near-decomposable structures, managing cospecialized assets, governance and knowledge management.

Teece (2014) sharpened the distinction between **ordinary capabilities**, which allow a firm to perform current activities efficiently ("doing things right"), and **dynamic capabilities**, which allow it to do the right things at the right time by combining good strategy with the capacity to orchestrate resources. Ordinary capabilities can often be bought or benchmarked. Dynamic capabilities

rest on signature processes and entrepreneurial management that are hard to imitate. Figure 5.1 summarises the logic.

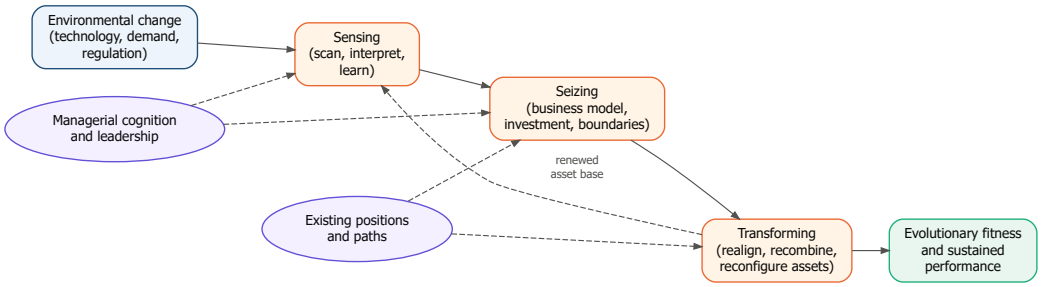


Figure 5.1. Microfoundations of dynamic capabilities. Sensing identifies opportunities and threats, seizing commits resources to address them, and transforming reconfigures the asset base. Managerial cognition and the firm's existing positions shape each step. Based on Teece (2007).

5.1.6 Dynamic capabilities as best practices

Eisenhardt and Martin (2000) offered an influential reinterpretation. They argued that dynamic capabilities are not vague, idiosyncratic abilities but **specific, identifiable processes**: product development, strategic decision making, alliancing, acquisition integration, resource allocation, and knowledge brokering. Because effective versions of these processes share common features across firms, dynamic capabilities exhibit **equifinality** (firms can reach them from different starting points) and substantial **commonality**: they resemble "best practices". It follows that dynamic capabilities are *necessary but not sufficient* for competitive advantage: advantage comes from the **resource configurations** that firms create with them, sooner, more astutely, or more fortuitously than rivals.

The form of effective dynamic capabilities, they argued, depends on **market dynamism**. In **moderately dynamic** markets, where change is frequent but predictable, dynamic capabilities resemble traditional routines: detailed, analytic, stable processes with predictable outcomes that rely on existing knowledge. In **high-velocity** markets, where change is nonlinear and unpredictable, effective dynamic capabilities are **simple rules** and experiential processes that rely on rapidly created new knowledge, prototyping, and iterative execution. Their outcomes are unpredictable, and they are inherently fragile. Learning mechanisms such as repeated practice, codification, learning from small failures, and pacing guide their evolution (Table 5.5).

Table 5.5. Two conversations about dynamic capabilities (see Peteraf, Di Stefano, & Verona, 2013).

Dimension	Teece, Pisano, & Shuen (1997), Teece (2007)	Eisenhardt & Martin (2000)
Nature of dynamic capabilities	Firm-specific, partly tacit, embedded in processes and managerial skill	Specific, identifiable processes with common features across firms
Heterogeneity	Idiosyncratic, hard to imitate	Commonalities across firms (best practices), equifinality
Link to advantage	Can be a direct source of sustained advantage	Necessary but not sufficient, advantage lies in resulting resource configurations
Role of market dynamism	Most valuable in rapidly changing environments	Form differs: detailed routines (moderate) vs. simple rules (high velocity)
Role of managers	Entrepreneurial sensing, seizing, and transforming	Designing and executing effective processes

Table 5.5. Two conversations about dynamic capabilities (see Peteraf, Di Stefano, & Verona, 2013). *(continued)*

Dimension	Teece, Pisano, & Shuen (1997), Teece (2007)	Eisenhardt & Martin (2000)
Sustainability	Durable if microfoundations are hard to replicate	Advantage often temporary in high-velocity markets

DEBATE AND EVIDENCE

Two conversations or one? M. Peteraf, Di Stefano, and Verona (2013) showed through a co-citation analysis that the dynamic capabilities literature had split into two largely separate conversations, anchored by Teece, Pisano, and Shuen (1997) and Eisenhardt and Martin (2000), which disagree about whether dynamic capabilities can underpin *sustained* advantage and under what market conditions. They proposed a contingent reconciliation: the two views can be integrated once boundary conditions such as market dynamism and the level of the capability are made explicit. Critics have also raised a **tautology** concern similar to the one levelled at the RBV: if a dynamic capability is defined by successful change, its link to performance is true by definition. A related **infinite-regress** problem asks whether there are capabilities to change dynamic capabilities, and capabilities to change *those*. Collis (1994) noted that any higher-order capability can be trumped by one at a still higher order. Schilke, Hu, and Helfat (2018), reviewing the empirical literature, found growing evidence that dynamic capabilities are associated with performance, often indirectly through changes in operational capabilities, while urging clearer measurement and stronger causal designs.

5.1.7 The capability hierarchy and learning mechanisms

Winter (2003) clarified the concept by distinguishing **zero-order** (ordinary) capabilities, those that permit a firm to make a living in the short term by producing and selling the same product to the same customers at the same scale, from **first-order** dynamic capabilities that change the product, the production process, the scale, or the customers served, and from **higher-order** capabilities that change the dynamic capabilities themselves. He made two further points that are easy to miss. First, not all change requires a dynamic capability: firms can change through **ad hoc problem solving**, which is non-routine and creative but not patterned. Second, dynamic capabilities are **costly**, they require long-term commitments to specialized resources and people, so whether they pay off depends on how often and how predictably the environment changes. A firm in a stable market may rationally rely on ad hoc change.

Where do dynamic capabilities come from? Zollo and Winter (2002) defined a dynamic capability as a learned and stable pattern of collective activity through which an organization systematically generates and modifies its operating routines in pursuit of improved effectiveness. They traced its origins to three **learning mechanisms** of increasing deliberateness and cost:

1. **Experience accumulation**: tacit learning by doing, as routines are repeated.
2. **Knowledge articulation**: making knowledge explicit through collective discussion, debriefings, and performance reviews.
3. **Knowledge codification**: writing knowledge down in manuals, decision tools, checklists, and software.

The more deliberate mechanisms are especially valuable when the relevant tasks are **infrequent**,

heterogeneous, and **causally ambiguous**: conditions under which experience alone teaches poorly. Post-acquisition integration is their example: few firms do enough acquisitions for experience alone to be a reliable teacher, and each deal differs from the last. Figure 5.2 combines Winter's hierarchy with Zollo and Winter's learning mechanisms.

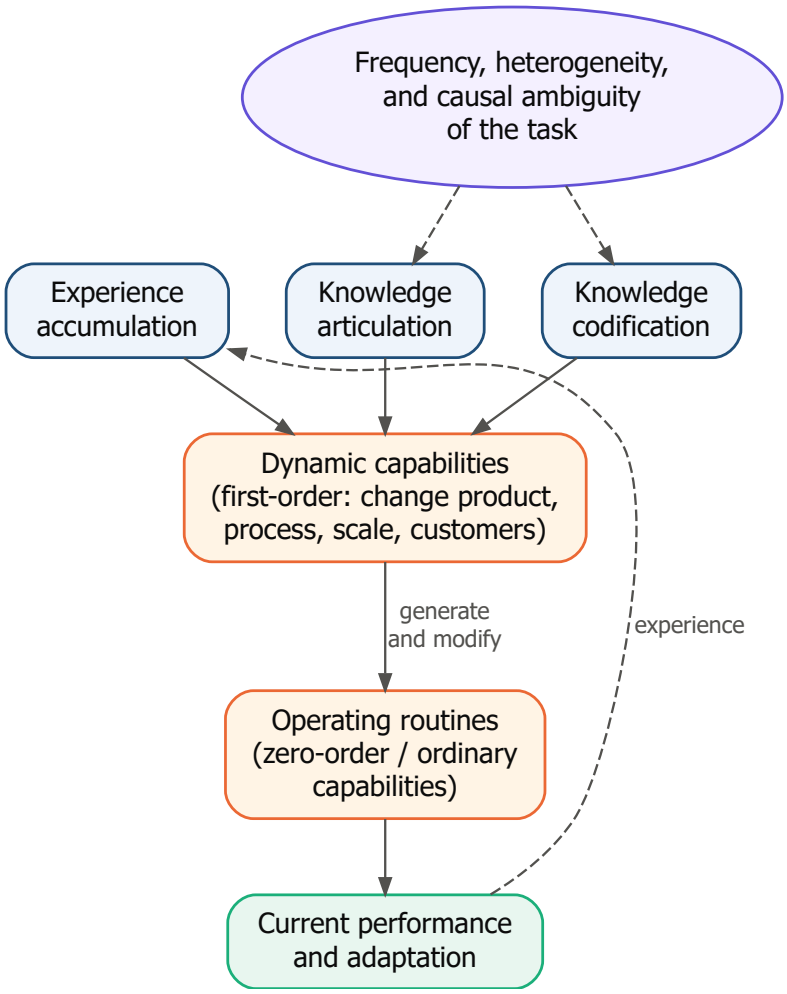


Figure 5.2. The capability hierarchy. Learning mechanisms generate and modify dynamic capabilities, which in turn modify the operating routines (ordinary capabilities) that produce current performance. Based on Winter (2003) and Zollo and Winter (2002).

Helfat and Peteraf (2003) added a life-cycle perspective. Capabilities are founded, developed, and reach maturity. When environments change, they may branch into **retirement**, **retrenchment**, **renewal**, **replication** (in another market), **redeployment** (to a related market), or **recombination** (with other capabilities). The capability life cycle explains why a firm's capability portfolio has a history and why the same capability can be an asset in one era and a burden in the next.

5.1.8 Core rigidities

Leonard-Barton (1992) identified the paradox at the heart of capability renewal. In a field study of new product development projects, she found that the **core capabilities** that distinguish a firm,

rooted in employee knowledge and skills, technical systems, managerial systems, and values and norms, simultaneously enhance projects that fit them and hinder projects that do not. The same capabilities become **core rigidities** when the firm needs to move in a new direction: skills that were celebrated become narrow, systems that were efficient become constraints, and values that motivated people become filters that screen out unfamiliar ideas.

Research on technological change reinforces the point. Discontinuities that are **competence-destroying**, rendering existing skills obsolete, tend to be introduced by new entrants and to hurt incumbents, whereas **competence-enhancing** discontinuities tend to favour incumbents (Tushman and Anderson 1986). Even seemingly modest **architectural innovations**, which change how known components are linked, can defeat incumbents whose communication channels and information filters are organized around the old architecture (R. M. Henderson and Clark 1990). And capability is not the only constraint: in their study of Polaroid's move into digital imaging, Tripsas and Gavetti (2000) showed that the firm developed substantial technical capability but that managerial beliefs about the business model, rooted in its instant-film heritage, shaped how it deployed that capability. **Cognition** as well as capability determines whether a firm can change.

BUSINESS CASE

Kodak and Fujifilm. Both firms faced the same competence-destroying discontinuity as digital imaging displaced chemical film. Kodak was not ignorant of the technology, its own engineer had built an early digital camera in 1975, and it introduced digital cameras and services. But its economics, organization, and identity were built around film, processing, and paper, and digital products offered nothing comparable to that business. The company filed for Chapter 11 protection in 2012 and emerged as a much smaller firm. Fujifilm, under long-serving chief executive Shigetaka Komori, restructured its film operations and asked which of its underlying competences, chemistry, thin films, coatings, precision manufacturing, and image processing, could be redeployed. It moved into medical imaging and pharmaceuticals (including the acquisition of Toyama Chemical in 2008), launched a cosmetics line built on its expertise in collagen and antioxidants, and supplied optical films for flat-panel displays. In Helfat and Peteraf's terms, Fujifilm **redeployed** and **recombined** capabilities. In Teece's terms, it sensed the threat and transformed its asset base, rather than defending a core that had become a rigidity.

5.1.9 Applications in business

BUSINESS CASE

Netflix: renewing the business model twice. Netflix was founded in 1997 as a DVD-by-mail rental service and built ordinary capabilities in logistics, subscription billing, and recommendation algorithms. In 2007 it launched streaming, a technology that would eventually cannibalize its own core. In 2011 it announced a plan to split the DVD business into a separate service called Qwikster, provoking a customer backlash. The plan was reversed within weeks: a reminder that seizing is risky and that reconfiguration has costs. Netflix then moved into original programming, beginning with *House of Cards* in 2013, which required capabilities in content development, talent relations, and global production that a distributor had not needed. It expanded into most countries in 2016 and closed its DVD business in 2023. Read through Teece's framework, Netflix sensed shifts in broadband and consumer behaviour, seized them

by committing to streaming and original content before they were profitable, and transformed its asset base from warehouses and discs to software, data, and intellectual property. Its rival Blockbuster, which filed for bankruptcy in 2010, illustrates the alternative path.

BUSINESS CASE

Microsoft's turn to the cloud. When Satya Nadella became chief executive in 2014, Microsoft's profits were anchored in Windows and Office licences, and it had struggled in mobile devices and search. Under the banner "mobile-first, cloud-first", the company shifted investment toward its Azure cloud platform and subscription software (Office 365), wrote down its phone business, made its software available on rival platforms, and embraced open-source technologies it had once opposed, later acquiring LinkedIn (2016) and GitHub (2018). Nadella emphasised that the strategic shift depended on a cultural one, built around a "growth mindset" (Nadella, Shaw, and Nichols 2017). In dynamic capabilities terms, Microsoft did not abandon its positions, enterprise relationships, developer ecosystems, and software engineering, but **recombined** them with new cloud infrastructure. The transformation of managerial systems and values, Leonard-Barton's deepest layers of capability, was as important as the technology.

5.1.10 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms with more deliberate learning mechanisms (articulation and codification) for acquisition integration will achieve better post-acquisition performance, especially when acquisitions are infrequent and heterogeneous (Zollo and Winter 2002).
- *H2.* The effect of dynamic capabilities on firm performance will be mediated by changes in operational capabilities (Schilke, Hu, and Helfat 2018).
- *H3.* Detailed, codified product development routines will improve performance in moderately dynamic markets but not in high-velocity markets, where simple rules will be more effective (Eisenhardt and Martin 2000).

Designs. Longitudinal case studies of renewal, panel studies of firms facing exogenous shocks (deregulation, technological discontinuities), event studies of reconfiguration announcements, simulation models of routine evolution, survey scales of sensing, seizing, and transforming validated against behavioural indicators.

Measures. Frequency and speed of product-line renewal. Acquisition and alliance experience and the presence of integration manuals or teams. R&D reallocation across technology classes. Top-management attention to emerging issues (text analysis of letters to shareholders or earnings calls).

Data sources. Patent and product-launch databases, SDC for acquisitions and alliances, annual reports and earnings-call transcripts, industry case archives.

COMMON PITFALL

Defining capability by outcome. It is tempting to infer that a firm that adapted successfully "had" dynamic capabilities and one that failed did not. This is circular. Measure the capability

before the change, the presence of sensing processes, investment rules, learning mechanisms, and test whether it predicts adaptation, ideally among firms facing the same shock. Remember also Winter's point that change can happen without a dynamic capability, through ad hoc problem solving or luck.

5.2 Absorptive Capacity

THEORY CARD

Core claim. A firm's ability to recognize the value of new external knowledge, assimilate it, and apply it to commercial ends, its **absorptive capacity**, is largely a function of its prior related knowledge, and therefore develops cumulatively and path-dependently.
Origins. Economics of innovation and organization theory, W. M. Cohen and Levinthal (1990), reconceptualized by Zahra and George (2002).
Level of analysis. Firm and business unit. Also individuals and teams who act as gatekeepers and boundary spanners.
Key constructs. Prior related knowledge, acquisition, assimilation, transformation, exploitation, potential vs. realized absorptive capacity, activation triggers, social integration mechanisms, regimes of appropriability.
Boundary conditions. Most important where valuable knowledge is generated outside the firm (by universities, suppliers, customers, or rivals) and is partly tacit or complex.
Related theories. Knowledge-based view (Section 4.3), Dynamic capabilities, Exploration and exploitation, Diffusion of innovations (Section 10.1).

5.2.1 Core logic

W. M. Cohen and Levinthal (1990) began from an observation about learning in individuals: the ability to learn new things depends on what one already knows, because new knowledge is understood by relating it to existing knowledge structures. Organizations are no different. A firm's **absorptive capacity** is its ability to recognize the value of new, external information, assimilate it, and apply it to commercial ends. Because it depends on prior related knowledge, absorptive capacity is **cumulative**, investing in it today makes tomorrow's learning easier, and **path-dependent**. A firm that stops investing in a field may find itself **locked out**: it can no longer recognize the significance of developments it once could have exploited.

Two implications have been especially influential. First, a firm's own R&D does more than produce inventions. It also builds the capacity to understand and use other organizations' inventions. Internal research and external sourcing are therefore complements rather than substitutes. Second, organizational absorptive capacity is not simply the sum of individuals' knowledge. It depends on **gatekeepers** and **boundary spanners** who translate external knowledge for internal use, and on internal communication structures that move knowledge to where it can be applied. Some overlap of knowledge across units helps communication, while diversity of backgrounds increases the range of external developments the firm can recognize.

5.2.2 Potential and realized absorptive capacity

Zahra and George (2002) reconceptualized absorptive capacity as a dynamic capability with four dimensions grouped into two components. **Potential absorptive capacity (PACAP)** comprises **acquisition** (identifying and acquiring external knowledge) and **assimilation** (analysing, interpreting, and understanding it). **Realized absorptive capacity (RACAP)** comprises **transformation** (combining new and existing knowledge) and **exploitation** (applying the result to create new or improved products, processes, and competences). The distinction explains why firms that are good at scanning the environment do not necessarily profit from what they find: PACAP without RACAP generates costs without returns, while RACAP without PACAP can produce short-term gains but leaves the firm exposed to competence traps. The ratio of realized to potential capacity is an **efficiency factor** that varies across firms.

The model also specifies contingencies (Figure 5.3). **Activation triggers**, crises, technological shifts, regulatory changes, intensify the search for external knowledge. **Social integration mechanisms**, formal and informal, reduce the gap between potential and realized capacity by spreading knowledge within the firm. **Regimes of appropriability**, the strength of intellectual-property protection and the inimitability of the resulting knowledge, determine whether realized capacity translates into sustained advantage.

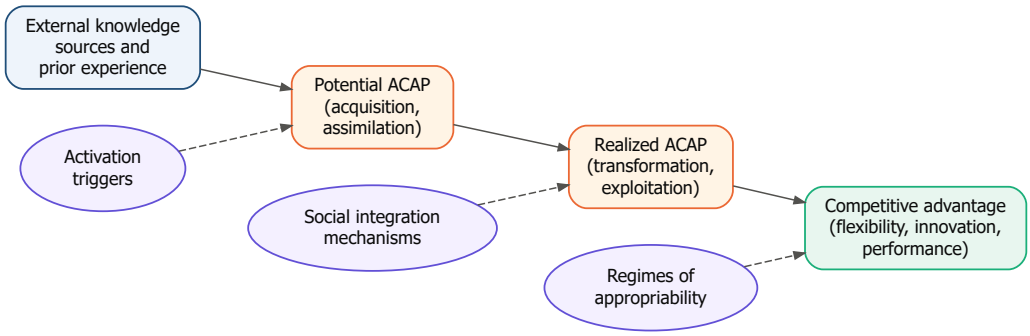


Figure 5.3. Potential and realized absorptive capacity. External knowledge sources and prior experience feed potential capacity (acquisition and assimilation). Social integration mechanisms convert it into realized capacity (transformation and exploitation), which yields advantage when appropriability regimes protect the results. Adapted from Zahra and George (2002).

BUSINESS CASE

Illustrative scenario: open innovation (Section 12.6) in a consumer-goods firm. A packaged-food company launches an open-innovation portal and a corporate venture unit to scout start-ups and university labs (potential absorptive capacity). Two years later, the scouting team has catalogued hundreds of technologies, yet few have reached products. Diagnosis using Zahra and George's model points to weak realized capacity: the scouts report to a corporate strategy office with no product responsibility, business-unit R&D teams lack the food-science expertise to evaluate fermentation-based proteins, and there is no forum in which scouts and product managers jointly prioritize opportunities. Remedies follow the model: rotate R&D staff through the scouting unit, create joint stage-gate reviews (a social integration mechanism), and invest selectively in internal research in the most promising fields to build the prior knowledge needed to evaluate them.

COMMON PITFALL

Measuring absorptive capacity with R&D intensity alone. R&D spending is the most common proxy, but it captures, at best, investment in potential capacity in technical domains. It misses non-technical knowledge (market, regulatory), the transformation and exploitation components, and the internal integration mechanisms that Cohen and Levinthal themselves emphasised. Combine input measures with process indicators (boundary-spanning roles, cross-unit forums) and outcome measures (citations to external knowledge in the firm's patents, speed of adopting external innovations).

5.3 Exploration and Exploitation

THEORY CARD

Core claim. Adaptive systems must divide resources between **exploration** of new possibilities and **exploitation** of old certainties. Because the returns to exploitation are surer, faster, and closer, learning processes tend to favour it, which improves short-run performance but can be self-destructive in the long run.
Origins. Organizational learning and behavioural theory of the firm, March (1991), Levinthal and March (1993), B. Levitt and March (1988).
Level of analysis. Individual, organization, and population of organizations.
Key constructs. Exploration, exploitation, competency trap, success trap, failure trap, temporal, spatial, and failure myopia, mutual learning.
Boundary conditions. The tension is sharpest when resources are scarce, when exploration and exploitation require different structures and cultures, and when environments change.
Related theories. Ambidexterity, Dynamic capabilities, Absorptive capacity, Bounded rationality (Section 23.1).

5.3.1 Core logic

March (1991) framed organizational adaptation as a choice between two kinds of activity. **Exploration** includes search, variation, risk taking, experimentation, play, flexibility, discovery, and innovation. **Exploitation** includes refinement, choice, production, efficiency, selection, implementation, and execution. Both are essential, but they compete for scarce resources and they have different return profiles: the returns to exploitation are positive, proximate, and predictable, whereas the returns to exploration are uncertain, distant, and often negative. Because organizations learn from feedback, and feedback from exploitation arrives sooner and more reliably, adaptive processes tend to refine exploitation faster than exploration. The result is effective in the short run and potentially self-destructive in the long run.

March illustrated the argument with a model of **mutual learning** between an organizational code (beliefs embodied in procedures and norms) and its members. The code learns from individuals whose beliefs are closer to reality, and individuals are socialized to the code. Fast socialization speeds convergence but eliminates the diversity from which the code learns. Slower learners, and a moderate turnover of personnel, preserve useful variety and improve long-run knowledge. Diversity, in other words, is an exploration resource.

B. Levitt and March (1988) had already identified the **competency trap**: favourable performance

with an inferior procedure leads an organization to accumulate more experience with it, which keeps experience with a superior procedure too low to make it rewarding to use. Levinthal and March (1993) generalized the problem as the **myopia of learning**. Learning tends to overlook the long run (**temporal myopia**), distant parts of the organization or environment (**spatial myopia**), and failures (**failure myopia**). Two dynamic traps follow. In a **success trap**, exploitation yields success, which encourages more exploitation, which drives out exploration. In a **failure trap**, exploration produces failure, which prompts further search for something new, which produces further failure, so the organization never stays long enough to become competent at anything.

5.3.2 Mechanism: an illustrative simulation

The exploration-exploitation problem has a precise analogue in the **multi-armed bandit** problem of decision theory and reinforcement learning (R. S. Sutton and Barto 2018). A decision maker repeatedly chooses among options with unknown, possibly changing payoffs. Choosing the option with the highest estimated payoff exploits current knowledge. Choosing another option explores. A simple policy, ε**-greedy**, exploits with probability $1 - \varepsilon$ and explores at random with probability ε.

Figure 5.4 applies the policy to a stylised firm. The firm knows its core business well. An emerging opportunity starts with low payoffs but improves over time, while the core business slowly declines. Unfamiliar options are initially valued pessimistically, as they often are inside established firms. A pure exploiter ($\varepsilon = 0$) earns the most early on and then falls into a competency trap: it never learns enough about the emerging option to switch. A heavy explorer ($\varepsilon = 0.3$) pays a large exploration tax while the environment is stable, but detects and moves to the emerging opportunity fastest once it improves. A moderate explorer ($\varepsilon = 0.1$) gives up a little early payoff and adapts more slowly, but it does adapt.

Figure 5.5 varies the exploration rate systematically and reports average payoffs in the stable early phase, after the environmental shift, and over the whole horizon. Early payoffs fall monotonically with exploration: exploration is a tax on current performance. Late payoffs rise sharply as soon as the firm explores at all and fall again when exploration becomes excessive. The whole-horizon payoff is an inverted U: too little exploration produces a competency trap. Too much produces the failure-trap pattern of never capitalizing on what has been learned. The optimal balance depends on how quickly and how much the environment changes: the core contingency in March's argument.

The simulation is deliberately simple, but it reproduces three of the theory's core claims: exploitation is rewarded sooner and more reliably. A firm that learns only from its own experience with familiar options can be trapped by its own competence, and the right balance is contingent on environmental change. It also illustrates why the balance is hard to manage in practice: the cost of exploration is visible in every quarterly report, whereas the cost of under-exploration is visible only after the environment shifts.

5.3.3 Evidence and extensions

Empirical and conceptual work has refined March's framing in several ways. Benner and Tushman (2003) showed that **process management** practices, such as quality programmes and process certification, increase exploitation and, if applied throughout the organization, crowd out exploration. They argued for insulating exploratory units from such practices. A. K. Gupta, Smith, and Shalley

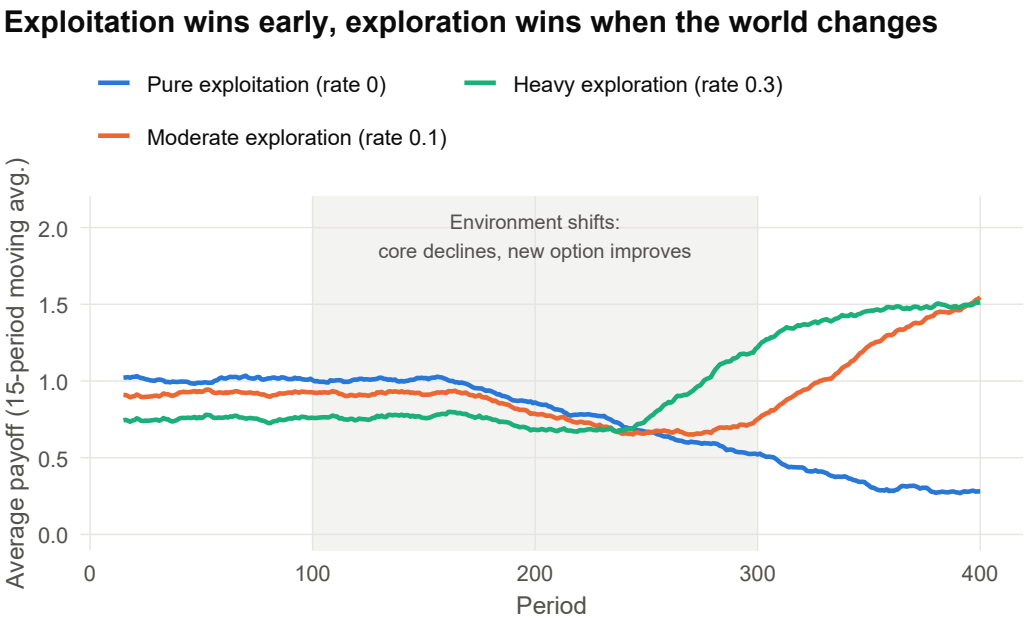


Figure 5.4. Illustrative bandit simulation of exploration and exploitation. A firm chooses each period among a known core business that slowly declines, an emerging opportunity that improves, and four other options. The pure exploiter wins early but falls into a competency trap. Exploration pays off once the environment shifts.

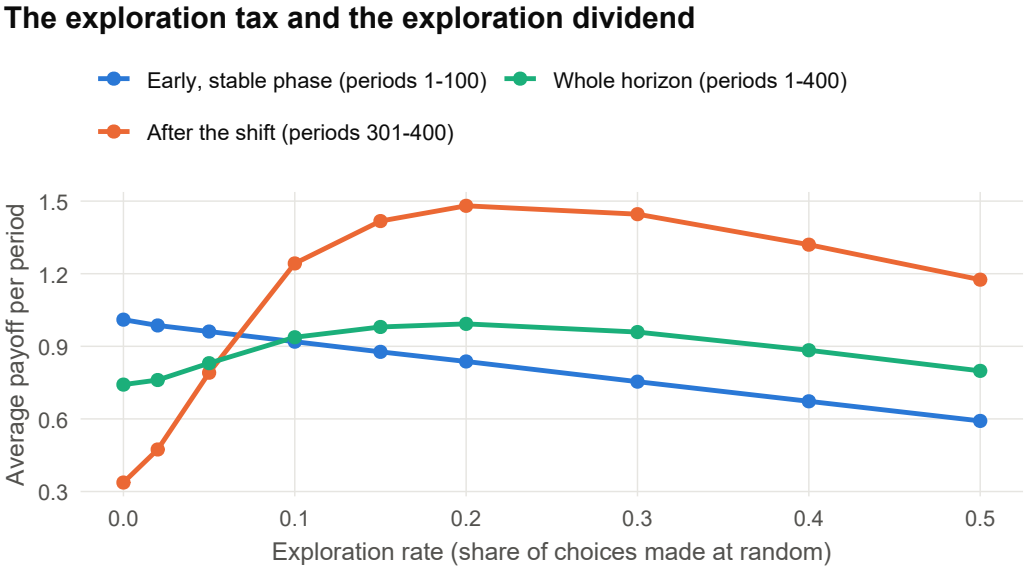


Figure 5.5. Illustrative trade-off between exploration and performance in the same stylised bandit environment. Exploration lowers early payoffs but raises payoffs after the environment shifts. Total payoff peaks at an intermediate exploration rate.

(2006) asked whether exploration and exploitation are two ends of one continuum (competing for the same resources) or **orthogonal** activities that can both be high, and concluded that the answer depends on the level of analysis and on whether the activities draw on the same scarce resources. Lavie, Stettner, and Tushman (2010) reviewed the literature and distinguished several ways firms balance the two: **organizational separation** (different units), **temporal separation** (alternating over time), and **domain separation** (exploring in some domains, such as alliances, while exploiting in others).

5.3.4 Disruptive innovation

Christensen (1997) provided a vivid, partly complementary account of why successful firms underexplore. In studies of the disk-drive industry and other sectors, he distinguished **sustaining** technologies, which improve products along dimensions mainstream customers already value, from **disruptive** technologies, which initially underperform on those dimensions but offer other benefits (simplicity, convenience, lower cost) valued by fringe or new customers. Incumbents typically lead in sustaining innovation. They lose to disruptive innovations not because they are badly managed but because good management, listening to the best customers, investing in the highest-margin opportunities, directs resources away from small, low-margin markets that later grow. In the language of this chapter, the firm's resource-allocation processes are an exploitation machine. Christensen's remedy, an autonomous organization with a cost structure suited to the new market, is a form of structural ambidexterity.

DEBATE AND EVIDENCE

How general is disruption? The theory of disruptive innovation (Section 12.2) has been widely applied, some argue over-applied, to any successful entrant. Scholars have questioned whether disruption can be identified *ex ante* rather than only in hindsight, and have noted incumbents that responded successfully to apparently disruptive threats. For research purposes, the theory is most useful when its specific conditions are stated and measured: an entrant whose initial product underperforms on mainstream attributes, a performance trajectory that eventually meets mainstream needs, and incumbent resource-allocation processes that systematically favour existing customers.

5.4 Organizational Ambidexterity

THEORY CARD

Core claim. Organizations can achieve long-run success by simultaneously exploiting existing competences and exploring new ones: through structurally separate units integrated by senior leadership, through a context that enables individuals to divide their own time, or through sequential shifts over time.

Origins. Organization theory and strategy, Tushman and O'Reilly (1996), I. O'Reilly Charles A. and Tushman (2004), Gibson and Birkinshaw (2004), reviewed in Raisch and Birkinshaw (2008) and I. O'Reilly Charles A. and Tushman (2013).

Level of analysis. Firm, business unit, senior team, and individual.

Key constructs. Structural ambidexterity, contextual ambidexterity (alignment and adaptability), sequential (temporal) ambidexterity, senior team integration, organizational context

(discipline, stretch, support, trust).
Boundary conditions. Benefits are greatest for firms with the resources to sustain both activities and in environments that are dynamic. Small firms and stable environments may benefit more from focus.
Related theories. Exploration and exploitation, Dynamic capabilities, Contingency theory (Section 7.6), Upper echelons theory (Section 6.4).

5.4.1 Three forms of ambidexterity

Tushman and O'Reilly (1996) argued that long-term success requires firms to compete in mature markets, where cost, efficiency, and incremental innovation matter, while also developing new products and services for emerging markets, where experimentation, speed, and flexibility matter. Because these require different structures, processes, and cultures, they proposed the **ambidextrous organization**: separate units for exploration and exploitation, each with its own alignment of people, structure, and culture, held together by a senior team and a common strategic intent. I. O'Reilly Charles A. and Tushman (2004) reported that firms using such structures were more successful at launching breakthrough innovations without sacrificing the existing business than firms that relied on functional designs, cross-functional teams embedded in the existing organization, or unsupported independent units. The key was integration at the top: separate units, but a senior team that shared resources and resolved conflicts between them.

Gibson and Birkinshaw (2004) proposed a different route. **Contextual ambidexterity** is the behavioural capacity to demonstrate both **alignment** (coherence among activities in pursuit of current goals) and **adaptability** (the capacity to reconfigure activities quickly to meet changing demands) across an entire business unit. It arises not from structural separation but from an organizational context that combines **performance management** (discipline and stretch) with **social support** (support and trust), allowing individuals to decide for themselves how to divide time between conflicting demands. In their study of business units in multinational firms, contextual ambidexterity mediated the relationship between organizational context and performance.

A third form, **sequential ambidexterity**, has firms shift their structures over time, exploring in one period, exploiting in the next, which may suit slower environments or smaller firms (I. O'Reilly Charles A. and Tushman 2013). Table 5.6 compares the three.

Table 5.6. Three forms of organizational ambidexterity.

Dimension	Structural ambidexterity	Contextual ambidexterity	Sequential ambidexterity
Key sources	Tushman & O'Reilly (1996), O'Reilly & Tushman (2004)	Gibson & Birkinshaw (2004)	O'Reilly & Tushman (2013)
How the tension is resolved	Separate units for exploration and exploitation	Individuals divide time between alignment and adaptability	The organization alternates between modes over time
Where integration happens	Senior team, shared vision, targeted linkages	Organizational context and individual judgement	Leadership-driven transitions
Key enablers	Distinct structures, metrics, and cultures, senior-team integration	Discipline, stretch, support, trust	Capacity to restructure, change management
Best suited to	Large firms facing discontinuous change	Business units with incremental, adjacent innovation	Slower-moving environments, smaller firms

Table 5.6. Three forms of organizational ambidexterity. *(continued)*

Dimension	Structural ambidexterity	Contextual ambidexterity	Sequential ambidexterity
Main risk	Isolation or starvation of exploratory units	Overload, exploitation crowds out exploration	Disruption and lost momentum at each shift

5.4.2 Evidence and extensions

The evidence broadly supports a positive association between ambidexterity and performance, with important qualifications. He and Wong (2004), in a study of manufacturing firms, found that the *interaction* of exploration and exploitation was positively related to sales growth, while a *relative imbalance* between them was negatively related: an early direct test of the ambidexterity hypothesis. Reviews by Raisch and Birkinshaw (2008) and I. O'Reilly Charles A. and Tushman (2013) conclude that the benefits of ambidexterity depend on environmental dynamism, competitive intensity, and firm size and resources, and that ambidexterity is better understood as a **dynamic capability** of senior leadership than as a structural formula. They also note persistent measurement debates: whether to combine exploration and exploitation additively (combined dimension) or as a difference (balance dimension), and at which level to measure them.

BUSINESS CASE

IBM's Emerging Business Opportunities process. IBM's history illustrates both the danger of core rigidities and the use of ambidextrous structures. In the early 1990s the company suffered heavy losses as the computer industry shifted away from mainframes. Louis Gerstner, who became chief executive in 1993, reversed plans to break the company into independent units and instead repositioned it around integrated solutions and services (Jr. Gerstner Louis V. 2002). Later, IBM created an **Emerging Business Opportunities (EBO)** process, begun in 2000, to protect promising new businesses from the short-term metrics of the established divisions: EBOs received dedicated leaders, separate funding, milestones suited to exploration, and regular senior-executive review (Harreld, O'Reilly, and Tushman 2007). Subsequent portfolio moves, selling the personal computer business to Lenovo in 2005, acquiring Red Hat in 2019, and spinning off its managed-infrastructure services business as Kyndryl in 2021, show sequential reconfiguration of the asset base. Harreld, O'Reilly, and Tushman (2007) interpret IBM's approach as an attempt to build dynamic capabilities in the senior leadership system rather than relying on any single reorganization.

5.4.3 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The positive effect of exploratory innovation on firm performance will be stronger in more dynamic environments, and the positive effect of exploitative innovation stronger in more competitive environments.
- *H2.* Structural separation of exploratory units will improve their innovation outcomes only when the senior team shares integrative goals and a common reward system.
- *H3.* Business units whose context combines high discipline and stretch with high sup-

port and trust will exhibit higher contextual ambidexterity, which will mediate the effect of context on unit performance (Gibson and Birkinshaw 2004).

Designs. Panel studies coding exploration and exploitation from patents (new versus familiar technology classes), product launches, or alliance types, multilevel surveys of business units, field experiments on time allocation, agent-based simulations in the tradition of March (1991).

Measures. Exploration and exploitation scales (e.g., following He and Wong (2004)). Patent-based measures of search distance. Share of R&D budget devoted to new domains. Balance and combined ambidexterity indices.

COMMON PITFALL

Treating ambidexterity as "more of everything". Ambidexterity does not mean maximizing both exploration and exploitation everywhere. The theory's core insight is that the two compete for scarce resources and require different structures and cultures, so they must be deliberately separated, sequenced, or contextually balanced, and integrated by leaders who manage the tension. Studies that simply add exploration and exploitation scores without modelling that tension, or that ignore resource constraints and environmental conditions, test a weaker claim than the theory makes.

5.5 Chapter Summary

KEY TAKEAWAYS

- **Dynamic capabilities** are the firm's ability to integrate, build, and reconfigure competences to address changing environments. They rest on **processes**, are shaped by **positions**, and are constrained by **paths** (Teece, Pisano, and Shuen 1997).
- Teece (2007) decomposes them into **sensing**, **seizing**, and **transforming**, each with microfoundations in managerial skill, processes, and structure.
- Eisenhardt and Martin (2000) view dynamic capabilities as identifiable **best practices** whose form depends on market dynamism. Advantage comes from the resource configurations they create.
- A **capability hierarchy** separates ordinary from dynamic capabilities (Winter 2003). Dynamic capabilities are built through **deliberate learning** (Zollo and Winter 2002) and evolve over a **life cycle** (Helfat and Peteraf 2003).
- Core capabilities can become **core rigidities** (Leonard-Barton 1992). Cognition as well as capability shapes adaptation (Tripsas and Gavetti 2000).
- **Absorptive capacity** depends on prior related knowledge. **Potential** capacity must be converted into **realized** capacity through integration mechanisms (W. M. Cohen and Levinthal 1990, Zahra and George 2002).
- Learning favours **exploitation** over **exploration**, creating success and competency traps (March 1991, Levinthal and March 1993). Disruptive innovation is a vivid case (Christensen 1997).
- **Ambidexterity**, structural, contextual, or sequential, is the main organizational answer, and is best seen as a leadership capability (I. O'Reilly Charles A. and Tushman 2013, Gibson and Birkinshaw 2004).

Table 5.7 presents the comparison of the chapter's theories.

Table 5.7. Comparison of the chapter's theories.

Theory	Unit of analysis	Key mechanism	Signature prediction	Typical business use
Dynamic capabilities (Teece)	Firm	Sensing, seizing, transforming	Firms with strong dynamic capabilities sustain performance through change	Strategic renewal, business-model change
Best-practice view (Eisenhardt & Martin)	Process	Effective processes create new resource configurations	Simple rules outperform detailed routines in high-velocity markets	Designing product-development and M&A processes
Capability hierarchy and learning	Routine	Learning mechanisms modify operating routines	Codification pays for infrequent, heterogeneous tasks	Building integration playbooks
Core rigidities	Firm / project	Strengths screen out misfit projects	Incumbents struggle with competence-destroying change	Diagnosing barriers to innovation
Absorptive capacity	Firm / unit	Prior knowledge enables learning from outside	Internal R&D raises returns to external knowledge	Open innovation, alliance and M&A strategy
Exploration and exploitation	Organization	Feedback favours sure, near returns	Success and competency traps	Portfolio and R&D allocation
Ambidexterity	Firm / business unit	Separation plus integration, or supportive context	Balanced, integrated firms outperform over time	Organizing for innovation

5.6 Discussion Questions

1. Compare the Teece and the Eisenhardt-and-Martin views of dynamic capabilities. Which is more convincing for a software company in a high-velocity market, and which for a utility in a regulated market? Why?
2. Kodak had valuable resources and early knowledge of digital technology. Using the concepts of core rigidities, managerial cognition, and exploration versus exploitation, explain why this was not enough. What could have been done differently, and at what cost?
3. Design a study that measures dynamic capabilities *before* an environmental shock and tests whether they predict adaptation afterwards. What shock would you use, and how would you avoid defining the capability by its outcome?
4. A firm's open-innovation unit identifies many promising technologies but commercializes few. Using the potential/realized absorptive capacity distinction, diagnose the problem and propose remedies.
5. In the bandit simulation, the best exploration rate depends on how the environment changes. Translate this into a managerial rule of thumb for R&D portfolio allocation, and identify the indicators that should trigger a shift.
6. When would contextual ambidexterity be preferable to structural separation? Use an example from a function such as sales, operations, or HR.

5.7 Further Reading

- Teece, Pisano, and Shuen (1997): the founding statement of the dynamic capabilities framework, including processes, positions, and paths.
- Teece (2007): the sensing-seizing-transforming framework and its microfoundations.
- Eisenhardt and Martin (2000): the influential alternative view of dynamic capabilities as identifiable processes whose form depends on market dynamism.
- Winter (2003) and Zollo and Winter (2002): short, precise papers on what dynamic capabilities are and where they come from.
- Leonard-Barton (1992): the classic account of how core capabilities become core rigidities.
- W. M. Cohen and Levinthal (1990) and Zahra and George (2002): the foundational and the reconceptualized statements of absorptive capacity.
- March (1991): one of the most cited papers in organization science. Essential for understanding exploration and exploitation.
- I. O'Reilly Charles A. and Tushman (2013): an accessible review of ambidexterity research and its open questions.

Chapter 6

Governance, Contracts, and Leadership

In 2016, regulators announced that employees at Wells Fargo, then one of the most admired retail banks in the United States, had opened large numbers of deposit and credit-card accounts that customers had never requested. No senior executive had ordered anyone to commit fraud. The bank had instead built an aggressive system of cross-selling targets, branch rankings, and incentive pay, and thousands of employees had responded to that system exactly as it rewarded them to. A few years earlier, scholars had begun asking a related question about the executives at the top of large firms: why do some chief executives pay enormous premiums for acquisitions that destroy their own shareholders' wealth, and why do boards let them?

Both puzzles are questions about **governance**: the arrangements by which one party gets another to act on its behalf when their interests, information, and attitudes toward risk differ. This chapter develops four theories that together form the core of organizational economics and strategic leadership research. **Agency theory** analyses how principals design contracts and monitoring to control self-interested agents. **Stewardship theory** argues that the agency model of human nature is too narrow and that many managers are better motivated by trust and identification than by incentives. **Transaction cost economics** asks a more fundamental question, why some transactions are organized inside firms rather than in markets, and answers it with the hazards of specific investment. **Upper echelons theory**, finally, opens the black box of the executive suite and argues that organizations become reflections of the people who lead them (Table 6.1).

Table 6.1. Chapter at a glance: theories of governance, contracts, and leadership.

Theory	Core question	Level of analysis	Key source
Agency theory	How can a principal align a self-interested agent's actions with its own goals?	Dyad (principal-agent), firm	Jensen & Meckling (1976), Eisenhardt (1989)
Stewardship theory	When do managers act as pro-organizational stewards rather than opportunistic agents?	Dyad, firm	Davis, Schoorman & Donaldson (1997)
Transaction cost economics	Which transactions should be organized by markets, hybrids, or hierarchies?	Transaction, firm boundary	Coase (1937), Williamson (1975, 1985)
Upper echelons theory	How do executives' experiences, values, and personalities shape strategy and performance?	Executive, top management team, firm	Hambrick & Mason (1984), Hambrick (2007)

6.1 Agency Theory

THEORY CARD

Core claim. When a principal delegates work to an agent whose goals differ and whose actions or information the principal cannot fully observe, the agent will not always act in the principal's interest. Contracts, monitoring, and incentives can reduce, but not eliminate, the resulting **agency costs**.

Origins. Economics and finance. The formal principal's problem (S. A. Ross 1973, Holmström 1979), the theory of the firm as a nexus of contracts (Jensen and Meckling 1976), and the separation of ownership and control (Berle and Means 1932, Fama and Jensen 1983). Imported into management by Eisenhardt (1989a).

Level of analysis. Dyad (shareholder-manager, employer-employee, franchisor-franchisee, client-supplier) and the firm as a set of such contracts.

Key constructs. Principal, agent, goal conflict, information asymmetry, moral hazard, adverse selection (Section 8.1), risk preferences, outcome- vs. behaviour-based contracts, monitoring, bonding, residual loss.

Boundary conditions. Goal conflict must be present. The agent's effort or information must be costly for the principal to observe. Agents are assumed self-interested, boundedly rational, and typically more risk-averse than principals.

Related theories. Information asymmetry (Section 8.1), Screening (Section 8.3), Stewardship theory, Transaction cost economics, Resource dependence (Section 7.1).

6.1.1 Origins and intellectual history

The worry that hired managers might not serve owners is older than modern economics, but it acquired its canonical form when Berle and Means (1932) documented that ownership of large American corporations had become dispersed among thousands of small shareholders, none with enough at stake to supervise management. Control had passed, in practice, to salaried executives. Four decades later, economists supplied the analytical machinery to study the problem.

S. A. Ross (1973) posed the **principal's problem** in general terms: a principal must choose a compensation schedule for an agent whose action affects an uncertain payoff, anticipating how the agent will respond. Holmström (1979) showed that when effort is unobservable, a contract that insures a risk-averse agent fully destroys the incentive to work, while a contract that ties pay entirely to output imposes inefficient risk on the agent. He also derived the **informativeness principle**: any additional measure that carries information about the agent's action, however noisy, should enter the contract, because conditioning pay on it reduces the risk the agent must bear for a given level of incentive.

Jensen and Meckling (1976) moved the analysis from abstract contracts to the corporation. They defined an **agency relationship** as a contract under which one or more principals engage an agent to perform a service on their behalf, delegating some decision-making authority. Because the parties are utility maximizers, the agent will not always act in the principal's best interest, and the divergence generates costs that shape ownership structure, capital structure, and the very boundaries of the firm. Fama (1980) added that the managerial labour market disciplines executives over time, and Fama and Jensen (1983) explained how complex organizations survive despite the separation of ownership and control: they separate **decision management** (initiating and implementing de-

cisions) from **decision control** (ratifying and monitoring them), with the board of directors at the apex of the control system.

Eisenhardt (1989a) introduced agency theory to a broad management audience. Her review distinguished two streams that remain useful today. The **positivist** stream, rooted in finance, identifies situations in which principals and agents are likely to have conflicting goals and describes the governance mechanisms, boards, markets for corporate control, equity ownership, that limit self-serving behaviour, usually in large public corporations. The **principal-agent** stream is more formal and more general: it asks which contract is most efficient given assumptions about people (self-interest, bounded rationality (Section 23.1), risk aversion), organizations (goal conflict), and information (a purchasable commodity).

6.1.2 Core logic and propositions

Agency theory rests on two conditions. First, there is **goal conflict**: the agent values leisure, perquisites, empire, or job security in ways the principal does not. Second, there is **information asymmetry**: the principal cannot costlessly verify what the agent knows or does. Remove either condition and the problem disappears: an agent with identical goals needs no monitoring, and a perfectly observed agent can be paid for the behaviour the principal wants.

Eisenhardt (1989a) separates two problems that arise from these conditions. The **agency problem** proper arises when goals conflict and the principal cannot verify the agent's behaviour. It comes in two forms, moral hazard and adverse selection. The **problem of risk sharing** arises when principal and agent have different attitudes toward risk, so they prefer different actions and different allocations of uncertain outcomes. Table 6.2 summarises these problems with business examples.

Table 6.2. The problems that agency theory addresses (following Eisenhardt, 1989).

Problem	What goes wrong	Timing	Business example	Typical remedy
Moral hazard (hidden action)	Agent shirks or diverts effort once the principal cannot observe it	After contracting	Sales staff push easy products rather than serving customers. Managers consume perquisites	Monitoring, outcome-based pay, performance audits
Adverse selection (hidden information)	Agent misrepresents ability or private information when contracting	Before contracting	Candidate overstates skills, supplier understates its costs, manager hides a project's risks	Screening, probation, credentials, self-selecting contract menus
Risk sharing	Principal and agent prefer different risk exposures and different actions	Throughout	Diversified shareholders want risky growth, an undiversified CEO prefers safe diversification	Equity and option grants, deferred pay, risk-adjusted targets

Because these problems cannot be eliminated without cost, the theory's central quantity is **agency cost**. Jensen and Meckling (1976) defined it as the sum of three components, shown in Table 6.3.

Table 6.3. The three components of agency costs (Jensen & Meckling, 1976).

Component	Definition	Who bears it initially	Examples
Monitoring expenditures	Resources the principal spends to observe and constrain the agent	Principal	External audits, board committees, budget controls, performance reviews, surveillance software
Bonding expenditures	Resources the agent spends to guarantee that it will not harm the principal (or will compensate if it does)	Agent	Financial reporting commitments, personal guarantees, deferred compensation subject to clawback, covenants
Residual loss	The value lost because the agent's decisions still differ from those that would maximize the principal's welfare	Principal	Value-destroying acquisitions, under-investment in risky innovation, excessive perquisites

A key insight is that agency costs are ultimately borne by whoever can be made to bear them in a competitive market. If investors anticipate that an owner-manager who sells equity will consume more perquisites, they price the shares accordingly. The owner therefore has an incentive to commit, *ex ante*, to monitoring and bonding arrangements that raise the share price (Jensen and Meckling 1976). Governance is not only imposed on managers. Managers often choose it.

The principal-agent stream generates sharper propositions about contract choice. Eisenhardt (1989a) frames the decision as a choice between **outcome-based contracts** (commissions, bonuses, stock options, profit sharing) and **behaviour-based contracts** (salaries combined with supervision and evaluation of behaviour). The following propositions summarise her argument:

- **P1 (Alignment).** When contracts are outcome-based, agents are more likely to act in the principal's interest.
- **P2 (Information).** When the principal has information systems that reveal agent behaviour, agents are more likely to act in the principal's interest, and behaviour-based contracts become more attractive.
- **P3 (Uncertainty and risk).** Outcome uncertainty, combined with agent risk aversion, favours behaviour-based contracts, because outcome-based pay would load costly risk on the agent.
- **P4 (Task characteristics).** Programmable tasks (whose appropriate behaviour can be specified in advance) favour behaviour-based contracts. Easily measured outcomes favour outcome-based contracts.
- **P5 (Relationship length).** In long relationships the principal learns about the agent, so behaviour-based contracts become more feasible (Table 6.4).

Table 6.4. Contingencies in contract choice: when outcome- versus behaviour-based contracts are more efficient (adapted from the propositions in Eisenhardt, 1989).

Condition	Favours outcome-based contract	Favours behaviour-based contract
Information systems (ability to observe behaviour)	Weak	Strong
Outcome uncertainty	Low	High
Agent risk aversion	Low	High
Principal risk aversion	High	Low

Table 6.4. Contingencies in contract choice: when outcome- versus behaviour-based contracts are more efficient (adapted from the propositions in Eisenhardt, 1989). *(continued)*

Condition	Favours outcome-based contract	Favours behaviour-based contract
Goal conflict	High	Low
Task programmability	Low (creative, non-routine work)	High (routine, specifiable work)
Outcome measurability	High	Low
Length of relationship	Short	Long

6.1.3 Mechanism

Figure 6.1 summarises the causal structure. Goal conflict and information asymmetry jointly create room for opportunism. Governance mechanisms are the levers principals use to narrow that room, at a cost.

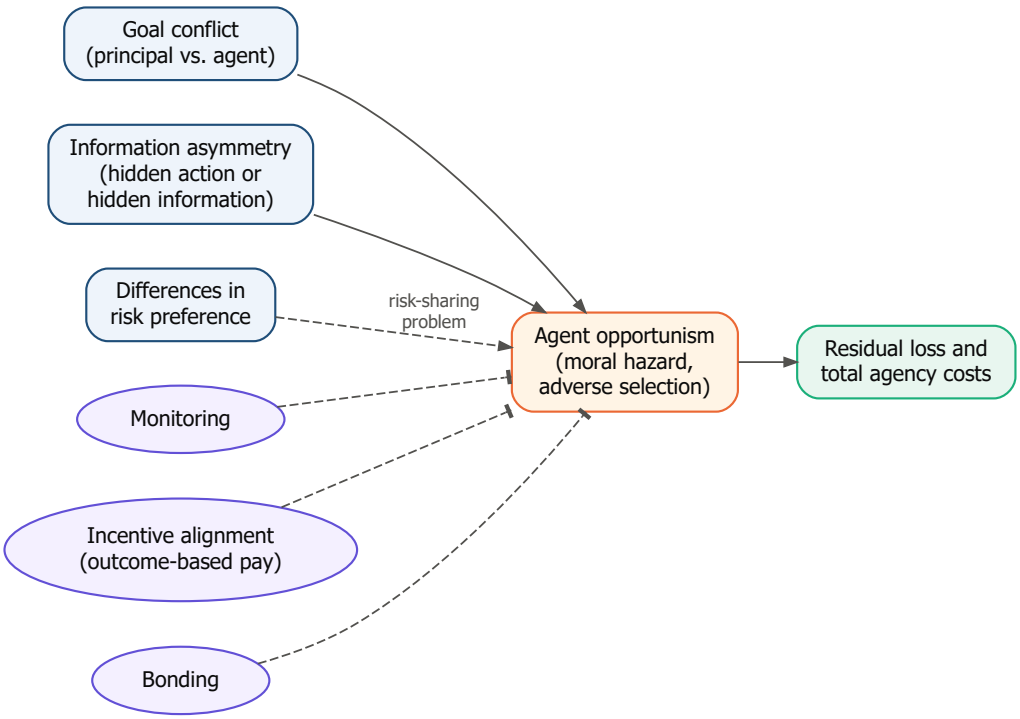


Figure 6.1. Mechanism of agency theory: goal conflict and information asymmetry enable agent opportunism (moral hazard, adverse selection), which produces residual loss. Monitoring, incentive alignment, and bonding reduce opportunism but are themselves costly.

The trade-off at the heart of incentive design can be made precise. In the linear model associated with Holmström (1979) and Holmström and Milgrom (1991), an agent's output equals effort plus noise, pay equals a fixed salary plus a share β of output, and the agent dislikes both effort and risk. A higher **pay-for-performance slope** β elicits more effort, but it also exposes the agent to more of the noise, for which a risk-averse agent must be compensated through a **risk premium**. The efficient slope balances the two, and it falls as performance measures become noisier and as agents become

more risk-averse. Figure 6.2 plots the components of joint surplus for one parameterisation.

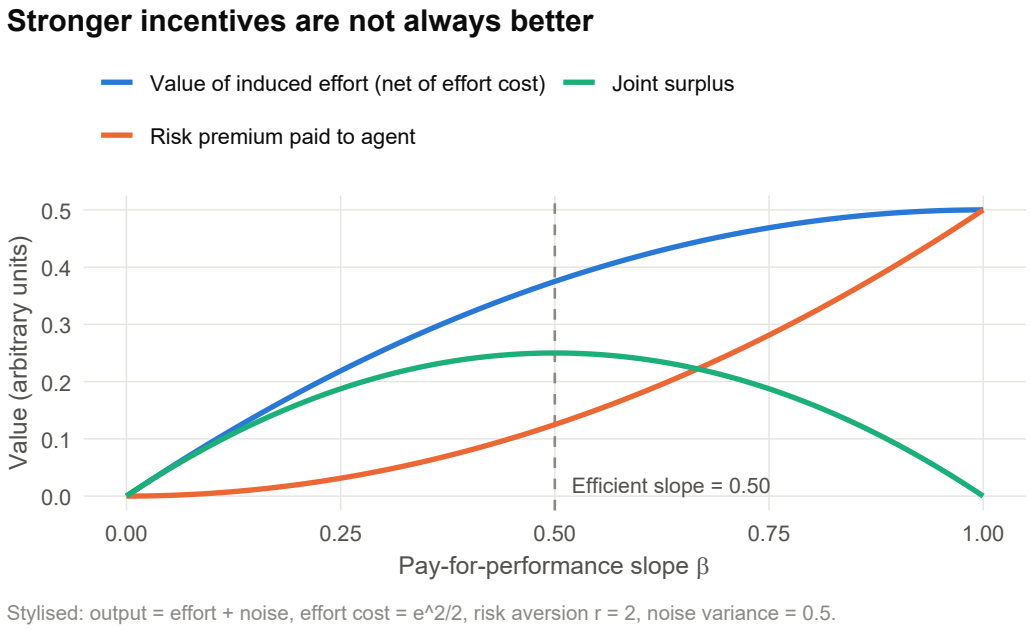


Figure 6.2. Illustrative incentive-intensity trade-off in a stylised linear principal-agent model. Raising the pay-for-performance slope increases the value created by effort but also the risk premium the agent must be paid. Joint surplus peaks at an interior slope (dashed line).

Figure 6.3 generalises the point: the efficient slope $\beta^* = 1/(1 + rk\sigma^2)$ declines as measurement noise rises, and it declines faster for more risk-averse agents. This is the formal reason why pure commission pay is common in transactional sales, where individual output is easy to measure, and rare for research scientists, whose output is noisy and delayed.

6.1.4 Evidence and extensions

Agency theory has generated one of the largest empirical literatures in management and finance. Several findings recur.

- **Pay-performance sensitivity.** Jensen and Murphy (1990) estimated that the wealth of chief executives of large U.S. firms changed by only a few dollars for every thousand-dollar change in shareholder wealth, and argued that pay was far less sensitive to performance than agency theory would recommend. The finding helped legitimise the shift toward equity-based compensation in the 1990s.
- **Unintended consequences of strong incentives.** Incentives tied to stock prices are associated with more aggressive earnings management (Bergstresser and Philippon 2006), and in a large survey many financial executives reported that they would sacrifice long-term economic value, for example, by delaying a positive net-present-value project, to meet short-term earnings targets (Graham, Harvey, and Rajgopal 2005). Lie (2005) showed that share prices tended to fall before and rise after executive option grants, a pattern consistent with grant dates being timed or backdated.
- **Multitasking.** Holmström and Milgrom (1991) showed that when an agent performs several tasks that differ in measurability, strong incentives on the measurable task draw effort away

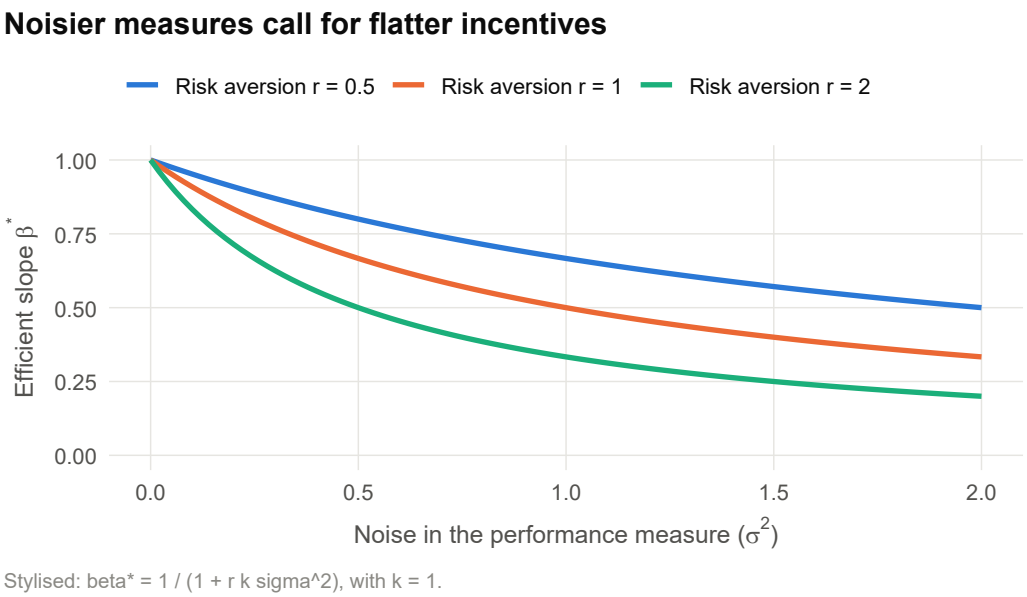


Figure 6.3. Illustrative comparative statics: the efficient pay-for-performance slope falls as performance measures become noisier, and falls faster for more risk-averse agents (stylised linear principal-agent model).

from the unmeasured ones, so low-powered incentives can be optimal. This formalises the older warning of S. Kerr (1975) against "rewarding A while hoping for B".

- **Boards and monitoring.** Meta-analyses find little systematic relationship between board independence or the separation of CEO and chair roles and firm financial performance (Dalton et al. 1998). Reviews of corporate governance conclude that agency theory's prescriptions are better supported for specific outcomes (e.g., takeover resistance, compensation design) than for overall performance (Daily, Dalton, and Cannella 2003).
- **Symbolic adoption.** Firms sometimes adopt incentive plans without using them: Westphal and Zajac (1994) found that many companies announced long-term incentive plans for their CEOs but did not actually grant awards under them, while still enjoying a favourable market reaction, an early demonstration that governance can be ceremonial (see Institutional theory (Section 7.2)).

The theory has also been extended in behavioural directions. The **behavioural agency model** of Wiseman and Gomez-Mejia (1998) combines agency logic with prospect theory (Chapter 24): because executives evaluate outcomes relative to reference points, the same option grant can make an executive more risk-seeking when the firm is below target and more risk-averse when accumulated option wealth is at stake.

DEBATE AND EVIDENCE

Is agency theory a self-fulfilling prophecy? Ghoshal (2005) argued that management education, by teaching that managers are opportunists who must be controlled, has helped create the behaviour it assumes: high-powered incentives and intensive monitoring can crowd out intrinsic motivation and signal distrust, and executives trained in the theory may come to see self-interest as legitimate. Defenders reply that the theory is a model of *contracting under risk*, not a description of human nature, and that its key predictions, that people respond to incen-

tives, including badly designed ones, are among the best-supported in social science. A second line of critique, L. Bebchuk and Fried (2004)'s **managerial power** view, accepts agency assumptions but argues that executive pay is itself an agency problem: powerful CEOs influence the boards that set their compensation, so pay packages reflect rent extraction rather than optimal contracting.

6.1.5 Applications in business

BUSINESS CASE

Executive stock options and short-termism. During the 1990s, U.S. boards shifted executive pay heavily toward stock options, in part on the agency argument that equity-based pay aligns executives with shareholders (Jensen and Murphy 1990). The results were mixed in ways the theory itself anticipates. Options reward share-price movements over a defined horizon, so they are an outcome-based contract with a noisy, manipulable measure. The 2005-2006 option-backdating revelations, in which grant dates were found to have been selected after the fact at numerous firms, followed academic evidence on suspicious timing (Lie 2005). Evidence linking equity incentives to earnings management (Bergstresser and Philippon 2006) and survey evidence of earnings-driven myopia (Graham, Harvey, and Rajgopal 2005) show the multitasking problem in practice: the measurable outcome (near-term stock price) can crowd out unmeasured long-term investment. Many boards responded by lengthening vesting periods, adding performance conditions and holding requirements, and adopting clawback provisions , in agency terms, combining outcome-based pay with bonding.

BUSINESS CASE

Wells Fargo's sales-practices scandal. In September 2016 the Consumer Financial Protection Bureau, the Office of the Comptroller of the Currency, and the Los Angeles City Attorney announced combined penalties of $185 million against Wells Fargo for sales practices in which employees had opened deposit and credit-card accounts without customers' authorization. The bank had fired roughly 5,300 employees over several years for related misconduct. Branch employees faced demanding cross-selling targets, and their pay and job security depended on meeting them. In agency terms, the bank used high-powered outcome-based incentives on a metric (accounts opened) that was easy to measure and easy to game, while the outcome it actually valued (profitable, durable customer relationships) was hard to measure, a textbook multitasking failure (Holmström and Milgrom 1991, S. Kerr 1975). The consequences extended well beyond the initial fine: the CEO resigned in 2016, the board clawed back compensation from senior executives, the Federal Reserve imposed a growth restriction on the bank in 2018, and in 2020 the bank agreed to pay $3 billion to resolve criminal and civil investigations by the U.S. Department of Justice and the Securities and Exchange Commission.

BUSINESS CASE

Franchising versus company-owned stores. A chain retailer or restaurant group must decide whether each outlet is run by a salaried manager (a behaviour-based contract with monitoring) or by a franchisee who owns the residual profit (an outcome-based contract). Agency theory predicts that franchising will be favoured where monitoring a salaried manager is

costly, for example, at outlets far from headquarters, because residual ownership makes the local operator self-monitoring. Brickley and Dark (1987) found exactly this pattern, and Lafontaine (1992) showed that franchising choices are better explained by incentive considerations than by the older idea that franchisors simply need franchisees' capital (see also Rubin 1978). Franchising creates a second agency problem in the other direction: franchisees may free-ride on the shared brand by cutting quality, which is why franchise contracts include detailed operating standards and inspection rights. McDonald's, which operates predominantly through franchisees, and Starbucks, which in its home market has relied mainly on company-operated and licensed stores, illustrate two different resolutions of the same trade-off.

6.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The share of franchised outlets in a chain will increase with the geographic dispersion of outlets from headquarters.
- *H2.* The positive association between CEO equity incentives and R&D intensity will be weaker when the CEO's accumulated option wealth is large relative to salary (behavioural agency model).
- *H3.* Sales commission intensity will be lower for products whose sales cycles are long and whose outcomes are noisy.

Designs. Panel data on compensation and governance (with firm fixed effects). Quasi-experiments from regulatory changes in disclosure, board composition, or takeover law. Field experiments on incentive schemes within firms. Structural models of contracts.

Measures. Pay-performance sensitivity (change in executive wealth per change in firm value), vesting horizon, ownership concentration, board independence and CEO duality, monitoring intensity (audit fees, analyst coverage), outcome measurability (noise in performance metrics).

Data sources. Proxy statements and compensation databases, franchise disclosure documents, personnel records and incentive-plan documents.

COMMON PITFALL

Treating every governance variable as a measure of "alignment". Board independence, CEO ownership, and option grants are often used as proxies for alignment without showing that they change the agent's behaviour. Such variables are themselves chosen, often by the agents they are supposed to discipline, so they are endogenous. State which agency problem (moral hazard, adverse selection, or risk sharing) a mechanism addresses, show the behavioural channel, and use designs that address selection.

6.2 Stewardship Theory

THEORY CARD

Core claim. Many managers are not opportunistic agents but **stewards** whose motives are aligned with the organization's objectives. For such managers, structures that empower and trust them produce better outcomes than structures that monitor and control them.
Origins. Organization theory and organizational psychology, L. Donaldson and Davis (1991), developed as a theory by J. H. Davis, Schoorman, and Donaldson (1997).
Level of analysis. Dyad (principal-manager) and firm (board and leadership structure).
Key constructs. Steward vs. agent "model of man", intrinsic motivation, organizational identification, personal vs. institutional power, involvement- vs. control-oriented management philosophy, mutual choice.
Boundary conditions. Both parties must choose a stewardship relationship. Predictions depend on the manager's motivation and on situational factors such as culture and management philosophy.
Related theories. Agency theory, Self-determination theory (Section 19.6), Social exchange theory (Section 9.1), Power distance (Section 18.3).

6.2.1 Core logic

L. Donaldson and Davis (1991) began with a governance puzzle. Agency theory implies that the roles of chief executive and board chair should be separated, so that the person who manages the firm is not the person who monitors management. Stewardship theory implies the opposite: if the executive is a steward, combining the roles gives a trustworthy leader clear authority and unity of command. Their empirical analysis of CEO duality reported results more consistent with the stewardship prediction than with the agency one.

J. H. Davis, Schoorman, and Donaldson (1997) generalized the argument. The **steward** derives utility from pro-organizational, collectivist behaviour. Achieving the organization's goals satisfies the steward's own higher-order needs for achievement, growth, and recognition. Where agency theory treats the manager as an economic actor who must be controlled, stewardship theory draws on a model of the person as self-actualizing. Table 6.5 contrasts the two.

Table 6.5. Agency theory and stewardship theory compared (organized following Davis, Schoorman, & Donaldson, 1997).

Dimension	Agency theory	Stewardship theory
Model of the person	Economic actor, self-serving	Self-actualizing, collective-serving
Motivation	Extrinsic (pay, perquisites, security)	Intrinsic (growth, achievement, affiliation)
Identification with organization	Low	High
Basis of power	Institutional (position, legitimate, coercive, reward)	Personal (expertise, referent)
Management philosophy	Control-oriented	Involvement-oriented
Cultural fit	Individualism, high power distance	Collectivism, low power distance
Governance prescription	Separate CEO and chair, independent monitoring, outcome-based pay	Empower executives, unify leadership, trust-based relationships

Main risk	Agent opportunism	Principal or steward betrayal

6.2.2 Mechanism: a game of mutual choice

The most influential element of J. H. Davis, Schoorman, and Donaldson (1997) is the idea that the relationship is a **mutual choice**. The principal chooses whether to treat the manager as an agent (with controls) or as a steward (with trust), and the manager chooses whether to behave as an agent or a steward. Figure 6.4 shows the four outcomes. When both choose stewardship, potential performance is maximized. When both choose agency, costs are minimized but potential is unrealized. The off-diagonal cells are the dangerous ones: a steward-type principal who trusts an opportunistic manager is betrayed, and a steward-type manager who is treated as an agent feels frustrated and may come to behave like one. Because each party is vulnerable to the other's choice, the structure resembles a prisoner's dilemma, and the safe mutual-agency outcome is common even when mutual stewardship would be better.

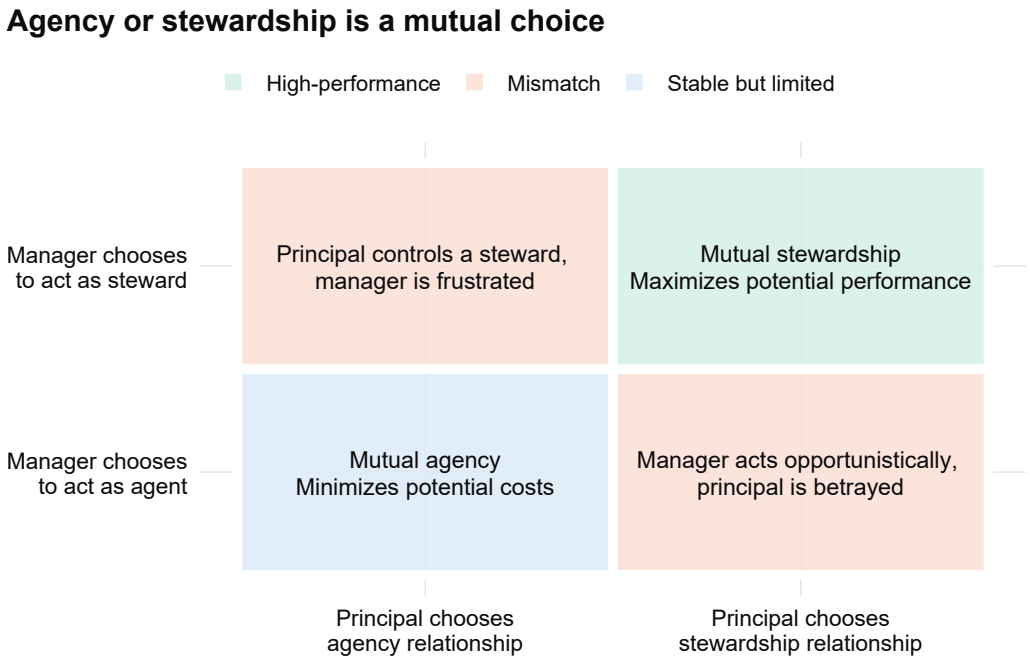


Figure 6.4. The principal-manager choice model, redrawn from the logic of Davis, Schoorman, and Donaldson (1997). Mutual stewardship maximizes potential performance, mutual agency minimizes potential costs, mismatched choices produce betrayal and frustration.

J. H. Davis, Schoorman, and Donaldson (1997) specify the factors that make each choice more likely. On the manager's side, **psychological factors** matter: stewards are more intrinsically motivated, identify more strongly with the organization, and rely on personal rather than institutional power. On the principal's side, **situational factors** matter: involvement-oriented management philosophies, collectivist cultures, and low power distance make stewardship relationships more likely.

6.2.3 Evidence and applications

Stewardship theory has been especially influential in research on family firms, nonprofit organizations, and boards, where identification with the organization is often high. Its prescriptions are

contingent rather than universal: it does not claim that all managers are stewards, only that governance should fit the type of manager and the situation. That contingency is both its strength and its main empirical challenge, because motivation and identification are difficult to observe at the top of large firms.

BUSINESS CASE

Board independence and the CEO-chair question. After the accounting scandals of 2001-2002, the Sarbanes-Oxley Act and new stock-exchange listing standards in the United States required majority-independent boards and fully independent audit committees, an explicitly agency-theoretic reform. Many U.K. and European firms had already separated the CEO and chair roles. Yet large U.S. companies have continued to differ in whether they combine the roles, and meta-analytic evidence finds no consistent performance advantage for either structure (Dalton et al. 1998). The stewardship perspective suggests why: the best structure depends on whether the executive is a steward and whether the board can tell. A practical compromise, a combined CEO-chair together with a strong lead independent director, tries to capture the unity-of-command benefits of stewardship while retaining an agency safeguard. Hillman and Dalziel (2003) add that independent directors provide not only monitoring but also **board capital**, advice, legitimacy (Section @ref(legitimacy (Section 7.3))), and links to external resources, which connects governance to resource dependence theory (Section 7.1).

COMMON PITFALL

Using stewardship theory as a post hoc label. Studies sometimes interpret any positive effect of executive autonomy as "support for stewardship theory". The theory makes a conditional claim: empowerment improves outcomes *when* managers are stewards. A credible test measures (or plausibly proxies) motivation and identification, for instance, founder or family status, tenure, or survey measures, and shows that the effect of governance structure depends on them.

6.3 Transaction Cost Economics

THEORY CARD

Core claim. Transactions differ in their attributes, above all in the specificity of the assets they require, and are assigned to governance structures (markets, hybrids, hierarchies) that economize on the costs of contracting, especially the hazards of opportunistic **hold-up**.

Origins. Economics, Coase (1937) asked why firms exist. Williamson (1975), Williamson (1979), Williamson (1985), and Williamson (1991) built the operational theory. Williamson (1981) introduced it to sociology and organization theory. Williamson shared the 2009 Nobel Memorial Prize in Economic Sciences, and Coase received the prize in 1991.

Level of analysis. The transaction is the unit of analysis. Outcomes are governance choices at the boundary of the firm.

Key constructs. Bounded rationality, opportunism, asset specificity, uncertainty, frequency, hold-up, appropriable quasi-rents, fundamental transformation, discriminating alignment, market, hybrid, and hierarchy.

Boundary conditions. Contracts are incomplete. Some parties are opportunistic some of the time. Courts cannot cheaply enforce all contingencies. Transacting parties can choose among governance alternatives.
Related theories. Agency theory, Resource-based theory (Section 4.1), Resource dependence (Section 7.1), Bounded rationality (Section 23.1).

6.3.1 Origins and intellectual history

Coase (1937) noticed a puzzle at the heart of economics. If markets allocate resources efficiently through prices, why do so many transactions take place inside firms, where they are directed by managerial authority instead? His answer was that using the price mechanism has costs: discovering the relevant prices, negotiating and concluding separate contracts for each exchange, and writing contracts for long periods under uncertainty. Firms exist because, for some transactions, authority is cheaper than repeated contracting. A firm grows until the cost of organizing one more transaction internally equals the cost of buying it in the market.

For decades the argument was admired but rarely used, because it did not say *which* transactions would be cheaper to organize internally. Williamson (1975) supplied the missing operational content. Drawing on Herbert Simon's concept of bounded rationality (Section 23.1) and on legal scholarship about contracts, Williamson argued that transaction costs arise from the combination of two behavioural assumptions and three attributes of transactions.

6.3.2 Core logic and propositions

The **behavioural assumptions** are bounded rationality and opportunism. **Bounded rationality** means that people intend to be rational but cannot foresee every contingency, so all complex contracts are incomplete. **Opportunism** means that at least some parties will pursue self-interest strategically, through selective disclosure, misrepresentation, or exploiting gaps in contracts, and that it is costly to know in advance who will (Williamson 1985). Neither assumption alone creates serious problems: with unbounded rationality, contracts could anticipate everything. Without opportunism, incomplete contracts could be filled in by good faith.

The **attributes of transactions** determine how much these behavioural limits matter (Williamson 1979, 1985):

- **Asset specificity**: the degree to which an investment made to support a transaction loses value in its next-best use. Williamson distinguished site specificity (a plant located next to a customer), physical asset specificity (custom dies and tooling), human asset specificity (relationship-specific know-how), and dedicated assets (capacity built for one buyer).
- **Uncertainty**: about the environment or about the other party's behaviour, which increases the number of contingencies a contract would need to cover.
- **Frequency**: how often the transaction recurs, which determines whether the fixed costs of specialised governance can be recovered.

The theory's key dynamic is the **fundamental transformation**. Before a specific investment is made, many suppliers may compete for a contract. Once one supplier has invested in assets that are valuable only in the relationship, the parties are locked into a bilateral relationship, and the large-numbers competition of the bidding stage becomes a small-numbers bargaining problem. Each side can then threaten to withhold cooperation to capture the **appropriable quasi-rents**: the difference

between an asset's value in the relationship and in its next-best use (B. Klein, Crawford, and Alchian 1978). This threat is the **hold-up problem**. Anticipating it, parties may under-invest in valuable specific assets, or they may choose a governance structure that protects those investments.

TCE's central principle is **discriminating alignment**: transactions, which differ in their attributes, are aligned with governance structures, which differ in their costs and competencies, in a way that economizes on transaction costs (Williamson 1991). The core propositions follow:

- **P1 (Asset specificity).** As asset specificity rises, governance shifts from markets toward hybrids (long-term contracts, alliances, franchising) and then to hierarchy (vertical integration).
- **P2 (Uncertainty).** Uncertainty amplifies the effect of asset specificity, because adapting an incomplete contract to unforeseen events invites haggling.
- **P3 (Frequency).** Recurrent transactions with high specificity justify the set-up costs of specialised governance. Occasional ones are more often governed by trilateral arrangements such as arbitration.
- **P4 (Remediableness).** A governance structure should be judged against feasible alternatives, not against an ideal that no organization can implement.

Williamson (1991) characterised the three generic governance structures by a small set of attributes (Table 6.6). Markets provide strong incentives and are good at **autonomous adaptation**, in which parties respond independently to price signals. Hierarchies provide weaker incentives but strong administrative controls and are good at **coordinated adaptation**, in which interdependent parties must adjust together. Hybrids sit in between.

Table 6.6. Attributes of generic governance structures (adapted from Williamson, 1991).

Attribute	Market	Hybrid	Hierarchy
Incentive intensity	Strong	Semi-strong	Weak
Administrative controls	Minimal	Moderate	Extensive
Autonomous adaptation (responding to prices)	Strong	Moderate	Weak
Coordinated (cooperative) adaptation	Weak	Moderate	Strong
Reliance on contract law and courts	Strong	Moderate	Weak (internal dispute resolution)
Typical business form	Spot purchase, competitive bidding	Long-term supply contract, alliance, franchise, joint venture	In-house production, wholly owned subsidiary

6.3.3 Mechanism

Figure 6.5 presents the make-or-buy logic. Transaction attributes create hazards, behavioural assumptions explain why those hazards cannot simply be contracted away, and governance choice follows.

Williamson's well-known heuristic representation expresses governance costs as a function of asset specificity. Markets have the lowest costs when assets are generic, but their costs rise steeply with specificity. Hierarchies carry high bureaucratic overhead but are least sensitive to specificity. Hybrids lie between. The efficient boundary of the firm follows the lower envelope of the three curves (Figure 6.6).

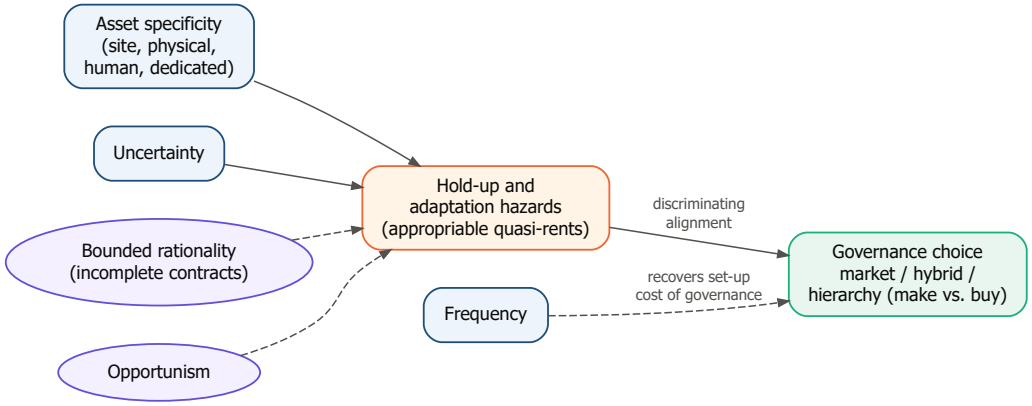


Figure 6.5. The make-or-buy mechanism in transaction cost economics: asset specificity, uncertainty, and frequency create hold-up and adaptation hazards, which bounded rationality and opportunism make costly to contract away. Governance is chosen to economize on those hazards.

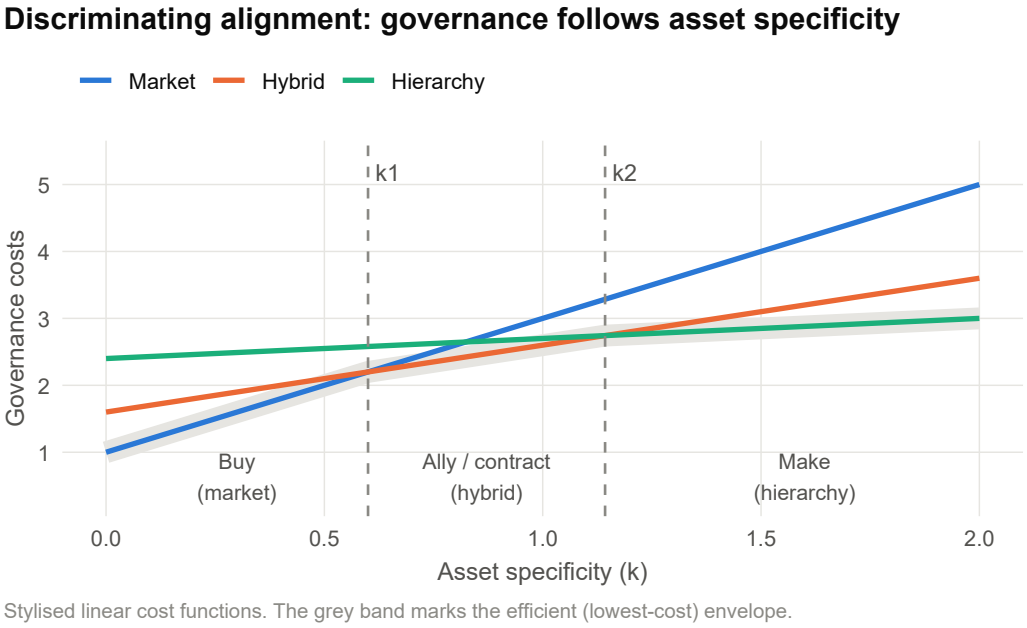


Figure 6.6. Stylised governance-cost curves after Williamson (1991). Each governance form is cheapest over a range of asset specificity. The efficient choice follows the lower envelope, switching from market to hybrid at k1 and from hybrid to hierarchy at k2. Functional forms are illustrative.

6.3.4 Evidence and extensions

TCE is one of the most extensively tested theories in the social sciences. Early studies of the automobile industry found that components requiring more specialised engineering know-how were more likely to be produced in-house (Monteverde and Teece 1982), and that make-or-buy decisions in a component division responded to supplier-market competition and volume uncertainty (Walker and Weber 1984). Reviews conclude that the evidence is broadly consistent with the theory, especially for asset specificity (Shelanski and Klein 1995), although a systematic assessment found the support uneven across constructs and noted disagreement over how key variables should be

measured (R. J. David and Han 2004).

DEBATE AND EVIDENCE

Did Fisher Body hold up General Motors? The most famous illustration of hold-up is the relationship between General Motors and Fisher Body, the supplier of closed car bodies. In 1919 GM acquired a majority stake in Fisher and signed a long-term exclusive supply contract with cost-plus pricing. In 1926 it acquired the rest of the company. B. Klein, Crawford, and Alchian (1978) argued that the contract left GM vulnerable once demand for closed bodies grew: Fisher allegedly exploited the cost-plus formula and resisted locating plants next to GM's assembly plants, and GM integrated fully to end the hold-up. The case became the canonical textbook example of TCE.

Seven decades after his 1937 article, Coase (2000) examined the historical record and rejected this account. He argued that relations between the two firms were close, that Fisher plants were in fact built near GM plants, and that the 1926 acquisition was motivated chiefly by GM's wish to secure the Fisher brothers' managerial talent and to coordinate production more closely. Two further studies in the same journal issue reached related conclusions: Casadesus-Masanell and Spulber (2000) called the hold-up story a "fable", and Freeland (2000) argued that, if anything, integration *created* hold-up potential by making GM dependent on the Fisher brothers' continued cooperation. B. Klein (2000) responded by defending a refined version of the hold-up interpretation.

What is at stake? Not the logic of hold-up, which no participant disputes, but the evidential standard for using historical cases to support theory. The debate is a reminder that canonical examples deserve the same scrutiny as statistical evidence, and that the same integration decision can be consistent with transaction-cost, capability, and coordination explanations.

TCE has also been criticised on conceptual grounds. Ghoshal and Moran (1996) argued that the assumption of opportunism can become self-fulfilling: governance designed for opportunists erodes trust and invites the behaviour it seeks to prevent, a critique parallel to the one levelled at agency theory. Capability-based theories offer a different answer to the make-or-buy question: firms integrate activities in which they have superior, hard-to-imitate capabilities and buy those in which others are more capable (Jay Barney 1991, Teece, Pisano, and Shuen 1997). Contemporary research often treats transaction-cost and capability logics as complements: specificity explains *why* an activity is hazardous to outsource, and capabilities explain *who* can perform it best.

6.3.5 Applications in business

TCE is a practical decision framework. Table 6.7 applies it to common boundary decisions.

Table 6.7. Transaction cost economics applied to boundary decisions across business functions.

Function	Decision	Key transaction attribute	TCE prediction
Operations	Outsource standard components or build them in-house?	Low physical asset specificity, many suppliers	Buy through competitive markets
Operations	Source a custom module requiring dedicated tooling	High physical and dedicated-asset specificity	Long-term contract with safeguards, or integrate
Marketing	Use independent distributors or a direct sales force?	Human asset specificity of product knowledge and customer relationships	Direct sales force when selling requires specialised, non-transferable knowledge

Table 6.7. Transaction cost economics applied to boundary decisions across business functions. *(continued)*

Function	Decision	Key transaction attribute	TCE prediction
IT	Move commodity workloads to a public cloud. Keep core data platforms?	Low specificity for commodity computing, high for proprietary data and processes	Buy commodity capacity. Govern core systems internally or through hybrid contracts
HR	Hire contractors or permanent employees?	Firm-specific know-how and tacit routines	Employment (hierarchy) for firm-specific roles
Strategy	Acquire a key supplier or form an alliance?	Specificity combined with high uncertainty	Hybrid when coordination needs are moderate, acquisition when hold-up hazards are severe

BUSINESS CASE

Outsourcing and re-integration in commercial aircraft. For the 787 Dreamliner, Boeing outsourced the design and manufacture of large structural sections to a global network of partners to a greater extent than in earlier programmes. The programme suffered extensive delays, and in 2009 Boeing acquired the South Carolina fuselage operations of one of its partners, Vought Aircraft Industries. In 2024 Boeing announced an agreement to reacquire Spirit AeroSystems, the fuselage maker it had spun off in 2005. Interpreted through TCE, large aircraft structures involve high physical and human asset specificity, substantial technological uncertainty, and intense need for coordinated adaptation: precisely the conditions under which market and hybrid governance are most exposed to hold-up and adaptation failures. Capability arguments point in the same direction: integration preserves architectural knowledge that may be lost when design work is dispersed.

BUSINESS CASE

Selective integration at Apple. Apple outsources most final assembly to contract manufacturers but has progressively brought chip *design* in-house, culminating in the transition of its Mac computers to Apple-designed processors beginning in 2020, while fabrication remains with external foundries. TCE explains the split: assembly is a large-scale but relatively codifiable activity that can be governed through long-term contracts, with the buyer at times investing in specialised equipment at suppliers' facilities to secure capacity. Chip architecture, by contrast, is deeply interdependent with Apple's software and product roadmap, making coordinated adaptation valuable and outsourcing hazardous.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The likelihood of in-house production of a component increases with the engineering effort specific to that component.
- *H2.* The effect of asset specificity on vertical integration is stronger under high technological uncertainty.
- *H3.* Alliances involving high relationship-specific investment are more likely to be structured as equity joint ventures than as contractual alliances.

Designs. Cross-sectional or panel analysis of make-or-buy choices at the transaction or com-

ponent level. Contract-level studies of duration, pricing terms, and safeguards. Event studies of supplier disruptions.
Measures. Asset specificity (engineer ratings, specialised tooling, co-location, relationship-specific training), uncertainty (demand volatility, technological change), frequency (order volume and repetition), governance (make/buy, contract length, equity sharing).

COMMON PITFALL

Measuring governance outcomes instead of transaction attributes. A frequent error is to infer asset specificity from the fact that a firm integrated, which makes the theory circular. Measure transaction attributes independently of the governance choice, compare against feasible alternatives (remediableness), and remember that the unit of analysis is the transaction, not the firm.

6.4 Upper Echelons Theory

THEORY CARD

Core claim. Organizational outcomes, strategic choices and performance, are partly predicted by the experiences, values, and personalities of top executives, because executives act on their *construed* reality rather than on objective situations.
Origins. Strategic management, Hambrick and Mason (1984), building on the Carnegie School's view of bounded rationality, updated in Hambrick (2007b), synthesised in Finkelstein, Hambrick, and Cannella (2009).
Level of analysis. Individual executive (especially the CEO), the top management team (TMT), and the firm.
Key constructs. Executive orientation (values, cognitive base, personality), observable characteristics (age, tenure, functional background, education), filtering (field of vision, selective perception, interpretation), managerial discretion, executive job demands, hubris and narcissism.
Boundary conditions. Executive characteristics matter more when managerial discretion is high and when job demands force executives to rely on mental shortcuts.
Related theories. Agency theory, Bounded rationality (Section 23.1), Heuristics and biases (Section 23.2), Glass cliff (Section 29.6).

6.4.1 Origins and core logic

For much of the twentieth century, organization theorists debated whether leaders matter. Population ecologists and institutional theorists emphasised environmental constraints, and a widely cited study found that leadership changes explained relatively little variance in company performance compared with industry and company factors (Lieberson and O'Connor 1972). Hambrick and Mason (1984) proposed a different starting point: since complex strategic decisions cannot be made through pure calculation, executives inevitably bring their own **cognitive base** (knowledge and assumptions) and **values** to bear. The organization therefore becomes, in Hambrick and Mason's phrase, a reflection of its top managers.

Two further moves made the idea researchable. First, because executives' cognitions are difficult to

measure directly, Hambrick and Mason proposed **observable demographic characteristics**, age, tenure, functional track, education, socioeconomic background, as imperfect proxies. Second, they argued that the **top management team**, not the CEO alone, is the appropriate unit, because strategy emerges from the shared and divergent views of a dominant coalition.

The theory's mechanism is a **filtering** process (Finkelstein, Hambrick, and Cannella 2009, Hambrick 2007b). An executive's orientation limits the field of vision (which parts of the environment are scanned), shapes selective perception (which stimuli within that field are noticed), and guides interpretation (what meaning is attached to them). The resulting construed reality drives strategic choices such as innovation, diversification, acquisitions, and leverage, which in turn affect performance (Figure 6.7).

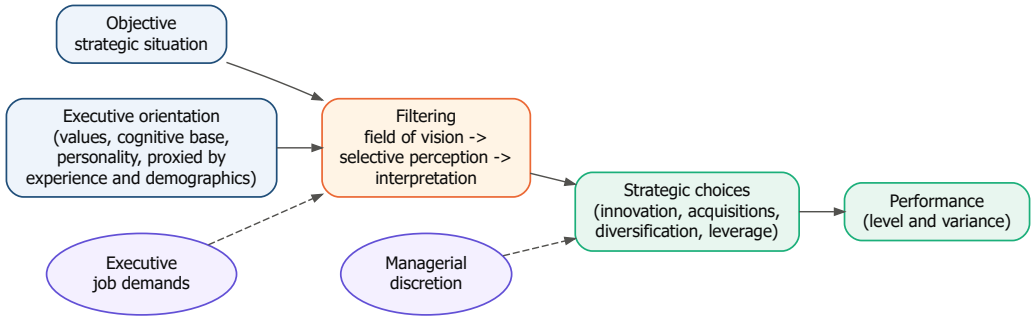


Figure 6.7. Upper echelons mechanism: executives' psychological orientation (proxied by observable experiences) filters the strategic situation into a construed reality that shapes strategic choices and performance. Managerial discretion and executive job demands moderate the strength of the link.

6.4.2 Moderators: discretion and job demands

Hambrick and Finkelstein (1987) introduced **managerial discretion**, latitude of action, to reconcile the view that leaders matter with the view that environments constrain them. Discretion is high in industries with product differentiability, rapid growth, and few regulatory constraints. In organizations with slack resources and weak internal resistance, and for executives with high tolerance for ambiguity and substantial power. Executive characteristics should predict outcomes much more strongly where discretion is high. Finkelstein and Hambrick (1990) provided early support: TMT tenure was associated with persistent strategies and strategies that conformed to industry averages, and these associations were stronger in high-discretion industries. Hambrick (2007b) added **executive job demands**: executives facing heavier demands rely more on experience and intuition, so their backgrounds should matter more.

Table 6.8 summarises common constructs and measures.

Table 6.8. Common upper echelons constructs, measures, and predictions (see Carpenter et al., 2004, Finkelstein et al., 2009).

Construct	Typical proxy or measure	Representative prediction
Tenure (organizational, CEO, team)	Years in firm or position	Longer tenure -> strategic persistence and conformity to industry norms
Functional background	Career track (output functions such as marketing vs. throughput functions such as operations)	Background shapes which problems and opportunities are noticed

Table 6.8. Common upper echelons constructs, measures, and predictions (see Carpenter et al., 2004, Finkelstein et al., 2009). *(continued)*

Construct	Typical proxy or measure	Representative prediction
Education	Level and type of degree	More formal education -> greater openness to innovation
Age	Executive or team mean age	Younger teams -> riskier strategies and more strategic change
TMT heterogeneity	Dispersion of backgrounds within the team	More diverse information but potentially less cohesion
Overconfidence / hubris	Option-exercise behaviour, media praise, relative pay	Larger acquisition premiums, investment sensitive to internal cash
Narcissism	Unobtrusive indicators (prominence in annual reports, first-person pronouns, relative pay)	Bolder, more dynamic strategies and more extreme performance

6.4.3 Evidence and extensions

Research on upper echelons has expanded from demographics to personality and psychology. In finance, Bertrand and Schoar (2003) tracked individual managers across firms and found that manager fixed effects explain a meaningful share of the variation in investment, financing, and organizational practices: strong evidence that "style" travels with the person. Quigley and Hambrick (2015) estimated that the share of performance variance attributable to CEOs has increased over recent decades, a partial answer to the early skepticism of Lieberson and O'Connor (1972).

The most vivid stream concerns **CEO hubris**. Roll (1986) proposed that bidding managers overestimate their ability to create value, so takeover premiums partly reflect errors of judgment. Hayward and Hambrick (1997) found that premiums paid for large acquisitions were higher when the acquiring firm had recently performed well, when the CEO had received favourable media praise, and when the CEO's pay was high relative to other executives, indicators of hubris, and that these effects were stronger when board vigilance was weak. Shareholder losses followed. Malmendier and Tate (2005) measured overconfidence by CEOs' reluctance to exercise in-the-money options and found that overconfident CEOs' investment was more sensitive to internal cash flow. Chatterjee and Hambrick (2007) showed that narcissistic CEOs pursue more dynamic and grandiose strategies and generate more extreme and fluctuating performance, without being systematically better or worse.

DEBATE AND EVIDENCE

Do demographics measure cognition? The theory's early reliance on demographic proxies is both its engine and its Achilles' heel. Critics note that the same demographic variable (say, tenure) can be linked to several psychological states, so demographic studies cannot identify which mechanism is at work. Hambrick (2007b) himself called for direct study of executive psychology and of the processes that link characteristics to choices. A second debate concerns **endogeneity**: boards select executives whose backgrounds fit the strategy they want, so an association between, for example, a CEO's finance background and a diversification strategy may reflect selection rather than influence. Executive successions caused by deaths or illnesses, and designs that follow managers across firms (Bertrand and Schoar 2003), are among the tools used to address this problem.

BUSINESS CASE

Acquisition premiums and board vigilance. When a celebrated CEO proposes a large acquisition at a substantial premium, upper echelons research suggests that the board should ask whether the premium reflects analysis or confidence born of recent success and favourable press. The governance implications link this theory to agency theory: hubris is not opportunism, the CEO may sincerely believe the deal will create value, but the remedies overlap. Independent valuation advice, a board that includes directors with acquisition experience, staged commitments, and compensation tied to post-deal performance rather than deal size are all ways of filtering executive overconfidence before it becomes residual loss.

BUSINESS CASE

Top-team composition in a strategic pivot. Illustrative scenario: a hardware-centred firm wants to move toward software subscriptions. Upper echelons theory predicts that a TMT drawn mainly from engineering and manufacturing will notice threats and opportunities in product reliability and cost, and may under-weight signals about recurring revenue and customer success. Firms facing such pivots often deliberately change TMT composition, appointing executives from the target business model, not only for their skills but because they change what the team notices and how it interprets ambiguous evidence.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The positive relationship between CEO overconfidence and acquisition premiums is weaker when the board includes more directors with prior acquisition experience.
- *H2.* TMT functional heterogeneity is positively related to strategic change, and this effect is stronger in high-discretion industries.
- *H3.* CEO narcissism increases the variance, but not the mean, of firm performance.

Designs. Panel data with executive and firm fixed effects. Manager-mobility designs, exogenous CEO turnover. Text analysis of letters to shareholders and earnings calls. Surveys of executives (with low response rates as a known constraint).

Measures. Demographics from proxy statements and biographies. Psychological traits from validated unobtrusive indicators or linguistic analysis. Discretion indices at the industry level.

COMMON PITFALL

Ignoring discretion and selection. Upper echelons effects are contingent. Testing them in heavily regulated, low-discretion contexts and finding null effects says little about the theory, and finding an association without addressing how executives were selected says little about influence. Model discretion explicitly and use designs that separate the effect of the executive from the choice of the executive.

6.5 Chapter Summary

KEY TAKEAWAYS

- **Agency theory** explains governance as a response to goal conflict and information asymmetry. Agency costs comprise monitoring, bonding, and residual loss (Jensen and Meckling 1976, Eisenhardt 1989a).
- Efficient incentives balance effort against risk: noisier measures and more risk-averse agents call for flatter pay-for-performance slopes, and strong incentives on measurable tasks can crowd out unmeasured ones (Holmström 1979, Holmström and Milgrom 1991).
- Outcome-based and behaviour-based contracts are substitutes whose relative efficiency depends on information systems, uncertainty, task programmability, and outcome measurability (Eisenhardt 1989a).
- **Stewardship theory** argues that governance should fit the manager: empowering structures outperform controls when both principal and manager choose a stewardship relationship (J. H. Davis, Schoorman, and Donaldson 1997).
- **Transaction cost economics** explains the boundaries of the firm: asset specificity, uncertainty, and frequency, combined with bounded rationality and opportunism, determine whether markets, hybrids, or hierarchies economize on hold-up and adaptation costs (Coase 1937, Williamson 1985, 1991).
- Canonical cases such as GM-Fisher Body remain contested (B. Klein, Crawford, and Alchian 1978, Coase 2000), a reminder to scrutinise illustrative evidence.
- **Upper echelons theory** shows that executives' experiences, values, and personalities shape strategy, especially when managerial discretion is high. Hubris and narcissism are its most consequential applications (Hambrick and Mason 1984, Hambrick 2007b, Hayward and Hambrick 1997).

Table 6.9 presents the comparison of the chapter's theories.

Table 6.9. Comparison of the chapter's theories.

Theory	Unit of analysis	View of the manager	Key mechanism	Signature prediction	Typical business use
Agency theory	Principal-agent contract	Self-interested, risk-averse agent	Incentive alignment and monitoring under information asymmetry	Outcome-based pay where outcomes are measurable, residual loss where they are not	Executive pay, sales compensation, franchising, board design
Stewardship theory	Principal-manager relationship	Pro-organizational steward (when chosen)	Identification, intrinsic motivation, trust	Empowerment outperforms control under mutual stewardship	Leadership structure, family firms, nonprofit governance

Table 6.9. Comparison of the chapter's theories. *(continued)*

Theory	Unit of analysis	View of the manager	Key mechanism	Signature prediction	Typical business use
Transaction cost economics	Transaction	Boundedly rational and potentially opportunistic	Hold-up hazards from asset specificity	Integration rises with asset specificity and uncertainty	Make-or-buy, outsourcing, alliances, contract design
Upper echelons theory	Executive and top management team	Boundedly rational person with a distinctive orientation	Filtering of information through values and experience	Executive traits predict strategy, especially under high discretion	Executive selection, TMT composition, M&A oversight

6.6 Discussion Questions

1. Agency theory and stewardship theory make opposite predictions about combining the CEO and board-chair roles. Design a study that could distinguish between them. What moderator would you measure, and how?
2. Using the linear incentive model in this chapter, explain why a hospital might pay surgeons largely by salary while a car dealership pays salespeople largely by commission. What other considerations (multitasking, ethics, teamwork) does the model omit?
3. The Wells Fargo case is often described as a failure of culture rather than of incentives. Is that distinction meaningful from an agency perspective? What evidence would help adjudicate between the two accounts?
4. Choose a firm that recently insourced an activity it had previously outsourced (or vice versa). Analyse the decision using asset specificity, uncertainty, and frequency. Would a capability-based explanation reach the same conclusion?
5. The GM-Fisher Body debate concerns the use of historical evidence. What standards should a historical case meet before it is used as support for a theory?
6. A board is evaluating a CEO's proposal for a large acquisition. Drawing on upper echelons research and agency theory, list the questions directors should ask and the governance safeguards they should consider.

6.7 Further Reading

- Jensen and Meckling (1976): the foundational statement of agency costs and the firm as a nexus of contracts.
- Eisenhardt (1989a): the clearest introduction to agency theory for management researchers, with testable propositions on contract choice.
- Holmström and Milgrom (1991): the multitask model. Essential for understanding why strong incentives can backfire.
- J. H. Davis, Schoorman, and Donaldson (1997): the theoretical development of stewardship

theory and the principal-manager choice model.

- Williamson (1991): a compact statement of discriminating alignment and the market-hybrid-hierarchy comparison.
- Coase (2000) and B. Klein (2000): read together, an instructive exchange on the historical evidence for hold-up.
- Hambrick (2007b): a short, candid update of upper echelons theory with an agenda for future research.
- Hayward and Hambrick (1997): a model empirical study linking executive psychology, governance, and acquisition outcomes.

Chapter 7

Organizations and Their Environments

Why did hospitals, universities, and city governments across the United States adopt nearly identical administrative structures in the twentieth century, even when those structures did little to improve their work? Why do thousands of suppliers seek the same quality and environmental certifications, why do software developers grumble about app-store commissions yet keep building for the same two platforms, and why did the chief executives of nearly two hundred of America's largest companies publicly redefine the purpose of the corporation in 2019? None of these questions can be answered by looking inside the firm alone. Each concerns the relationship between an organization and the **environment** on which it depends: suppliers, customers, regulators, professions, competitors, and the wider society that grants or withholds approval.

This chapter presents the major "open-systems" theories of organizations. **Resource dependence theory** treats the environment as a set of other organizations that control critical resources, and organizational action as the management of the resulting power relations. **Institutional theory** treats the environment as a set of rules, norms, and taken-for-granted beliefs, and explains why organizations conform to them, often at the expense of efficiency. **Legitimacy** is the currency of that institutional environment. **Stakeholder theory** asks whose interests the organization should and does serve. **Organizational ecology** shifts the focus from individual adaptation to the selection of whole populations of organizations. **Contingency theory**, finally, argues that there is no one best way to organize: effective structures fit their environments (Table 7.1).

Table 7.1. Chapter at a glance: theories of organizations and their environments.

Theory	Core question	Level of analysis	Key source
Resource dependence theory	How do organizations manage dependence on others for critical resources?	Organization, dyad, organization set	Pfeffer & Salancik (1978)
Institutional theory	Why do organizations in a field become similar, and why do they adopt practices that may not improve efficiency?	Organizational field, organization	Meyer & Rowan (1977), DiMaggio & Powell (1983)
Legitimacy	How do organizations gain, maintain, and repair social acceptance?	Organization, audience	Suchman (1995), Deephouse (1999)
Stakeholder theory	Whose interests should and do organizations serve?	Organization, stakeholder relationships	Freeman (1984), Donaldson & Preston (1995)

Table 7.1. Chapter at a glance: theories of organizations and their environments. *(continued)*

Theory	Core question	Level of analysis	Key source
Organizational ecology	Why do populations of organizations change through founding and failure rather than adaptation?	Population, community	Hannan & Freeman (1977, 1984)
Contingency theory	Which organizational structures fit which environments?	Organization, subunit	Burns & Stalker (1961), Lawrence & Lorsch (1967)

7.1 Resource Dependence Theory

THEORY CARD

Core claim. Organizations are not self-sufficient. They depend on other actors for critical resources. Dependence creates power for those who control the resources, and organizations act, through co-optation, alliances, mergers, political action, and leadership choices, to reduce their own dependence and increase others' dependence on them.
Origins. Organizational sociology. Pfeffer and Salancik (1978, reviewed in Stern 1979), building on the power-dependence theory of Emerson (1962) and the analysis of co-optation by Selznick (1949). Reviewed by Hillman, Withers, and Collins (2009).
Level of analysis. The focal organization and its exchange partners (dyads and the organization set).
Key constructs. Resource importance, discretion over resources, concentration of control (availability of alternatives), dependence, power, power imbalance, mutual dependence, constraint absorption, co-optation.
Boundary conditions. Resources must be critical and not freely available. The focal organization must have some strategic discretion to restructure its dependencies.
Related theories. Social exchange theory (Section 9.1), Network exchange theory (Section 9.2), Agency theory (Section 6.1), Transaction cost economics (Section 6.3), Institutional theory.

7.1.1 Origins and intellectual history

The basic premise of resource dependence theory (RDT) is that external resources, and the actors who control them, shape organizational behaviour. Organizations must acquire inputs such as capital, materials, labour, information, customers, and legitimacy from an environment populated by other organizations. Because resources are power and no organization can survive alone, **power and resource dependence are interlocked** (Pfeffer and Salancik 1978).

The theory's micro-foundation comes from Emerson (1962), who defined power in relational terms: the power of actor A over actor B is equal to the dependence of B on A. B's dependence, in turn, increases with B's motivational investment in goals that A mediates and decreases with the availability of those goals outside the A-B relationship. Two implications follow. First, **power is relational**: it is not an attribute of an actor but a property of a relationship, so a firm can be powerful in one relationship and weak in another. Second, **power is situational**: it changes when the importance of a resource or the availability of alternatives changes, so yesterday's dominant buyer can become today's dependent partner.

Pfeffer and Salancik (1978) translated this into a theory of organizations. The dependence of a focal organization on another actor depends on three factors: the **importance** of the resource (how much of it the organization needs and how critical it is to operations), the other actor's **discretion** over the resource's allocation and use, and the **concentration** of control over the resource (the extent to which alternative sources exist). Organizations facing such dependence experience **external constraint**, and managers' central task is to manage it.

7.1.2 Core logic and propositions

RDT can be summarised in four propositions:

- **P1 (Dependence and power).** The more an organization depends on another for important, discretionary, and concentrated resources, the more power the other has over it (Emerson 1962, Pfeffer and Salancik 1978).
- **P2 (Compliance).** Organizations are more likely to comply with the demands of actors on whom they depend.
- **P3 (Restructuring).** Organizations attempt to reduce dependence and uncertainty by restructuring their exchange relationships: absorbing, co-opting, or diversifying away from powerful actors.
- **P4 (Internal power).** Subunits and executives who manage the organization's most critical external dependencies gain power inside the organization.

Emerson (1962) identified four generic **balancing operations** by which a weaker actor can rebalance a dependent relationship: reduce its motivational investment in what the stronger actor provides (withdrawal). Cultivate alternative sources (extension of the power network). Increase the stronger actor's dependence on it (for example, by offering status or unique value), and form coalitions that deny the stronger actor its alternatives. Pfeffer and Salancik (1978) and later research mapped these onto organizational tactics (Table 7.2).

Table 7.2. Tactics for managing resource dependence.

Tactic	How it manages dependence	Business example	Key evidence
Co-optation	Brings representatives of a constraining actor into the organization's decision structure	Appointing a banker, regulator, or major customer's executive to the board	Selznick (1949), Pfeffer (1972)
Board interlocks	Provide information, advice, access, and legitimacy, signal commitment	Directors who sit on boards of suppliers or financial institutions	Pfeffer (1972), Hillman & Dalziel (2003)
Alliances and joint ventures	Partially absorb interdependence while preserving autonomy	Joint development agreements, long-term supply partnerships	Hillman et al. (2009)
Mergers and acquisitions	Absorb the constraint completely (vertical, horizontal, or diversifying)	Acquiring a critical supplier, diversifying away from a dominant customer	Casciaro & Piskorski (2005)
Political action	Shapes the rules and resources that government controls	Lobbying, political contributions, participation in standard-setting	Hillman et al. (2009)
Executive succession	Aligns leadership with the organization's most critical dependencies	Appointing a CEO with expertise in the firm's binding constraint	Pfeffer & Salancik (1978)
Diversifying sources	Reduces concentration of resource control	Dual or multiple sourcing, multi-homing across platforms or channels	Emerson (1962)

7.1.3 Mechanism

Figure 7.1 presents the core causal chain, from the characteristics of resources to dependence, power, and organizational response.

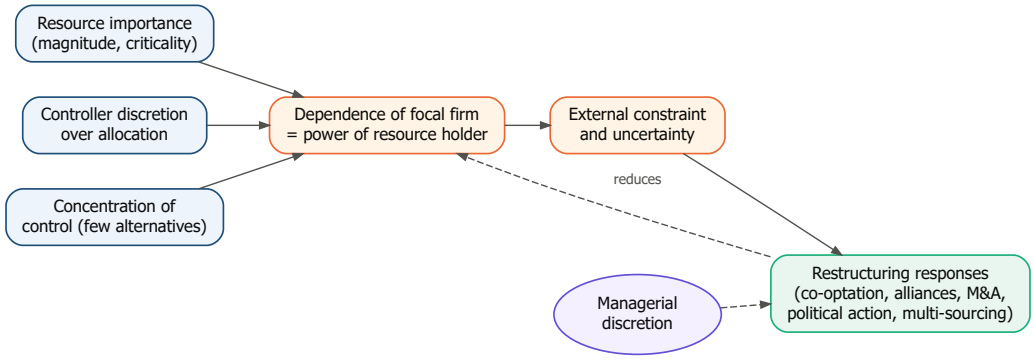


Figure 7.1. Mechanism of resource dependence theory: resource importance, the controller's discretion, and the concentration of control create dependence, which gives power to resource holders and constrains the focal organization. The organization responds by restructuring its dependencies, with managerial discretion moderating its options.

7.1.4 Power imbalance versus mutual dependence

A major theoretical refinement came from Casciaro and Piskorski (2005). In any dyad, each party depends on the other to some degree. Pfeffer and Salancik's concept of "interdependence" combined two distinct quantities: **power imbalance**, the difference between the two parties' dependence on each other, and **mutual dependence**, the total dependence in the relationship. Casciaro and Piskorski argued that these have opposite effects on **constraint absorption**, the use of mergers and acquisitions to eliminate a dependence. Mutual dependence makes absorption more attractive to both parties, because both stand to gain from removing the constraint. Power imbalance makes it less likely, because the more powerful party has little reason to give up its advantage and can resist. Their analysis of acquisitions across U.S. industries supported both predictions.

Figure 7.2 illustrates the logic. Moving up the diagonal (both parties more dependent) raises the likelihood of absorption. Moving toward the corners (one party much more dependent than the other) lowers it.

7.1.5 Evidence and extensions

RDT is among the most-cited theories of organizations, and its evidence base is broad. Pfeffer (1972) found that board size and the composition of boards varied with firms' need to manage their environments, including their reliance on external capital. Hillman, Withers, and Collins (2009) reviewed research in five areas, mergers and vertical integration, joint ventures and other interorganizational relationships, boards of directors, political action, and executive succession, and concluded that the theory's predictions are broadly supported, though often tested with loosely specified measures of dependence. A meta-analysis by Drees and Heugens (2013) found that resource dependence is associated with the formation of interorganizational arrangements, and that such arrangements are linked to both organizational autonomy and legitimacy. Hillman and Dalziel (2003) integrated RDT with agency theory by showing that boards both monitor management and provide

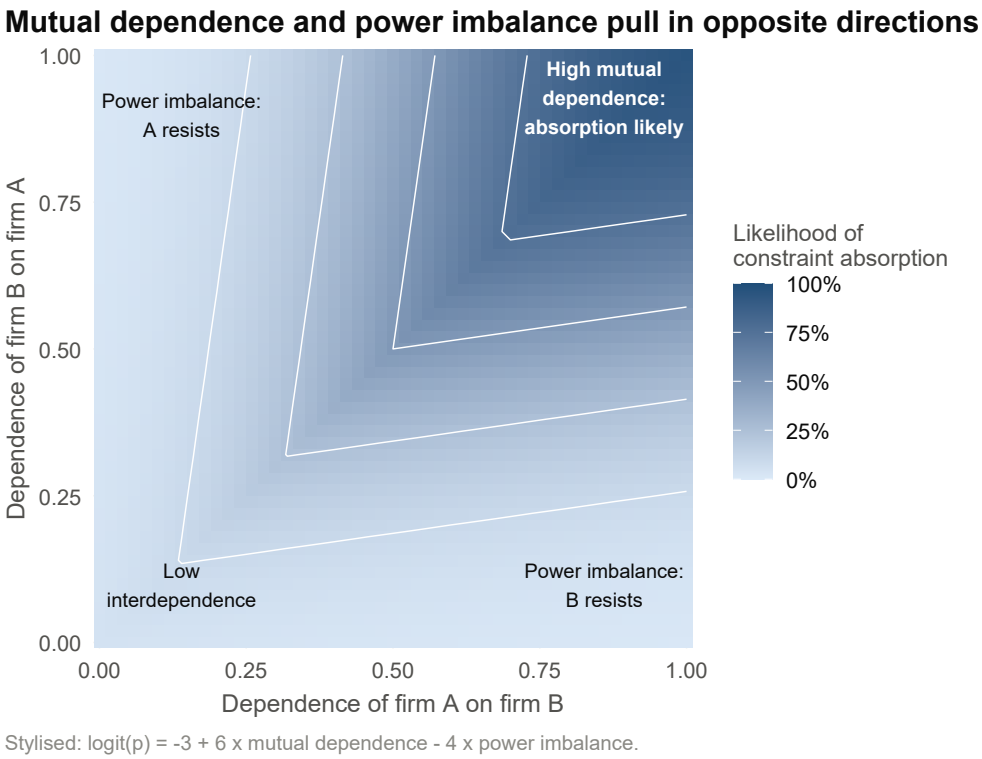


Figure 7.2. Illustrative surface showing the distinct effects of mutual dependence and power imbalance on the likelihood of constraint absorption (e.g., merger). Probability rises with mutual dependence (along the diagonal) and falls with power imbalance (toward the off-diagonal corners), following the argument of Casciaro and Piskorski (2005). Values are stylised.

board capital: the human and social capital (Section 9.5) of directors that supplies resources (see Stewardship theory (Section 6.2)).

DEBATE AND EVIDENCE

Is RDT a theory or a metaphor? RDT's influence rests partly on its intuitive appeal: everyone recognises that dependence confers power. Critics note that the theory was for many years used more as a general orienting framework than as a source of precise, falsifiable predictions, and that "dependence" was often measured with crude proxies such as industry-level input-output flows. The refinement by Casciaro and Piskorski (2005) shows how much sharper predictions become when constructs are disaggregated. A second debate concerns RDT's relationship to transaction cost economics (Section 6.3): both predict that firms absorb problematic exchange partners, but TCE emphasises efficiency (economizing on hold-up) while RDT emphasises power and autonomy. Distinguishing them empirically requires measures that separate dependence from asset specificity.

7.1.6 Applications in business

BUSINESS CASE

Supplier dependence and dual sourcing. Single sourcing reduces purchasing costs and deepens supplier relationships, but it concentrates control of a critical resource. Two widely reported supply disruptions illustrate the trade-off. In February 1997, a fire at an Aisin Seiki plant that was the main source of a brake-system valve used across Toyota's vehicles in Japan halted much of Toyota's domestic assembly. Production resumed within days because a network of group and non-group suppliers rapidly improvised replacement production: a case of dependence buffered by dense relational ties. In March 2000, a fire at a Philips semiconductor plant in New Mexico affected two mobile-phone makers differently: Nokia, which moved quickly to secure alternative supply, was much less affected than Ericsson, which had relied on the plant as a single source for key components. In RDT terms, multiple sourcing reduces the **concentration** of resource control, and strong relationships increase the supplier's **dependence** on the buyer (and thus its willingness to prioritise the buyer in a crisis). Many firms now combine both: a primary partner with deep integration and a qualified second source.

BUSINESS CASE

Platform dependence of app developers. Developers of mobile applications depend on two app-store operators for access to most smartphone users, for payment processing, and for discoverability: resources that are important, concentrated, and subject to the platform's discretion through review guidelines and commission rates. Cutolo and Kenney (2021) describe such firms as **platform-dependent entrepreneurs** who face pronounced power asymmetries. Platforms can also enter complementors' markets: F. Zhu and Liu (2018) found that Amazon was more likely to enter product spaces in which third-party sellers had demonstrated strong demand. Developers respond with balancing operations predicted by Emerson: multi-homing across platforms (extending the power network), building direct relationships with customers through web sales and subscriptions (reducing the platform's discretion), and forming coalitions and pursuing litigation and regulation. Epic Games' 2020 lawsuits against Apple and Google, filed after its game *Fortnite* was removed from both stores for offering direct payments, and the European Union's Digital Markets Act are prominent examples of the coalition and political-action routes.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Mutual dependence between two firms increases the likelihood that they form an equity alliance or merge. Power imbalance decreases it.
- *H2.* Firms that depend more heavily on a single customer are more likely to appoint directors with ties to that customer's industry.
- *H3.* Complementors that multi-home across platforms experience smaller performance declines after adverse platform policy changes.

Designs. Dyadic panel data on interorganizational ties (alliances, interlocks, supply relationships). Event studies of exogenous changes in dependence (supplier disasters, platform policy changes). Network analysis of organization sets.

Measures. Dependence (share of sales or purchases with a partner, input-output coefficients,

availability of alternatives). Power imbalance and mutual dependence (Casciaro and Piskorski 2005). Tactics (board appointments, alliances, acquisitions, lobbying expenditures).

COMMON PITFALL

Measuring dependence without measuring alternatives. Dependence is not the same as transaction volume. A firm that buys most of its inputs from one supplier is not dependent if it could switch cheaply. Measure importance, discretion, *and* the availability of alternatives, and remember that dependence is two-sided: always ask how dependent the partner is on the focal firm.

7.2 Institutional Theory

THEORY CARD

Core claim. Organizations conform to the rules, norms, and taken-for-granted beliefs of their institutional environments because conformity brings legitimacy, resources, and survival, even when it does not improve technical efficiency. As a result, organizations in the same field become structurally similar.
Origins. Organizational sociology, J. W. Meyer and Rowan (1977), DiMaggio and Powell (1983), synthesised in Scott (2014), strategic responses by C. Oliver (1991).
Level of analysis. Organizational field (all organizations that constitute a recognized area of institutional life) and the organization.
Key constructs. Institution, organizational field, isomorphism (coercive, mimetic, normative), regulative, normative, and cultural-cognitive pillars, decoupling, legitimacy, strategic response.
Boundary conditions. Strongest where technical performance is hard to measure, goals are ambiguous, and organizations depend on external audiences for resources. Weaker in highly competitive markets with clear performance criteria.
Related theories. Legitimacy, Resource dependence, Social norms (Section 22.5), Conformity (Section 22.4), Diffusion of innovations (Section 10.1).

7.2.1 Origins and core logic

J. W. Meyer and Rowan (1977) challenged the prevailing view that formal organizational structures exist to coordinate and control technical work. Many structures, positions, policies, programmes, and procedures, instead reflect **rationalized institutional myths**: widely shared beliefs that certain structures are the proper, rational way to organize. Organizations that incorporate these elements gain legitimacy and resources, irrespective of whether the elements improve efficiency. Because institutional rules may conflict with technical demands, organizations often **decouple** formal structure from actual practice, maintaining ceremonial conformity while work proceeds on a different basis, and relying on a logic of confidence and good faith that discourages close inspection.

DiMaggio and Powell (1983) asked why organizations are so similar. Their answer was **institutional isomorphism**: once an **organizational field** is established, suppliers, consumers, regulators, and competitors that constitute a recognized area of activity, powerful forces lead its members to become more alike. They identified three mechanisms (Table 7.3).

Table 7.3. Three mechanisms of institutional isomorphism (DiMaggio & Powell, 1983).

Mechanism	Source	Driving condition	Business example
Coercive isomorphism	Formal and informal pressure from organizations on which the focal organization depends, and from cultural expectations	Resource dependence, legal and regulatory requirements	Suppliers adopting a buyer's mandated codes of conduct. Firms adopting compliance functions required by law
Mimetic isomorphism	Imitation of organizations perceived as successful or legitimate	Uncertainty, ambiguous goals, poorly understood technologies	Copying a leading rival's organizational structure, AI strategy, or market entry
Normative isomorphism	Professionalization: shared education, professional networks, and personnel movement	Reliance on credentials, strong professional associations	Similar HR, marketing, and finance practices spread by MBA programmes, consultancies, and professional bodies

DiMaggio and Powell also offered testable hypotheses. Among them: the greater an organization's dependence on another organization, and the more centralized its resource supply, the more similar it will become to that organization. The more uncertain the relationship between means and ends, and the more ambiguous an organization's goals, the more it will model itself on organizations it perceives as successful, and the greater the reliance on academic credentials and the participation of managers in professional associations, the more similar organizations will become.

7.2.2 Mechanism

Figure 7.3 shows how the three mechanisms produce isomorphism, legitimacy, and survival, and how decoupling can separate adopted structures from practice.

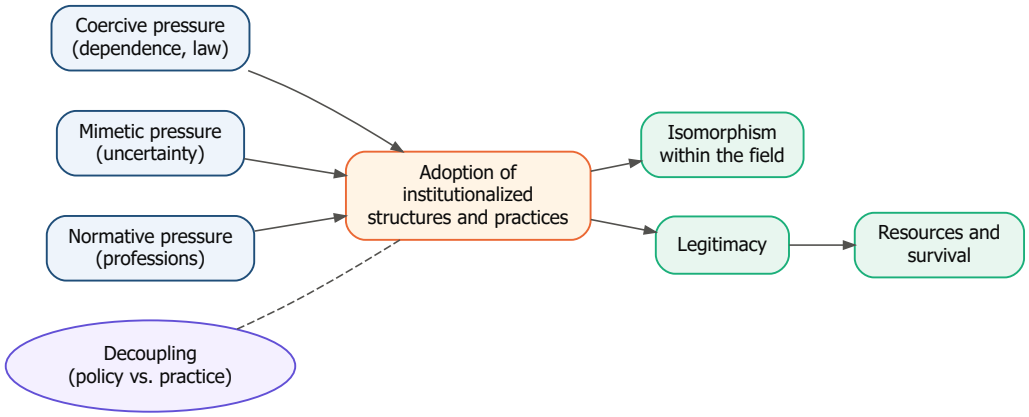


Figure 7.3. Mechanism of institutional theory: coercive, mimetic, and normative pressures lead organizations to adopt institutionalized structures, producing isomorphism and legitimacy and, through legitimacy, access to resources and survival. When institutional and technical demands conflict, adoption may be decoupled from practice.

7.2.3 Scott's three pillars

Scott (2014) synthesized institutional theory by distinguishing three **pillars** of institutions, each with a different basis of compliance and legitimacy (Table 7.4). The **regulative** pillar emphasises

rules, monitoring, and sanctions. Actors comply out of expedience. The **normative** pillar emphasises values and obligations. Actors comply because it is appropriate. The **cultural-cognitive** pillar emphasises shared conceptions of reality. Actors comply because alternatives are literally unthinkable. Most real institutions combine all three, but the pillars help researchers specify which mechanism they are studying.

Table 7.4. Three pillars of institutions (adapted from Scott, 2014).

Dimension	Regulative	Normative	Cultural-cognitive
Basis of compliance	Expedience	Social obligation	Taken-for-grantedness, shared understanding
Basis of order	Regulative rules	Binding expectations	Constitutive schemas
Mechanism	Coercive	Normative	Mimetic
Logic	Instrumentality	Appropriateness	Orthodoxy
Indicators	Rules, laws, sanctions	Certification, accreditation	Common beliefs, shared logics of action, isomorphism
Basis of legitimacy	Legally sanctioned	Morally governed	Comprehensible, recognizable, culturally supported
Business example	Data-protection law, financial reporting standards	Professional codes of ethics, industry certification	The assumption that a firm needs a website, a CFO, and a strategy

7.2.4 Evidence and extensions

The classic empirical demonstration is Tolbert and Zucker (1983)'s study of civil-service reform in U.S. cities. Early adoption was predicted by city characteristics that made the reform technically useful, whereas later adoption was not: once the reform had become legitimated, cities adopted it because it was the accepted way to organize. Where state law mandated the reform, adoption was immediate. The same two-stage pattern, early adoption for technical reasons, later adoption for legitimacy, has been documented for total quality management, where Westphal, Gulati, and Shortell (1997) found that early adopters customized practices to gain efficiency while later adopters conformed to a normative template that conferred legitimacy but fewer efficiency gains. Haveman (1993) showed mimetic isomorphism in market entry: savings-and-loan associations followed large and profitable peers into new markets.

Research has also moved beyond pure conformity. C. Oliver (1991) integrated institutional and resource dependence perspectives by arguing that organizations choose among a range of **strategic responses** to institutional pressures, from passive acquiescence to active manipulation (Table 7.5). Which response an organization chooses depends on why the pressure is exerted, who exerts it, what the pressure is, how it is exerted, and the context in which it occurs. Bromley and Powell (2012) distinguished **policy-practice decoupling** (policies adopted but not implemented) from **means-ends decoupling** (policies implemented but loosely connected to the outcomes they are meant to achieve), arguing that the latter has become more common as audit and evaluation pressures make pure ceremonial adoption harder to sustain.

Table 7.5. Strategic responses to institutional processes (Oliver, 1991).

Strategy	Tactics	Business example
Acquiesce	Habit (follow taken-for-granted norms), imitate (mimic institutional models), comply (obey rules)	Adopting a mandated disclosure format without debate

Table 7.5. Strategic responses to institutional processes (Oliver, 1991). *(continued)*

Strategy	Tactics	Business example
Compromise	Balance (reconcile multiple constituents), pacify (placate), bargain (negotiate with stakeholders)	Negotiating phased compliance timelines with a regulator
Avoid	Conceal (disguise nonconformity), buffer (loosen attachment), escape (change goals or domain)	Relocating an activity to a less regulated jurisdiction
Defy	Dismiss (ignore), challenge (contest), attack (assault the source)	Publicly contesting a proposed industry standard
Manipulate	Co-opt (import influential constituents), influence (shape values and criteria), control (dominate)	Leading an industry consortium that writes the standard

7.2.5 Applications in business

BUSINESS CASE

ISO and ESG certification as isomorphism. Management-system certifications such as ISO 9000 (quality) and ISO 14001 (environmental management) spread across tens of thousands of firms worldwide within a few decades. Guler, Guillén, and Macpherson (2002) showed that the international spread of ISO 9000 reflected coercive pressures from states and multinational buyers as well as normative and mimetic pressures transmitted through trade networks among countries. A. A. King, Lenox, and Terlaak (2005) found that firms used ISO 14001 certification in part to communicate unobservable environmental practices to distant buyers: an institutional mechanism that also performs a signaling (Section 8.2) function. The same dynamics are visible in environmental, social, and governance (ESG) reporting frameworks, sustainability ratings, and certifications such as B Corp status. Institutional theory supplies two practical warnings. First, adoption driven by legitimacy rather than technical need is vulnerable to decoupling, so audiences increasingly demand evidence of implementation (Bromley and Powell 2012). Second, when nearly everyone in a field is certified, certification ceases to differentiate and becomes a cost of entry.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Late adopters of a management practice will implement it with less customization than early adopters.
- *H2.* Suppliers whose largest customers require a certification will adopt it earlier (coercive isomorphism), and suppliers in uncertain markets will adopt it after observing peers (mimetic isomorphism).
- *H3.* Policy-practice decoupling will be lower when third-party audits are frequent and their results public.

Designs. Event-history models of practice adoption. Diffusion models with network or spatial contagion. Comparative case studies of fields. Text analysis of organizational documents to measure field-level logics.

Measures. Adoption timing, prevalence in the field, ties to adopters, professionalization (credentials, association membership), decoupling (gap between stated policy and observed practice or outcomes).

COMMON PITFALL

Calling every diffusion process "institutional". Practices also spread because they work, because of competitive imitation, or because of social learning about quality. To claim an institutional explanation, show that adoption is driven by legitimacy concerns rather than expected efficiency: for example, by showing that predictors of adoption shift from technical to social over time (Tolbert and Zucker 1983), or that adoption is decoupled from implementation.

7.3 Legitimacy

THEORY CARD

Core claim. Organizations need not only resources but social acceptance. Legitimacy is the generalized perception that an organization's actions are desirable, proper, or appropriate, and organizations act to gain, maintain, and repair it.
Origins. Institutional and strategic perspectives, Suchman (1995), Aldrich and Fiol (1994) on new industries, Deephouse (1999) on strategic balance.
Level of analysis. Organization or industry as evaluated by audiences.
Key constructs. Pragmatic, moral, and cognitive legitimacy, legitimacy management (gaining, maintaining, repairing), strategic balance, optimal distinctiveness.
Related theories. Institutional theory, Optimal distinctiveness (Section 22.3), Stakeholder theory, Signaling theory (Section 8.2).

7.3.1 Core logic

Suchman (1995) provided the standard definition: legitimacy is "a generalized perception or assumption that the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions" (Suchman 1995, 574). He also reconciled two traditions. The **strategic** approach treats legitimacy as an operational resource that managers can extract from the environment through symbolic and substantive action. The **institutional** approach treats legitimacy as a set of constitutive beliefs that shape organizations from outside. Both views are useful, and managers face the tension between them: legitimacy can be managed, but overt efforts to manage it can themselves look illegitimate.

Suchman distinguished three types of legitimacy (Table 7.6).

Table 7.6. Three types of organizational legitimacy (Suchman, 1995).

Type	Basis of evaluation	Subtypes	Business example
Pragmatic	Self-interested calculation by the organization's most immediate audiences	Exchange, influence, dispositional	A supplier is accepted because it reliably delivers value to customers
Moral	Normative judgment of whether the activity is the right thing to do	Consequential, procedural, structural, personal	A bank is judged by whether its lending practices are fair and transparent
Cognitive	Comprehensibility and taken-for-grantedness	Comprehensibility, taken-for-grantedness	A ride-hailing service becomes an unremarkable part of urban transport

Cognitive legitimacy is the most powerful and the hardest to manipulate: once an organizational form is taken for granted, alternatives become hard even to imagine. It is also what new industries lack. Aldrich and Fiol (1994) argued that founders of new industries face both **cognitive** legitimacy deficits (audiences do not understand what the venture is) and **sociopolitical** deficits (key stakeholders and authorities do not yet accept it), and must build trust, reliability, and collective action (Section 17.2) across firms to overcome them.

7.3.2 Strategic balance and optimal distinctiveness

Deephouse (1999) connected legitimacy to competitive strategy. Competitive theories counsel firms to differentiate in order to reduce rivalry. Institutional theory counsels them to conform in order to be legitimate. His **strategic balance** theory holds that firms should be "as different as legitimately possible": moderately differentiated firms avoid both the intense competition faced by conformers and the legitimacy challenges faced by deviants. His study of commercial banks found a curvilinear relationship between strategic similarity and performance consistent with this argument.

The idea parallels, at the firm level, Brewer's theory of optimal distinctiveness (Section 22.3) at the individual level (Brewer 1991), in which people balance needs for assimilation and differentiation. E. Y. Zhao et al. (2017) extended the argument, proposing that the optimal point is not fixed but depends on the stage of a market category's development, the audiences evaluating the firm, and the multiple dimensions on which firms can conform or differ.

BUSINESS CASE

Legitimacy repair after a scandal. When an organization's legitimacy is threatened, Suchman's framework suggests matching the response to the type of legitimacy lost. A product-safety failure damages pragmatic legitimacy (customers no longer trust the exchange) and calls for substantive fixes and guarantees. A misconduct scandal damages moral legitimacy and calls for structural responses, leadership changes, new oversight bodies, independent investigations, that demonstrate that the organization's procedures have changed. Purely symbolic responses are risky because audiences increasingly test whether policies are implemented, the same pressure that reduces policy-practice decoupling.

COMMON PITFALL

Treating legitimacy as reputation or status. Legitimacy concerns whether an organization is acceptable. Reputation concerns how well it is expected to perform. Status concerns its rank in a social hierarchy. A firm can be fully legitimate but have a poor reputation (a reliable but mediocre utility), or be high-status but of contested legitimacy. Define which construct you mean and measure it from the appropriate audience's perspective.

7.4 Stakeholder Theory

THEORY CARD

Core claim. Firms create value through relationships with multiple groups, customers, employees, suppliers, communities, financiers, and others, that affect or are affected by the firm. Managing these relationships well is both an ethical obligation and a source of long-run performance.
Origins. Strategic management and business ethics, R. E. Freeman (1984), T. Donaldson and Preston (1995), R. K. Mitchell, Agle, and Wood (1997), reviewed in Parmar et al. (2010).
Level of analysis. The firm and its stakeholder relationships.
Key constructs. Stakeholder, descriptive, instrumental, and normative stakeholder theory, stakeholder salience (power, legitimacy, urgency), shareholder primacy.
Boundary conditions. Instrumental claims depend on stakeholders' ability to reward or punish firms. Normative claims rest on ethical arguments that cannot be settled by data alone.
Related theories. Resource dependence, Legitimacy, Agency theory (Section 6.1), Social exchange theory (Section 9.1).

7.4.1 Core logic

R. E. Freeman (1984) defined a **stakeholder** broadly as any group or individual that can affect or is affected by the achievement of an organization's objectives, and argued that strategic management should systematically account for all such groups rather than treating them as constraints on maximizing shareholder value. The idea was simple to state but generated decades of debate about what kind of theory it was.

T. Donaldson and Preston (1995) clarified the debate by distinguishing three uses of stakeholder theory. **Descriptive** stakeholder theory claims that firms and managers in fact attend to multiple stakeholders. **Instrumental** stakeholder theory claims that firms that manage stakeholder relationships well perform better on conventional measures. **Normative** stakeholder theory claims that stakeholders' interests have intrinsic value and that managers ought to consider them for their own sake. Donaldson and Preston argued that the three are nested and that the normative core is fundamental. T. M. Jones (1995) developed the instrumental strand most fully, arguing that firms that contract with stakeholders on the basis of mutual trust and cooperation reduce opportunism and monitoring costs and thus gain competitive advantage: a direct challenge to the assumptions of agency theory and TCE.

7.4.2 Stakeholder salience

Not all stakeholders receive equal attention. R. K. Mitchell, Agle, and Wood (1997) proposed that the **salience** of a stakeholder to managers depends on three attributes: **power** (the ability to impose its will on the firm), **legitimacy** (the perceived appropriateness of its relationship and claims), and **urgency** (the extent to which its claims are time-sensitive and critical). Salience increases with the number of attributes a stakeholder possesses, yielding the typology in Figure 7.4 and Table 7.7.

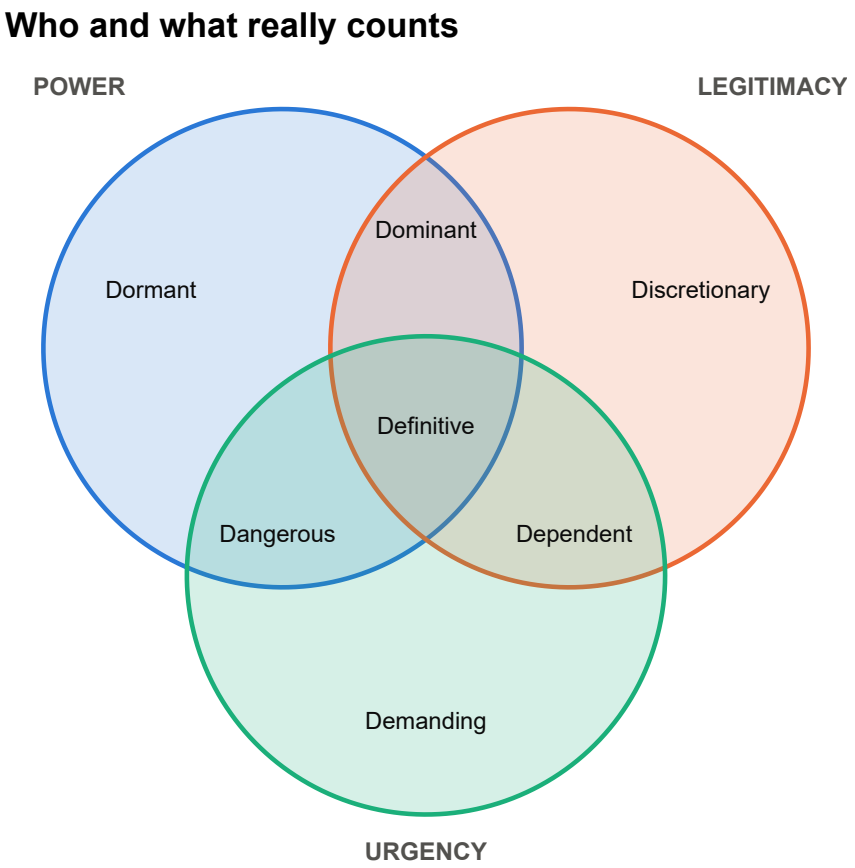


Figure 7.4. Stakeholder salience model, redrawn after Mitchell, Agle, and Wood (1997). Stakeholders possessing one attribute are latent (dormant, discretionary, demanding), two attributes, expectant (dominant, dangerous, dependent), all three, definitive.

Table 7.7. Stakeholder typology by salience (Mitchell, Agle, & Wood, 1997).

Class	Type	Attributes	Illustrative stakeholder
Latent (low salience)	Dormant	Power	A large investor that has never engaged with management
Latent (low salience)	Discretionary	Legitimacy	Beneficiaries of a firm's philanthropy
Latent (low salience)	Demanding	Urgency	A lone activist with a pressing but unsupported claim
Expectant (moderate)	Dominant	Power + legitimacy	Major institutional shareholders, key customers
Expectant (moderate)	Dangerous	Power + urgency	Groups using coercive tactics such as blockades
Expectant (moderate)	Dependent	Legitimacy + urgency	Residents affected by an industrial accident who rely on others to advocate
Definitive (high salience)	Definitive	Power + legitimacy + urgency	A dominant shareholder coalition demanding action after a crisis

The model is dynamic: stakeholders move between classes as they gain or lose attributes. A de-

pendent stakeholder becomes definitive when it acquires power: for instance, when community groups win the support of a regulator or a large investor.

7.4.3 Evidence and the shareholder-primacy debate

The instrumental claim has been tested extensively through studies of the relationship between corporate social performance and financial performance. Margolis and Walsh (2003) reviewed this literature and found a generally positive but methodologically fragile association, and argued that researchers should ask how firms manage the tensions between social and economic objectives rather than only whether "doing good" pays. Barnett and Salomon (2006) found that the financial performance of socially responsible investment funds was curvilinear in the number of social screens they applied, suggesting that the link between social and financial performance is not monotonic.

DEBATE AND EVIDENCE

Shareholder primacy versus stakeholder capitalism. In an influential 1970 essay, Friedman (1970) argued that a business has one social responsibility: to use its resources to increase profits while staying within the rules of the game, engaging in open and free competition without deception or fraud. Spending shareholders' money on social goals, in his view, amounted to managers imposing taxes without accountability. Agency theorists add that a single objective makes managers accountable, whereas multiple objectives let them justify almost any decision.

On August 19, 2019, the Business Roundtable, an association of chief executives of major U.S. companies, released a revised **Statement on the Purpose of a Corporation** signed by 181 CEOs (Business Roundtable 2019). It committed signatory companies to delivering value to customers, investing in employees, dealing fairly and ethically with suppliers, supporting the communities in which they work, and generating long-term value for shareholders, and it replaced earlier Business Roundtable principles that had described serving shareholders as management's paramount duty. Supporters saw the statement as recognition that long-run value creation requires attention to all stakeholders. Skeptics questioned whether it would change behaviour: L. A. Bebchuk and Tallarita (2020) examined signatory companies and argued that the statement was largely rhetorical, finding little evidence of corresponding changes in corporate governance documents or executive incentives, and warning that stakeholder rhetoric can insulate managers from accountability. The debate remains open, and it turns on both empirical questions (does stakeholder orientation improve long-run value?) and normative ones (whose interests *should* the corporation serve?) that T. Donaldson and Preston (1995) argued cannot be resolved by evidence alone.

BUSINESS CASE

Stakeholder mapping for a plant closure. Illustrative scenario: a manufacturer plans to consolidate production and close a plant in a small town. A salience analysis identifies the union and the regional government as *dominant* or *definitive* (power and legitimacy, with urgency rising as the announcement approaches), local suppliers and residents as *dependent* (legitimate, urgent claims but limited power), and a national advocacy group as potentially *dangerous* if it mobilizes coercive tactics. Instrumental stakeholder theory suggests that investing in transition support, early communication, and supplier assistance lowers the costs

of conflict and protects the firm's reputation with employees at other sites. Normative stakeholder theory asks what the firm owes the community regardless of those returns.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Stakeholders possessing more of power, legitimacy, and urgency receive faster and more substantive managerial responses.
- *H2.* Firms with stronger employee-relations practices experience lower turnover and higher innovation, and this effect is stronger in knowledge-intensive industries (instrumental stakeholder theory).
- *H3.* Signatories of public stakeholder commitments do not differ from matched non-signatories in subsequent changes in employee or community outcomes.

Designs. Matched comparisons and difference-in-differences around stakeholder commitments or regulatory changes. Surveys of managers' attention. Text analysis of disclosures and earnings calls.

Measures. Stakeholder attributes coded from documents or expert ratings. Social-performance ratings (with attention to rater disagreement). Stakeholder outcomes (wages, safety, emissions, supplier terms).

COMMON PITFALL

Conflating the three versions of the theory. Instrumental evidence (does stakeholder management pay?) neither confirms nor refutes normative claims (should firms serve stakeholders?), and descriptive evidence (do managers attend to stakeholders?) supports neither. State which version you are testing.

7.5 Organizational Ecology

THEORY CARD

Core claim. Organizational change occurs mainly through the selection of organizations, founding and failure, rather than through the adaptation of individual organizations, because organizations are subject to strong **structural inertia**.

Origins. Sociology, Hannan and Freeman (1977), Hannan and Freeman (1984), Hannan and Freeman (1989), density dependence developed in Hannan and Carroll (1992).

Level of analysis. Populations of organizations sharing a common form, and communities of interacting populations.

Key constructs. Organizational form, population, niche (generalist vs. specialist), structural inertia, liability of newness (Section 12.10), density dependence (legitimation and competition), resource partitioning.

Boundary conditions. Applies to populations whose members share a form and draw on common resources. Less informative about the strategic choices of individual firms.

Related theories. Contingency theory, Institutional theory, Exploration and exploitation (Section 5.3), Dynamic capabilities (Section 5.1).

7.5.1 Origins and core logic

Hannan and Freeman (1977) asked a question borrowed from biology: why are there so many kinds of organizations? Most organization theories explained diversity through adaptation: organizations change to fit their environments. Hannan and Freeman argued that adaptation is limited and that environmental fit is achieved largely by **selection**: organizational forms poorly suited to an environment fail and are replaced by new organizations with better-suited forms.

Hannan and Freeman (1984) explained why adaptation is limited. Modern societies favour organizations that are **reliable** (able to produce collective products of consistent quality) and **accountable** (able to document how resources are used). Reliability and accountability require that structures be highly reproducible: that is, inert. Selection therefore favours inertia. Key propositions follow:

- **P1 (Inertia).** Structural inertia increases with age and, for some features, with size.
- **P2 (Core vs. periphery).** Inertia is strongest in **core** features, stated goals, forms of authority, core technology, and marketing strategy, and weaker in peripheral ones.
- **P3 (Hazard of change).** Attempts to change core features increase the risk of failure in the short run, partly because change resets the organization's "clock" and exposes it again to the **liability of newness** (Stinchcombe 1965, J. Freeman, Carroll, and Hannan 1983).

7.5.2 Density dependence

The best-known empirical regularity of organizational ecology is **density dependence**: the relationship between the number of organizations in a population (its density) and the rates at which organizations are founded and fail (Hannan and Freeman 1989, Hannan and Carroll 1992). Two processes operate. **Legitimation**, the taken-for-grantedness of the organizational form, rises with density, but at a decreasing rate: the first few organizations do much to establish a form, later ones little. **Competition** for limited resources also rises with density, but at an increasing rate, because each additional organization competes with all the others. The combination yields an inverted-U relationship between density and founding rates and a U-shaped relationship between density and failure rates (Figure 7.5).

Density dependence has been documented in many populations, including labour unions, newspapers, banks, breweries, and automobile manufacturers (Hannan and Carroll 1992). A related theory, **resource partitioning**, explains a striking pattern in concentrated markets: as a few large generalists come to dominate the centre of a market, resources at the periphery are freed up for small specialists (G. R. Carroll 1985). G. R. Carroll and Swaminathan (2000) used this logic to explain the rise of microbreweries and brewpubs in the United States: consolidation among mass-production brewers coincided with, and, the theory argues, enabled, a proliferation of specialist producers whose identity as craft producers was valued precisely because it contrasted with the mass producers.

DEBATE AND EVIDENCE

Selection or adaptation? Organizational ecology was a provocation to strategy research, which assumes that managers can adapt their firms. Critics argue that the theory underestimates managerial agency and that many populations show substantial adaptation by incumbents. Ecologists reply that the theory does not deny change, only that change is hazardous and that selection does much of the work at the population level. The debate has become more

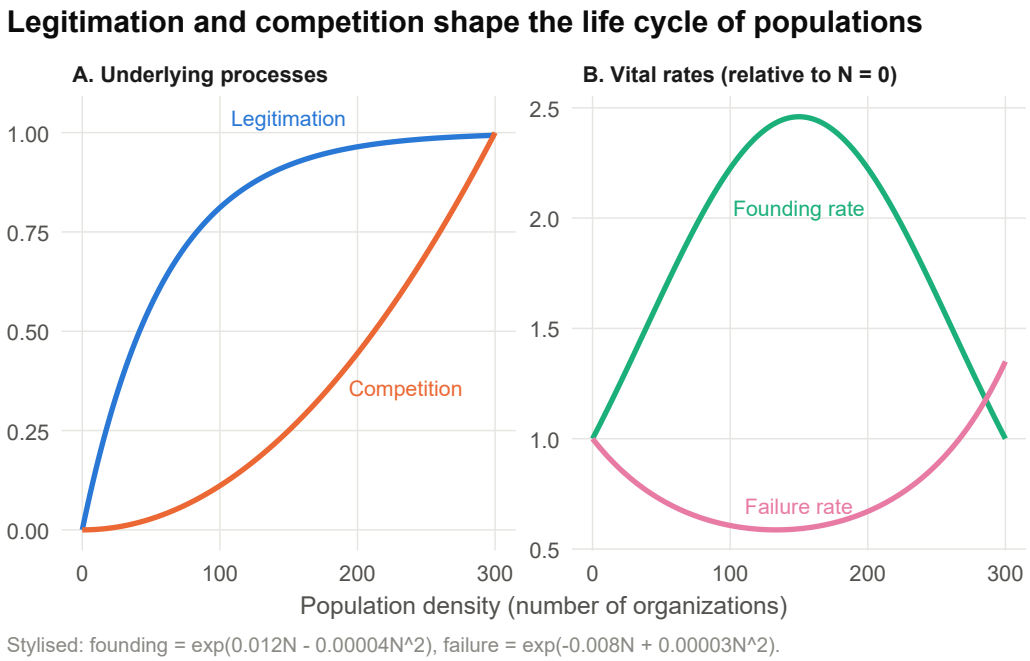


Figure 7.5. Stylised density-dependence model. Left: legitimation rises with density at a decreasing rate while competition rises at an increasing rate. Right: the resulting founding rate is an inverted U and the failure rate a U in population density. Functional forms follow the log-quadratic specification used in density-dependence research. Parameter values are illustrative.

productive as researchers study *when* adaptation succeeds, for example, how the timing and sequence of change affect survival, linking ecology to research on dynamic capabilities (Section 5.1). A second critique concerns measurement: density may proxy for many unmeasured processes, and legitimation is rarely measured directly.

BUSINESS CASE

Inertia in incumbents facing technological change. Public accounts of incumbents such as Kodak in digital photography and Blockbuster in video rental are often told as stories of managerial blindness. Organizational ecology offers a structural interpretation: both firms had core features, business models, assets, and organizational routines, that were reliable and accountable precisely because they were stable, and changing those features was hazardous. The theory implies two practical lessons. First, core change should be undertaken while resources are abundant, because the hazard of change is highest when the organization is already weak. Second, new forms are often better pursued in separate units, where they face the liability of newness without disrupting the core.

7.6 Contingency Theory

THEORY CARD

Core claim. There is no one best way to organize. The effectiveness of an organizational structure depends on its **fit** with contingencies such as environmental uncertainty, technology, size, and strategy.
Origins. Organization theory, T. Burns and Stalker (1961), Woodward (1965), Lawrence and Lorsch (1967), J. D. Thompson (1967), synthesised by L. Donaldson (2001).
Level of analysis. Organization and subunit.
Key constructs. Mechanistic vs. organic structure, environmental uncertainty, differentiation and integration, technology, fit, structural adaptation.
Boundary conditions. Fit is defined relative to a specific contingency. Performance consequences may appear only over time. Managers have some choice in structure.
Related theories. Organizational ecology, Resource dependence, Ambidexterity (Section 5.4), Upper echelons theory (Section 6.4).

7.6.1 Origins and core logic

T. Burns and Stalker (1961) studied firms, many of them Scottish companies attempting to enter the electronics industry, and found two contrasting management systems. **Mechanistic** organizations have specialized tasks, hierarchical control, vertical communication, and an emphasis on loyalty and obedience. **Organic** organizations have fluid task definitions, network-like control, lateral communication, and an emphasis on commitment to the task. Neither is superior in general: mechanistic systems suit stable conditions, while organic systems suit changing conditions in which problems cannot be broken down and assigned in advance (Table 7.8).

Table 7.8. Mechanistic and organic management systems (after Burns & Stalker, 1961).

Feature	Mechanistic	Organic
Task definition	Specialized, precisely defined, relatively fixed	Fluid, continually redefined through interaction
Coordination and control	Hierarchical, rules and supervision	Network-like, shared commitment to the task
Communication	Vertical, instructions and decisions	Lateral, information and advice
Location of knowledge	Concentrated at the top	Distributed throughout the organization
Basis of commitment	Loyalty and obedience to superiors	Commitment to the task and the organization's progress
Fitting environment	Stable, predictable	Dynamic, uncertain
Business example	High-volume assembly, regulated back-office processing	R&D labs, product teams in fast-moving software markets

Lawrence and Lorsch (1967) extended the argument to subunits. Departments facing different sub-environments, research facing scientific uncertainty, production facing technical stability, sales facing market change, should develop different structures and orientations, a condition they called **differentiation**. Yet differentiated units must also work together, which requires **integration**. In their comparative study of firms in industries with different degrees of uncertainty, high-performing firms in uncertain environments achieved both high differentiation and high

integration, often through integrating roles and cross-functional teams. Woodward (1965) showed that effective structure also varied with production technology, J. D. Thompson (1967) linked structure to the type of interdependence among units, and Galbraith (1973) reformulated the theory in information-processing terms: the greater the task uncertainty, the more information must be processed, and structures must either reduce the need for information processing or increase the capacity for it.

7.6.2 Fit and performance

The central proposition is that performance depends on the **fit** between structure and contingencies. Figure 7.6 illustrates the classic interaction: under low environmental dynamism, mechanistic structures outperform organic ones. Under high dynamism, the ranking reverses.

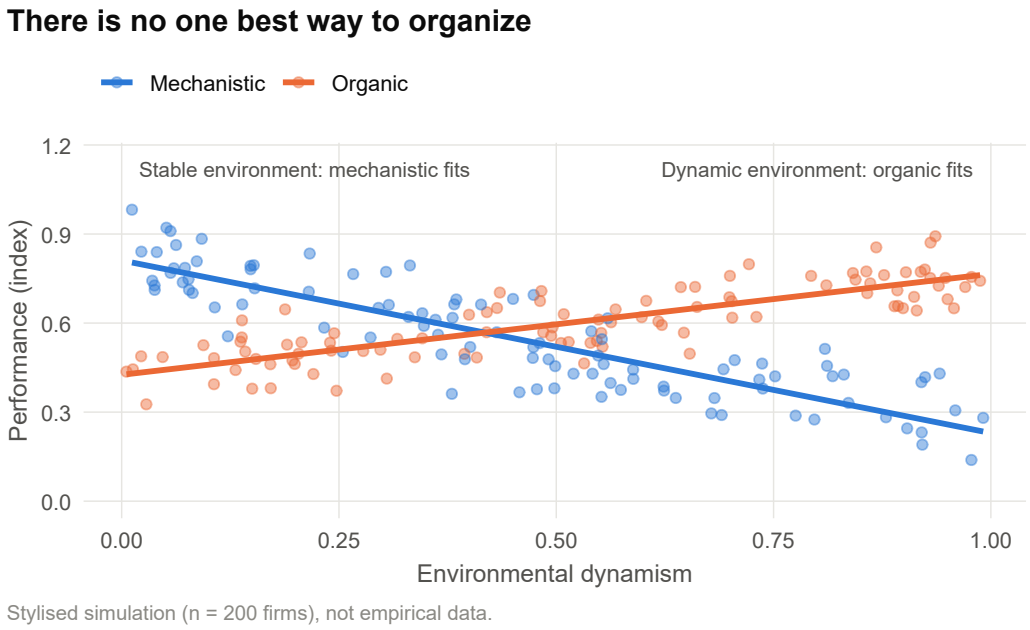


Figure 7.6. Illustrative contingency-fit interaction using simulated firms: mechanistic structures perform better in stable environments and organic structures in dynamic ones. Points are stylised simulated firms, lines are linear fits.

L. Donaldson (2001) proposed that organizations move in and out of fit over time: as contingencies such as size or diversification change, a once-fitting structure becomes a misfit, performance suffers, and the organization eventually adapts its structure to regain fit. This **structural adaptation** model gives contingency theory a dynamic form.

> **DEBATE AND EVIDENCE**
>
> **Determinism and strategic choice.** Early contingency theory was criticized for implying that environments determine structures. Child (1972) argued that this neglects **strategic choice**: managers select which environments to operate in, can influence those environments, and may accept some misfit for other goals such as control. Later critiques noted that fit is hard to measure, that multiple contingencies may pull in different directions, and that equally effective configurations may exist for the same contingency (equifinality). Contemporary research

on ambidexterity (Section 5.4) can be read as a response: firms facing both stable and dynamic demands build differentiated units and integrate them at the top, Lawrence and Lorsch's insight applied to exploration and exploitation.

BUSINESS CASE

Structuring a firm with two environments. Illustrative scenario: a payments company runs a highly regulated transaction-processing business and a fast-moving consumer app. Contingency theory predicts that the processing unit should be relatively mechanistic, standardized procedures, clear accountability, change control, while the app unit should be organic, with cross-functional teams and rapid iteration. Following Lawrence and Lorsch (1967), the critical design problem is integration: shared platforms, integrating roles such as product-operations liaisons, and senior forums that resolve conflicts between the units' different time horizons and risk tolerances.

COMMON PITFALL

Testing fit with main effects. Contingency theory predicts interactions, not main effects: the effect of structure on performance depends on the contingency. Specify the form of fit (moderation, matching, or configuration), test it with an appropriate model, and measure contingencies independently of structure to avoid circularity.

7.7 Corporate Social Responsibility Pyramid

THEORY CARD

Core claim. Corporate social responsibility comprises four kinds of societal expectation: economic, legal, ethical, and discretionary (philanthropic). A responsible firm meets all four at once, with economic performance as the foundation on which the others rest.
Origins. Business and society scholarship. A. B. Carroll (1979) defined the four categories as part of a three-dimensional model of corporate social performance. A. B. Carroll (1991) recast them as a pyramid. M. S. Schwartz and Carroll (2003) revised the model as three overlapping domains.
Level of analysis. The firm and the societal expectations it faces.
Key constructs. Economic, legal, ethical, and philanthropic responsibilities, social issues, social responsiveness (reaction, defense, accommodation, proaction), corporate social performance.
Boundary conditions. The content of each layer depends on the institutional context: what is a legal duty in one country may be voluntary in another. The ordering describes logical foundations, not a sequence in which firms should fulfil the layers.
Related theories. Stakeholder theory, Legitimacy, Institutional theory, Creating shared value.

7.7.1 Core logic

Debates about corporate social responsibility (CSR) in the 1960s and 1970s often treated profit and responsibility as opposites. A. B. Carroll (1979) proposed a definition that included both: the

social responsibility of business encompasses the economic, legal, ethical, and discretionary expectations that society holds of organizations at a given point in time. Economic responsibilities come first because society expects business to produce goods and services and to sell them at a profit. Legal responsibilities reflect society's codified expectations. Ethical responsibilities cover behaviour that society expects but has not written into law. Discretionary responsibilities are voluntary contributions that society values but does not demand. Carroll embedded these categories in a three-dimensional model that also specified the **social issues** a firm faces (such as consumerism, the environment, discrimination, and product and occupational safety) and its **philosophy of responsiveness**, ranging from reaction and defense to accommodation and proaction.

A. B. Carroll (1991) presented the four categories as a **pyramid**, shown in Figure 7.7, and renamed the top layer **philanthropic**. The pyramid conveys that the economic layer is the foundation on which the others rest, and that philanthropy, although desired by society, is less important than the other three. Carroll stressed that the layers are not sequential: the responsible firm strives to make a profit, obey the law, be ethical, and be a good corporate citizen simultaneously. The pyramid also links CSR to Stakeholder theory: each layer can be read against the expectations of specific stakeholders.



Figure 7.7. Carroll's pyramid of corporate social responsibility, redrawn after Carroll (1991). Layers indicate logical foundations, not a sequence: the responsible firm meets all four expectations at once.

Table 7.9 summarizes each layer with examples. The examples show that the same issue can fall in different layers depending on whether it is codified in law.

Table 7.9. The four responsibilities of Carroll's pyramid (after Carroll, 1979, 1991).

Responsibility	Societal expectation	Illustrative practices	Typical failure
Economic	Produce valued goods and services at a profit	Efficient operations, sustained returns, employment	Insolvency, value destruction for investors and employees
Legal	Operate within the law and regulation	Compliance programmes, product safety standards, accurate disclosure	Fraud, regulatory violations, cheating on emissions or safety tests
Ethical	Do what is right and fair beyond legal minimums	Fair supplier terms, honest marketing, respect for human rights in supply chains	Exploitative but legal practices, deceptive framing
Philanthropic	Contribute voluntarily to community well-being	Donations, employee volunteering, product donation programmes	Philanthropy used to offset or distract from harm

7.7.2 Evidence and extensions

Aupperle, Carroll, and Hatfield (1985) developed a forced-choice instrument to measure managers' orientation toward the four categories. Their respondents weighted the domains in roughly the order Carroll proposed, from economic to discretionary, and the study found no significant relationship between social orientation and profitability. M. S. Schwartz and Carroll (2003) argued that the pyramid's geometry invites misreading, as if the top layers mattered less or could wait until the base was secure. They proposed a **three-domain model** of overlapping economic, legal, and ethical circles, folding philanthropy into the ethical and economic domains, so that activities can be classified by the combinations of motives they serve. Matten and Moon (2008) argued that the meaning of CSR varies across institutional settings. In much of continental Europe, responsibilities that U.S. firms address through voluntary **explicit** CSR policies are handled through **implicit** CSR embedded in law, collective agreements, and national institutions. The layers of the pyramid are therefore not fixed: the boundary between "legal" and "ethical" or "philanthropic" shifts with the institutional context, a point that connects the pyramid to Institutional theory.

BUSINESS CASE

Philanthropy and legal failure at two ends of the pyramid. In 1987 Merck announced that it would donate its river-blindness drug Mectizan (ivermectin) to affected communities for as long as needed, a widely cited example of the philanthropic layer that builds on the economic and ethical capabilities of the firm. At the other extreme, in September 2015 the U.S. Environmental Protection Agency issued a notice of violation alleging that Volkswagen had installed software in diesel cars to defeat emissions tests. The episode shows how a failure in the legal and ethical layers can overwhelm a firm's other CSR activities: no volume of philanthropy or sustainability reporting offsets deliberate violation of the law. In pyramid terms, the upper layers are credible only when the lower ones are secure.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Firms whose CSR disclosures emphasise philanthropic over legal and ethical responsibilities will experience larger reputational penalties after misconduct is revealed. *H2.* The share of CSR activities that firms present as voluntary will be higher in liberal than in coordinated market economies (Matten and Moon 2008).

Designs and measures. Content analysis of CSR reports coded into Carroll's categories or the three-domain model. Forced-choice surveys of managerial orientation (Aupperle, Carroll, and Hatfield 1985). Cross-country comparisons that hold industry constant.

7.8 Creating Shared Value

THEORY CARD

Core claim. Firms can create economic value in ways that also create value for society by addressing social needs and challenges. Such **shared value** should be integral to strategy and profit-making rather than a separate CSR activity.
Origins. Strategy. Michael E. Porter and Kramer (2006) distinguished responsive from strategic CSR. Michael E. Porter and Kramer (2011) introduced the concept of creating shared value (CSV). Crane et al. (2014) offered the most prominent critique.
Level of analysis. The firm, its value chain, and its local competitive context.
Key constructs. Shared value, responsive versus strategic CSR, inside-out and outside-in linkages, reconceiving products and markets, redefining productivity in the value chain, local cluster development.
Boundary conditions. Applies where social problems impose costs on the firm or represent unmet demand. Does not resolve cases in which social and economic goals genuinely conflict.
Related theories. Stakeholder theory, Corporate social responsibility pyramid, Resource-based theory (Section 4.1), Legitimacy.

7.8.1 Core logic

Michael E. Porter and Kramer (2006) argued that much CSR was fragmented and disconnected from strategy: firms responded to external pressure with philanthropy and reporting that were unrelated to their businesses. They distinguished **responsive CSR**, acting as a good citizen and mitigating harm from value-chain activities, from **strategic CSR**, which selects a small number of social issues where the firm can make a distinctive contribution that also strengthens its competitive position. Their analysis linked business and society in two directions. **Inside-out** linkages run from the firm's value chain to society (for example, emissions, labour practices, or product safety). **Outside-in** linkages run from social conditions to the firm's competitive context (for example, the skills of the local workforce or the quality of local suppliers and infrastructure).

Michael E. Porter and Kramer (2011) generalized the argument into the concept of **creating shared value**: policies and operating practices that enhance a firm's competitiveness while simultaneously advancing economic and social conditions in the communities where it operates. They identified three routes. Firms can **reconceive products and markets** to meet social needs, such as health, nutrition, or financial inclusion. They can **redefine productivity in the value chain**, for instance by reducing energy use, waste, and logistics costs, or by improving the productivity of suppliers. And they can **enable local cluster development** by strengthening the suppliers, institutions, and infrastructure around their operations. Porter and Kramer presented CSV as a successor to CSR and contrasted the two, as in Table 7.10.

Table 7.10. Conventional CSR versus creating shared value, with critics' objections (after Porter and Kramer, 2011, and Crane et al., 2014).

Dimension	Conventional CSR (as characterized by Porter and Kramer)	Creating shared value	Objection raised by Crane et al. (2014)
Core idea	Doing good, citizenship, philanthropy	Joint economic and social value relative to cost	Repackages long-standing ideas from CSR and stakeholder research
Relation to profit	Separate from or a cost to profit	Integral to profit and competitiveness	Overlooks genuine trade-offs between social and economic goals
Driver	External pressure and reputation	Company-specific strategy	Assumes firms already comply with laws and ethical norms
Budget and scope	Limited CSR budget	Whole company budget and value chain	Retains a narrow, firm-centred view of the corporation's role in society

Figure 7.8 locates CSV relative to other activities. Shared value occupies the upper-right region, where an activity raises both economic and social value. The concept's critics focus on the lower-right region, where activities are profitable but socially harmful, and on the large set of situations in which no option lies in the upper right.

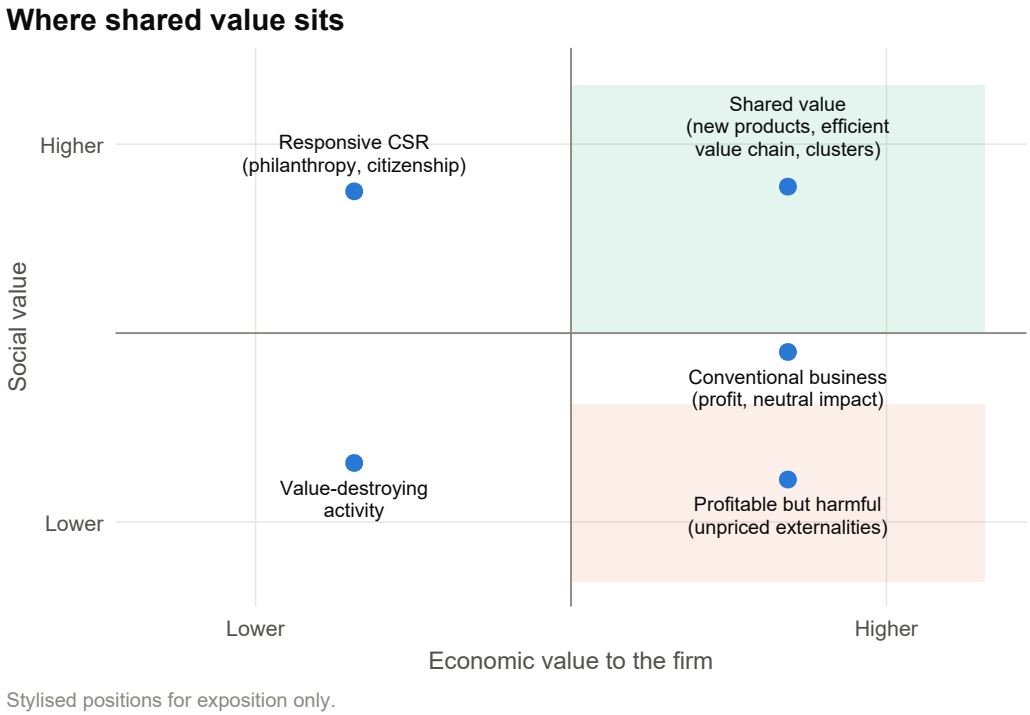


Figure 7.8. Stylised map of corporate activities by economic value to the firm and social value. Creating shared value targets the upper-right region. Positions are illustrative, not measured.

7.8.2 Evidence and critique

Porter and Kramer illustrate the concept with company examples, among them Nestlé's work with dairy farmers in India, where investment in collection, refrigeration, and veterinary support raised

farmers' output and secured the company's milk supply (Michael E. Porter and Kramer 2006). Nestlé subsequently adopted "Creating Shared Value" as the organizing label for its corporate responsibility reporting. The concept spread quickly among managers and consultancies, but systematic evidence that CSV initiatives outperform conventional CSR is limited, partly because "shared value" is usually identified after the fact from successful cases.

DEBATE AND EVIDENCE

Is shared value new, and is it enough? Crane et al. (2014) praised CSV for elevating social goals to the level of strategy, but raised four objections. It is **unoriginal**, presenting ideas long developed in stakeholder and CSR research as new. It **ignores the tensions** between social and economic goals, focusing on win-win cases when many decisions involve real trade-offs. It is **naive about compliance**, assuming that firms already obey the law and meet ethical standards. And it rests on a **shallow conception of the corporation's role in society**, leaving the profit motive and corporate power unexamined. In the terms of Carroll's pyramid, CSV concentrates on the economic layer and treats the legal and ethical layers as given. Defenders reply that CSV is a strategic tool, not a complete theory of corporate responsibility, and that its value lies in mobilizing operating budgets rather than philanthropic ones.

BUSINESS CASE

Illustrative scenario: a food company's sourcing programme. A packaged-food company sources a key ingredient from smallholder farmers whose low yields and inconsistent quality raise its costs. A CSV programme would invest in agronomic training, shared processing facilities, and guaranteed purchase contracts, improving farmers' incomes while lowering the company's input costs and supply risk. The critique of Crane et al. (2014) suggests the questions to ask before calling it shared value: Are farmers' gains measured independently, or only through the company's own reporting? What happens when market prices fall and the company's interest diverges from the farmers'? Does the programme coexist with practices, such as aggressive marketing of unhealthy products, that destroy social value elsewhere?

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Social initiatives embedded in a firm's value chain will persist longer through budget cuts than philanthropic initiatives of similar size. *H2.* The financial returns to CSV initiatives will be larger when the targeted social problem directly raises the firm's operating costs.

Designs and measures. Program evaluation with independent measurement of social outcomes (randomized or staggered roll-outs of supplier programmes). Panel studies linking value-chain sustainability investments to cost and revenue outcomes. Text analysis of how firms label initiatives (CSR, CSV, ESG) and whether labels predict substantive differences.

7.9 Chapter Summary

KEY TAKEAWAYS

- **Resource dependence theory**: dependence on others for important, discretionary, and concentrated resources confers power on them. Power is relational and situational, and organizations restructure dependencies through co-optation, alliances, mergers, and political action (Pfeffer and Salancik 1978, Emerson 1962).
- Power imbalance and mutual dependence have opposite effects on constraint absorption (Casciaro and Piskorski 2005).
- **Institutional theory**: organizations adopt rationalized structures to gain legitimacy, producing coercive, mimetic, and normative isomorphism and, often, decoupling (J. W. Meyer and Rowan 1977, DiMaggio and Powell 1983, Scott 2014).
- Organizations respond to institutional pressures strategically, from acquiescence to manipulation (C. Oliver 1991).
- **Legitimacy** can be pragmatic, moral, or cognitive (Suchman 1995). Firms benefit from being as different as legitimately possible (Deephouse 1999).
- **Stakeholder theory** has descriptive, instrumental, and normative versions (T. Donaldson and Preston 1995). Salience depends on power, legitimacy, and urgency (R. K. Mitchell, Agle, and Wood 1997). The shareholder-stakeholder debate remains unresolved.
- **Organizational ecology**: structural inertia makes selection a central driver of change. Density dependence reflects legitimation and competition (Hannan and Freeman 1984, Hannan and Carroll 1992).
- **Contingency theory**: effective structures fit their environments, mechanistic for stable, organic for dynamic conditions (T. Burns and Stalker 1961, Lawrence and Lorsch 1967).

Table 7.11 presents the comparison of the chapter's theories.

Table 7.11. Comparison of the chapter's theories.

Theory	View of the environment	Key mechanism	Signature prediction	Typical business use
Resource dependence	Other organizations controlling critical resources	Power from dependence, restructuring of exchange	Firms co-opt or absorb actors on whom they depend	Supplier strategy, board design, alliances, platform strategy
Institutional theory	Rules, norms, and shared beliefs in a field	Coercive, mimetic, and normative isomorphism	Organizations in a field converge, adoption may be decoupled	Certification, compliance, practice adoption, ESG reporting
Legitimacy	Audiences evaluating social acceptability	Pragmatic, moral, and cognitive evaluation	Moderate differentiation outperforms extremes	Positioning, crisis management, new-category creation
Stakeholder theory	Groups that affect or are affected by the firm	Salience, trust-based contracting	Salient stakeholders receive attention, trust lowers costs	Corporate purpose, stakeholder mapping, CSR strategy
Organizational ecology	Resource niches supporting populations	Selection under structural inertia	Founding and failure depend on density, core change is hazardous	Industry evolution, entry timing, incumbent renewal

Table 7.11. Comparison of the chapter's theories. *(continued)*

Theory	View of the environment	Key mechanism	Signature prediction	Typical business use
Contingency theory	Uncertainty, technology, and size as contingencies	Fit between structure and contingency	Organic structures outperform in dynamic settings	Organizational design, unit structure, integration roles
CSR pyramid	Society holding economic, legal, ethical, and philanthropic expectations	Layered societal expectations of business	Upper layers are credible only when lower layers are met	CSR strategy, compliance, philanthropy design
Creating shared value	Social needs as conditions of competitiveness	Joint creation of economic and social value	Value-chain social investments raise profits	Sourcing, product innovation, cluster development

7.10 Discussion Questions

1. Resource dependence theory and transaction cost economics both predict that firms absorb problematic exchange partners. Design a study that could distinguish a power-based from an efficiency-based explanation of a vertical acquisition.
2. Choose a practice that has diffused widely in your industry (for example, agile methods, ESG reports, or AI governance policies). Which of the three isomorphic mechanisms best explains its spread? What evidence would show decoupling?
3. Apply Suchman's three types of legitimacy to a new industry such as autonomous vehicles or cultivated meat. Which type is the binding constraint today, and what collective strategies could build it?
4. Map the stakeholders of a firm facing a controversial decision using the salience model. How would the map change if one dependent stakeholder gained power?
5. Does the 2019 Business Roundtable statement represent a change in corporate behaviour or in corporate rhetoric? What data would you collect to find out?
6. A large incumbent is considering a radical change to its core business model. What do organizational ecology and contingency theory each advise, and where do they conflict?

7.11 Further Reading

- Pfeffer and Salancik (1978): the founding statement of resource dependence theory. Still the best source for its full argument.
- Casciaro and Piskorski (2005): a model of how disaggregating a core construct sharpens a theory's predictions.
- Hillman, Withers, and Collins (2009): a concise review of three decades of resource dependence research.
- DiMaggio and Powell (1983): the classic statement of institutional isomorphism and its testable hypotheses.
- Suchman (1995): the standard framework for analyzing and managing legitimacy.
- R. K. Mitchell, Agle, and Wood (1997): the stakeholder salience model and typology.
- Hannan and Freeman (1984): the theory of structural inertia and why change is hazardous.
- Lawrence and Lorsch (1967): the foundational study of differentiation, integration, and fit.

Chapter 8

Information, Signals, and Markets

Why does a used-car dealer offer a 12-month warranty it hopes never to honour? Why do start-ups pay premium fees to be underwritten by prestigious investment banks, and why do some of the wealthiest consumers buy handbags with *no* visible logo? Each puzzle has the same structure. One party knows something valuable, the quality of a car, the prospects of a firm, the wealth of a buyer, and another party cannot observe it directly. The informed party therefore *does* something observable that the uninformed party can use to infer what is hidden.

This chapter introduces the family of theories that explain how markets, organizations, and people cope with **information asymmetry**. We begin with the problem itself (Akerlof's "lemons"), then develop **signaling theory**, its mirror image **screening theory**, the biology-inspired idea of **costly signaling**, and two consumer-behaviour extensions: **identity signaling** and **countersignaling**. Together they form one of the most widely used theoretical toolkits in management, marketing, finance, and economics (Table 8.1).

Table 8.1. Chapter at a glance: theories of information and signals.

Theory	Core question	Level of analysis	Key source
Information asymmetry	What happens when one side of a transaction knows more?	Market / dyad	Akerlof (1970)
Signaling theory	How can the informed party credibly reveal hidden quality?	Dyad / market	Spence (1973)
Screening theory	How can the uninformed party induce disclosure?	Dyad / market	Rothschild & Stiglitz (1976)
Costly signaling	Why must honest signals be expensive to fake?	Individual / population	Zahavi (1975)
Identity signaling	Why do consumers converge or diverge in their choices?	Individual / group	Berger & Heath (2007)
Countersignaling	Why do the highest types sometimes refuse to signal?	Individual / market	Feltovich et al. (2002)

8.1 Information Asymmetry

THEORY CARD

Core claim. When sellers know more about quality than buyers, buyers pay only for *average* quality. High-quality sellers withdraw, average quality falls, and the market can unravel.
Origins. Economics. Akerlof (1970), building on earlier work on uncertainty and insurance. Akerlof, Spence, and Stiglitz shared the 2001 Nobel Memorial Prize in Economic Sciences for analyses of markets with asymmetric information.
Level of analysis. Market and dyad (buyer-seller, investor-firm, employer-applicant).
Key constructs. Hidden information (adverse selection), hidden action (moral hazard), quality uncertainty, pooling.
Boundary conditions. Quality is unobservable *before* purchase, is costly to verify, and matters to the buyer. No cheap third-party verification exists.
Related theories. Signaling, Screening, Agency theory (Section 6.1), Transaction cost economics (Section 6.3).

8.1.1 Core logic

Akerlof (1970) used the market for used cars to show how hidden quality can destroy a market. Suppose half of the cars are good ("peaches") and half are bad ("lemons"). Buyers cannot tell them apart, so they are willing to pay only the expected value of a random car. Owners of good cars find that price too low and keep their cars. The share of lemons in the market rises. Buyers anticipate this and lower their offers further. In the limit, only lemons trade. The general lesson is that **adverse selection**, the tendency of hidden-information markets to attract the "wrong" types, can make mutually beneficial trades disappear.

Economists distinguish two forms of asymmetry (Stiglitz 2000, Connelly et al. 2011) (Table 8.2):

Table 8.2. Two forms of information asymmetry.

Dimension	Hidden information	Hidden action
Also called	Adverse selection	Moral hazard
What is hidden	A characteristic (quality, ability, risk type)	Behaviour after contracting (effort, care)
Timing	Before the transaction	After the transaction
Classic example	Used-car quality, applicant ability	Insured driver's care, manager's effort
Typical remedy	Signals, screens, certification, warranties	Monitoring, incentives, bonding
Home theory	Signaling / screening theory	Agency theory

8.1.2 Mechanism

Figure 8.1 simulates the unravelling process. Quality is uniformly distributed, buyers value a car at 1.2 times its quality, and sellers only sell if the price is at least their car's quality. Each round, buyers offer the expected value of the cars that remain on the market. The market shrinks round after round.

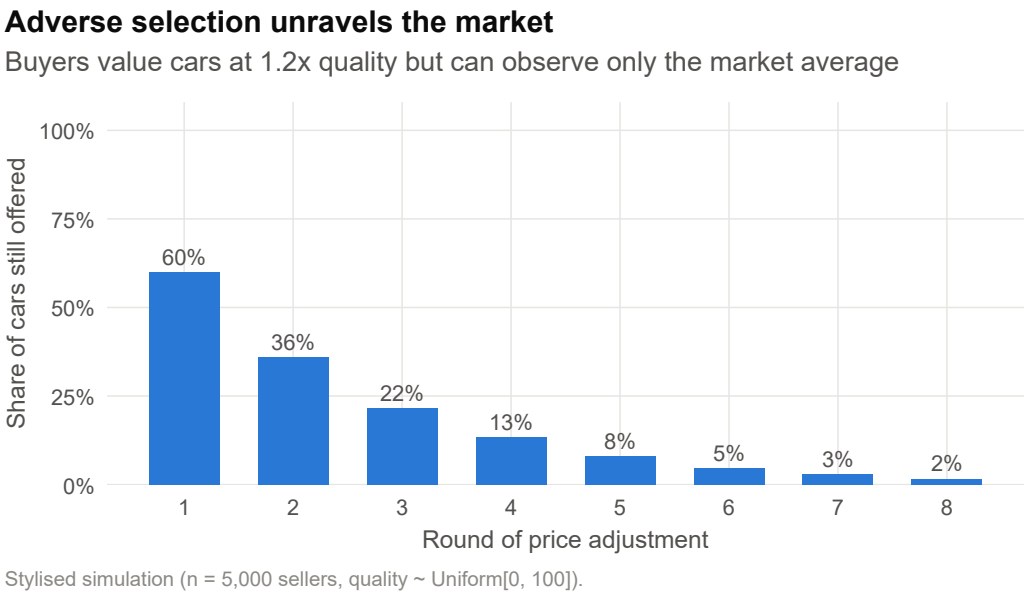


Figure 8.1. Illustrative simulation of Akerlof's lemons problem: when buyers can only pay for average quality, the best sellers exit and the market shrinks.

8.1.3 Applications in business

BUSINESS CASE

Certified pre-owned programs. The used-car market is the textbook lemons market, and the industry's best-known response is the manufacturer-backed *certified pre-owned* (CPO) program: a multi-point inspection plus an extended warranty. Both elements attack the asymmetry directly: inspection converts hidden information into verifiable information, and the warranty makes a low-quality car expensive for the *seller* to certify. Similar logic underlies home-inspection contingencies in real estate, escrow services in online marketplaces, and "satisfaction guaranteed" return policies in retailing.

BUSINESS CASE

Mandatory disclosure. Information can also be disclosed by regulation. When Los Angeles County required restaurants to post hygiene grade cards in their windows, health inspection scores rose and foodborne-illness hospitalizations fell (Jin and Leslie 2003). A hidden attribute (kitchen hygiene) became a visible one, and consumers and restaurants both responded. See Dranove and Jin (2010) for a review of disclosure and certification.

8.2 Signaling Theory

THEORY CARD

Core claim. An informed party (the *signaler*) can credibly convey unobservable quality by taking a costly, observable action that low-quality types would find too expensive to imitate.
Origins. Labour economics. Spence (1973) on education as a job-market signal. Reviewed in Spence (2002) and, for management, Connelly et al. (2011).
Level of analysis. Dyad and market: job applicant-employer, firm-investor, brand-consumer, start-up-venture capitalist.
Key constructs. Signaler, receiver, signal, signal cost, signal fit, signal observability, signal consistency, separating vs. pooling equilibrium, feedback (countersignals).
Boundary conditions. Quality is unobservable *ex ante*. The cost of the signal is *negatively correlated with quality*. Receivers can observe the signal and know (or learn) how to interpret it.
Related theories. Information asymmetry, Screening, Costly signaling, Agency theory (Section 6.1), Institutional theory (Section 7.2).

8.2.1 Origins and intellectual history

Spence (1973) asked why employers pay more to workers with more education even if education teaches nothing productive. His answer: education is a **signal**. If acquiring a degree is less costly, in effort, time, or psychic cost, for high-ability people, then only high-ability people will find it worthwhile to acquire it, and employers can rationally use the degree to sort applicants. The signal need not *create* the quality it reveals. It only needs to be harder to obtain for those who lack it.

The idea spread quickly. In finance, debt levels and dividends were modelled as signals of managers' private information (S. A. Ross 1977, S. Bhattacharya 1979, Myers and Majluf 1984). In marketing, advertising spending, price, warranties, and brand names were analysed as signals of unobservable product quality (P. Nelson 1974, P. Milgrom and Roberts 1986, Kirmani and Wright 1989, Boulding and Kirmani 1993, Erdem and Swait 1998, Kirmani and Rao 2000). In management, signaling theory became a standard lens for IPOs, board composition, corporate social responsibility, and the hiring process (Certo 2003, Connelly et al. 2011, Bergh et al. 2014).

8.2.2 Core logic and propositions

Connelly et al. (2011) organize signaling theory around a timeline with four elements. Figure 8.2 redraws that logic.

Formally, a signal is informative when it supports a **separating equilibrium**: different types choose different signals, so the receiver can infer type from the signal. When all types send the same signal, the equilibrium is **pooling** and the signal carries no information. The theory's central propositions can be stated as follows:

- **P1 (Differential cost).** A signal separates types only if its cost is lower for high-quality than for low-quality signalers (the *single-crossing* condition).
- **P2 (Observability).** The more observable a signal is to receivers, the stronger its effect on receiver beliefs (Connelly et al. 2011).

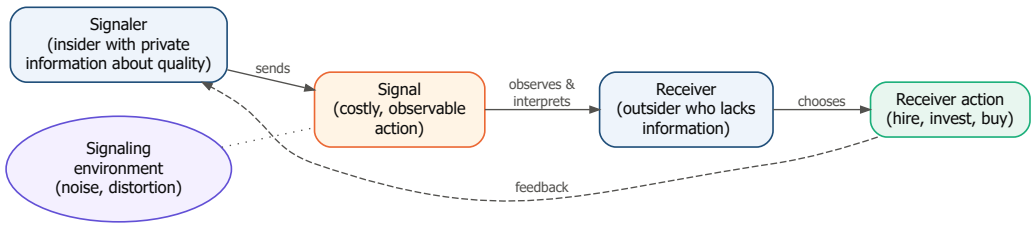


Figure 8.2. The signaling timeline: a signaler with private information sends a signal. The receiver observes and interprets it, chooses an action, and may send feedback. Adapted from the framework in Connelly et al. (2011).

- **P3 (Fit).** Signals are more effective when they are logically connected to the unobservable quality being signalled (e.g., a warranty fits durability. A celebrity endorsement fits less well).
- **P4 (Consistency).** Consistent signals across time and channels are more credible than one-off or conflicting signals.
- **P5 (Receiver attention and interpretation).** Signal effects depend on whether receivers attend to the signal and on how they calibrate it. Signals are socially interpreted, not mechanically decoded.

Figure 8.3 shows the single-crossing logic for Spence's education model. Because the high-ability type's cost line is flatter, there is a range of education levels (shaded) that high types are willing to acquire but low types are not. Any education requirement in that range separates the two types.

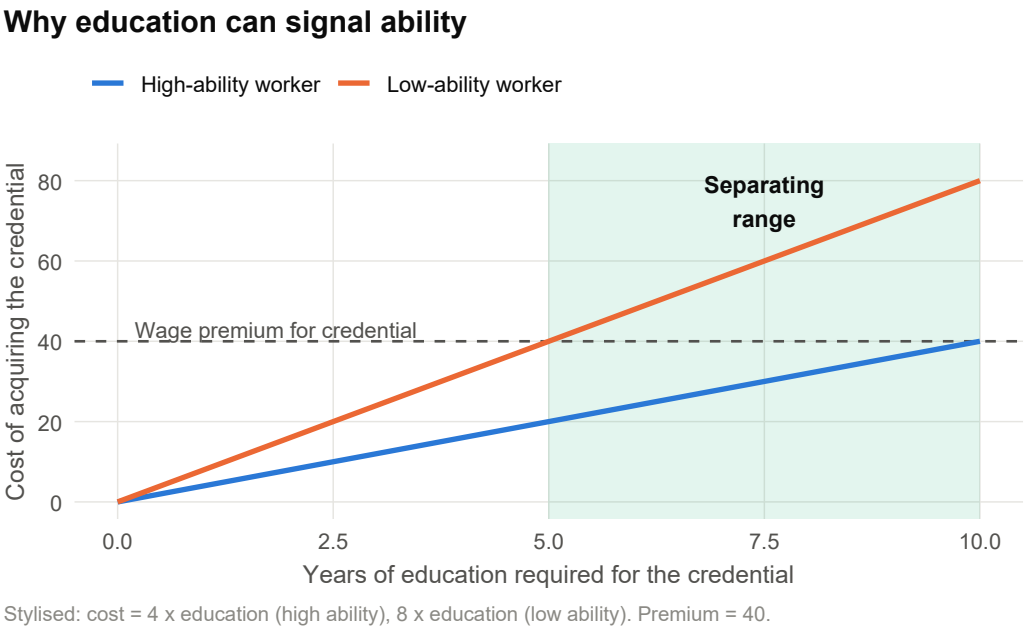


Figure 8.3. Stylised Spence model. Education costs more for low-ability workers. A credential level in the shaded range is worth acquiring for high-ability workers (wage premium exceeds cost) but not for low-ability workers, producing a separating equilibrium.

8.2.3 Characteristics of effective signals

Table 8.3 summarises the constructs that empirical studies most often operationalize.

Table 8.3. Key constructs in signaling research (organized following Connelly et al., 2011).

Element	Construct	Definition	Example operationalization
Signaler	Signaler quality	The unobservable ability of the signaler to meet receivers' needs	Post-IPO survival, product failure rate
Signaler	Honesty / credibility	Whether the signaler actually has the signalled quality	Restatements, litigation, returns
Signal	Signal cost	Expense of producing the signal, ideally higher for low types	Advertising spend, warranty length
Signal	Observability	Extent to which receivers can notice the signal	Media coverage, logo size
Signal	Fit	Correspondence between signal and the underlying quality	Warranty to durability, patents to innovativeness
Signal	Consistency	Agreement among signals over time and sources	Variance in ratings, repeated certification
Receiver	Attention	Receiver vigilance in scanning for signals	Analyst coverage, search behaviour
Receiver	Interpretation	How receivers translate signals into beliefs	Price paid, valuation, choice
Environment	Noise / distortion	Factors that reduce signal clarity	Number of competing signals, market volatility

8.2.4 Evidence and extensions

The empirical record is large and spans disciplines:

- **Product quality.** Consumers infer quality from warranties (Boulding and Kirmani 1993), perceived advertising expense (Kirmani and Wright 1989), and price-advertising combinations (P. Milgrom and Roberts 1986). Kirmani and Rao (2000) synthesise this literature and argue that "no pain, no gain": signals work when a low-quality firm would lose money by sending them.
- **Brands.** Erdem and Swait (1998) conceptualise **brand equity (Section 27.3) itself as a signal**: a brand is a credible promise because the firm has invested in it and stands to lose that investment if it cheats.
- **Capital markets.** Prestigious boards and underwriters are associated with better IPO outcomes (Certo 2003). In corporate finance, issuing equity can signal that managers believe shares are *over*valued, which helps explain negative stock-price reactions to equity issues (Myers and Majluf 1984).
- **Management research.** Bergh et al. (2014) note that many studies assume rather than test the differential-cost condition, and call for research on equilibrium: whether signals actually separate types in the population studied.

> **DEBATE AND EVIDENCE**
>
> **Is it really a signal?** Many papers label any observable firm attribute a "signal". Critics argue that a variable is a signal in the theoretical sense only if (a) it is chosen by the signaler, (b) its cost differs by type, and (c) it changes receiver beliefs. Without evidence on (b), a "signal"

may simply be a correlate of quality, and the theory is doing little explanatory work (Bergh et al. 2014). A second debate concerns the *human capital* alternative: education may raise productivity and also signal it. Separating these explanations requires careful designs (e.g., sheepskin effects, natural experiments).

8.2.5 Applications in business

Signals appear in every business function. Table 8.4 maps common signals to the hidden attributes they communicate.

Table 8.4. Signals across business functions.

Function	Signaler	Hidden attribute	Signal	Why it is costly to fake
Marketing	Manufacturer	Product durability	Long warranty	Low-quality products generate expensive claims
Marketing	New brand	Quality and staying power	Heavy launch advertising	Spend is recouped only through repeat purchase
Finance	IPO firm	Firm prospects	Prestigious underwriter / board	Reputational intermediaries screen and risk their name
Finance	Mature firm	Stable cash flows	Regular dividends	Cutting dividends later is penalised
HR	Job applicant	Ability and diligence	Degree, certifications	Harder to complete for lower-ability candidates
HR	Employer	Good workplace	Employer awards, transparent pay	Audits and employee reviews expose false claims
Strategy	Incumbent	Resolve to fight entry	Capacity investment	Sunk costs are lost if the firm retreats
Platforms	Seller	Reliability	Verified badges, long review history	Reputation takes time to build and is lost by cheating

BUSINESS CASE

Venture funding and the start-up signaling stack. An early-stage start-up has little revenue history, so investors rely heavily on signals: the founders' track records, affiliation with selective accelerators, patents, and the identity of earlier investors. Each signal is costly to obtain for a weak venture: an accelerator screens out most applicants. A top-tier lead investor stakes its reputation. A practical implication for founders is **signal fit**: patents are a strong signal of technical novelty but a weak signal of go-to-market ability, whereas paying pilot customers signal demand. Investors, in turn, should ask whether a signal is *differentially* costly: a polished pitch deck is not.

BUSINESS CASE

Illustrative scenario: warranty design. A mid-market appliance maker has a lower failure rate than its rivals but cannot convince retailers. Moving from a one-year to a five-year warranty is a credible signal precisely because it would bankrupt a firm with a high failure rate. A back-of-the-envelope test: expected warranty cost per unit = failure probability × repair cost × years covered. If that cost is small for the focal firm but large for imitators, the warranty separates. If every competitor can match it cheaply, the warranty will be copied and the market

will pool.

8.2.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms that offer longer warranties will receive higher consumer quality ratings, and this effect will be stronger for experience goods than for search goods.
- *H2.* The positive effect of underwriter prestige on IPO valuation will be weaker when analyst coverage of the industry is high (lower information asymmetry).
- *H3.* Signal consistency (low variance across third-party ratings) will strengthen the effect of average rating on sales.

Designs. Field data with plausibly exogenous variation in signal cost (policy changes, certification thresholds). Regression discontinuity around rating cut-offs. Lab or online experiments manipulating signal presence and cost. Structural models that estimate types and equilibria.

Measures. Signal cost (spend, warranty length, credential difficulty). Observability (media mentions, prominence), receiver response (price, choice, valuation, hiring). Information asymmetry (analyst dispersion, firm age, intangibles ratio).

Contexts. IPOs and crowdfunding, online marketplaces and reviews, hiring and credentials, ESG disclosure, product launches.

COMMON PITFALL

Treating every attribute as a signal. Theory requires that the signal be *chosen* and *differentially costly*. If you cannot argue, and ideally show, that low-quality signalers would find the signal more expensive, you are probably studying a correlate of quality, not a signal. State the cost structure explicitly and test for heterogeneity in signal effects across levels of information asymmetry.

8.3 Screening Theory

THEORY CARD

Core claim. When the uninformed party moves first, it can design a menu of contracts that induces informed parties to *self-select*, revealing their type through the option they choose.

Origins. Insurance economics, Rothschild and Stiglitz (1976), Stiglitz (2000).

Level of analysis. Dyad and market (insurer-insured, lender-borrower, employer-applicant, firm-customer).

Key constructs. Self-selection, incentive compatibility, menu of contracts, information rents.

Boundary conditions. The uninformed party can commit to a menu. Types differ in preferences or costs in a way the menu can exploit.

Related theories. Signaling, Agency theory (Section 6.1).

Signaling and screening are mirror images. In **signaling**, the informed party moves first. In

screening, the uninformed party moves first by offering choices. An insurer that offers both a low-premium/high-deductible plan and a high-premium/low-deductible plan lets customers sort themselves: low-risk drivers prefer the high deductible because they expect few claims (Rothschild and Stiglitz 1976) (Table 8.5).

Table 8.5. Signaling versus screening.

Feature	Signaling	Screening
Who moves first	Informed party (signaler)	Uninformed party (screener)
Instrument	Costly action or credential	Menu of contracts or options
Revelation through	Choice of signal	Self-selection among options
Classic model	Spence (1973) job market	Rothschild & Stiglitz (1976) insurance
Business examples	Warranties, degrees, certifications	Deductibles, versioning, probation periods, freemium tiers

BUSINESS CASE

Versioning and freemium. Software firms routinely screen customers by offering tiers: a free plan with limits, a professional plan, and an enterprise plan with security and compliance features. Customers with high willingness to pay self-select into higher tiers because the features they need are only available there. The design problem is **incentive compatibility**: each tier must be the best choice for the segment it targets, or high-value customers will "trade down".

8.4 Costly Signaling

THEORY CARD

Core claim. Signals are honest when they are costly in a way only high-quality senders can afford: the **handicap principle**.
Origins. Evolutionary biology (Zahavi 1975). Anticipated in economics by Veblen's analysis of conspicuous consumption (Veblen 1899).
Level of analysis. Individual (organism or consumer) within a population.
Key constructs. Handicap, honest signal, conspicuous consumption, status.
Related theories. Signaling, Snob effect (Section 29.5), Identity signaling.

Zahavi (1975) proposed that the peacock's tail is an honest signal of fitness *because* it is a handicap: only a healthy bird can survive while carrying it. Veblen had made a parallel argument about the "leisure class", whose conspicuous waste of time and money demonstrated wealth (Veblen 1899). In consumer research, costly signaling explains luxury consumption, charitable giving with public recognition, and expensive rituals that demonstrate commitment.

As Berger (2016) puts it, the more costly something is, the more likely it is to retain its value as a clear signal. Costs need not be monetary: they can be **time** (mastering an esoteric hobby), **opportunity** (forgoing alternatives), or **knowledge** (insider tastes that take years to acquire).

8.5 Identity Signaling

THEORY CARD

Core claim. Consumers diverge from others in product domains that signal identity. When outgroups adopt a taste, ingroup members abandon it to avoid signalling an undesired identity.
Origins. Consumer psychology, Berger and Heath (2007), popularised in Berger (2016).
Level of analysis. Individual consumer within social groups.
Key constructs. Identity-relevance of the domain, observability, outgroup adoption, misidentification cost.
Related theories. Optimal distinctiveness (Section 22.3), Social identity theory (Section 22.2), Snob effect (Section 29.5).

People do not diverge everywhere. Berger and Heath (2007) show that divergence is concentrated in domains seen as **identity-relevant**, music, clothing, hairstyles, cars, rather than in functional domains such as dish soap. Two features make a domain a good identity signal: it is **observable** to others, and it is chosen for its **symbolic** rather than functional value. When a dissimilar outgroup adopts a taste, ingroup members abandon it because continued use risks **misidentification** (Berger and Heath 2007, Berger 2016).

This produces a cycle that marketers know well: identity signaling causes tastes to *catch on* within groups and also to *die out* once they diffuse too broadly. Counterfeits accelerate the cycle by diluting a signal's exclusivity, pushing early adopters toward the next thing (Berger 2016) (Figure 8.4).

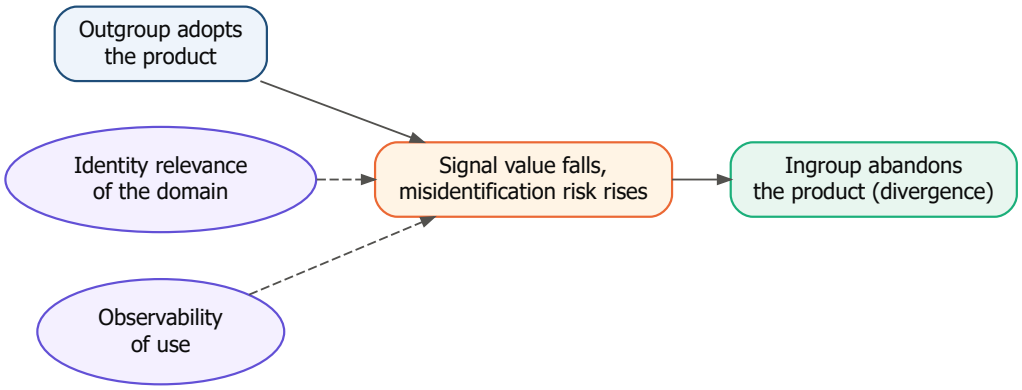


Figure 8.4. Mechanism of identity signaling: outgroup adoption lowers a product's value as an identity signal in observable, identity-relevant domains, prompting ingroup abandonment.

BUSINESS CASE

Managing the fashion cycle. Brands with strong identity value face a paradox: growth dilutes the signal that made them attractive. Common responses include limited editions and "drops" that ration supply, sub-brands for different segments, and pricing architecture that keeps a flagship line exclusive. Each is a way of preserving signal value while expanding reach.

8.6 Countersignaling

THEORY CARD

Core claim. High types sometimes *avoid* the signal that medium types use, because they are confident that other information already separates them from low types.
Origins. Economics, Feltovich, Harbaugh, and To (2002), consumer research on brand prominence, Y. J. Han, Nunes, and Drèze (2010).
Level of analysis. Individual and market.
Key constructs. Low / medium / high types, extraneous information, brand prominence, "quiet" signals.

Feltovich, Harbaugh, and To (2002) explain why the most talented students sometimes do not bother to show off, why the wealthiest dress down, and why the most secure firms forgo flashy marketing. If receivers have other, noisy information that already sets high types apart from low types, medium types still need to signal to separate themselves from low types, but high types can separate from *medium* types by **not** signaling.

Y. J. Han, Nunes, and Drèze (2010) found a consumer analogue in luxury goods. They classified consumers by wealth and need for status. Status-seeking consumers without old wealth ("parvenus") prefer loud, prominently branded goods. Wealthy consumers with a low need for status ("patricians") prefer quiet goods whose subtle cues are recognisable only to fellow insiders. The result is a non-monotonic relationship between price and logo prominence, sketched in Figure 8.5.

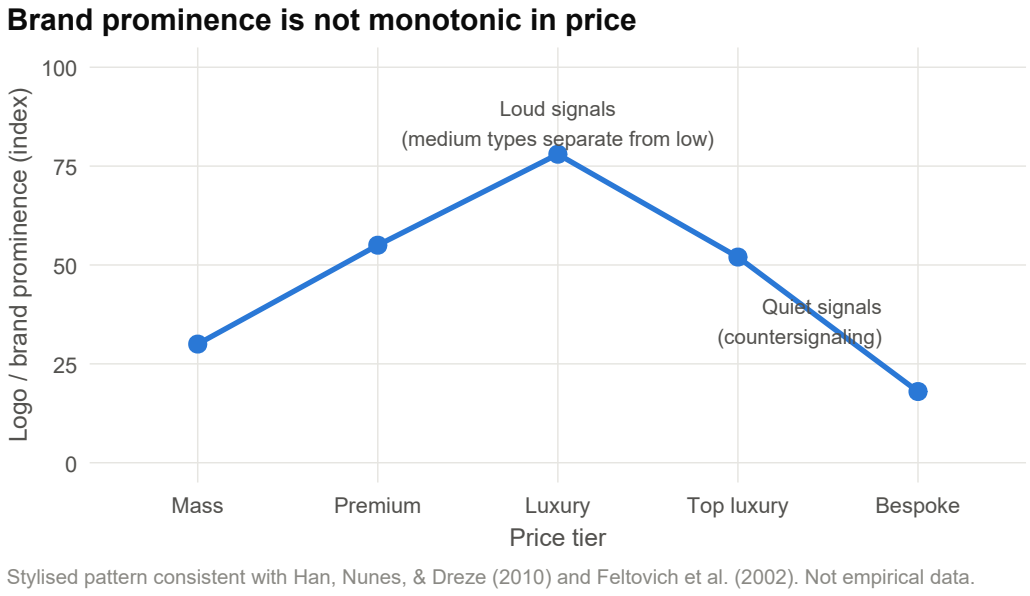


Figure 8.5. Stylised relationship between price tier and brand prominence implied by countersignaling: prominence rises with price for status-seeking buyers but falls at the very top, where quiet signals speak only to insiders.

8.7 Chapter Summary

KEY TAKEAWAYS

- Information asymmetry can unravel markets through **adverse selection** (Akerlof 1970). Remedies include signals, screens, certification, and disclosure.
- A **signal** is a chosen, observable action whose cost is lower for high-quality types. Without differential cost, it cannot separate types (Spence 1973).
- Effective signals are **observable, costly to fake, consistent, and fit** the hidden attribute (Connelly et al. 2011).
- **Screening** inverts the logic: the uninformed party designs a menu so that types self-select (Rothschild and Stiglitz 1976).
- **Costly signaling** and **identity signaling** extend the theory to status and social identity. Tastes catch on and die out as signals diffuse (Berger and Heath 2007).
- **Countersignaling** explains why the highest types sometimes refuse to signal (Feltovich, Harbaugh, and To 2002, Y. J. Han, Nunes, and Drèze 2010).

Table 8.6 presents the comparison of the chapter's theories.

Table 8.6. Comparison of the chapter's theories.

Theory	Who acts	Key mechanism	Signature prediction	Typical business use
Information asymmetry	Market	Adverse selection	High-quality sellers exit	Diagnosing market failure
Signaling	Informed party	Differential cost	Signals separate types	Warranties, credentials, IPO design
Screening	Uninformed party	Self-selection	Menus reveal types	Pricing tiers, insurance, hiring
Costly signaling	Signaler	Handicap	Wasteful displays are honest	Luxury, philanthropy, commitment
Identity signaling	Consumer	Misidentification cost	Divergence after outgroup adoption	Brand and fashion cycles
Countersignaling	High types	Other info already separates	Top types under-signal	Quiet luxury, understated positioning

8.8 Discussion Questions

1. Spence's model implies that education can be valuable even if it teaches nothing. What evidence would distinguish a signaling account from a human-capital account of the college wage premium?
2. Choose a product category you know well. List three signals sellers use and evaluate each on cost, observability, fit, and consistency.
3. Online reviews are often described as solving the lemons problem. Under what conditions might reviews themselves become a pooling signal?
4. Design a screening menu for a B2B software firm with three customer segments. How would you test whether the menu is incentive-compatible?
5. When would you predict *countersignaling* in a labour market? Formulate a testable hypothesis.
6. A firm announces an ESG certification. What would you need to show to argue that the certi-

fication is a signal rather than a correlate of firm quality?

8.9 Further Reading

- Akerlof (1970): the foundational statement of adverse selection. Short and readable.
- Spence (1973): the original signaling model, the logic behind every later application.
- Connelly et al. (2011): the standard review of signaling theory in management, with a useful construct glossary.
- Kirmani and Rao (2000): a marketing-focused review of quality signals and when they work.
- Bergh et al. (2014): a critical assessment of how management research uses (and misuses) signaling theory.
- Y. J. Han, Nunes, and Drèze (2010): an elegant demonstration of countersignaling in luxury consumption.
- Berger (2016): an accessible book on social influence and identity signaling.

Chapter 9

Social Exchange and Networks

Why do employees who feel valued by their employer work harder than their contracts require? Why does a supplier that could earn more by switching customers stay with a long-standing buyer, and why do job seekers so often hear about the position they eventually take from an acquaintance rather than a close friend? Why are some managers, with no formal authority, consistently the source of the ideas that senior executives adopt? Each question concerns resources that do not move through prices or hierarchies alone. They move through **relationships**: through chains of favours, obligations, trust, and information that link people and organizations to one another.

This chapter develops the family of theories that treat social life as **exchange embedded in networks**. We begin with **social exchange theory**, the foundational account of how reciprocity, dependence, and trust sustain relationships between two parties. We then move from the dyad to the structure in which dyads sit: **network exchange theory** explains how a position in a network of exchange opportunities creates power. The **strength of weak ties** and **embeddedness** arguments explain how tie strength shapes the flow of information and trust. **Structural holes** theory explains the advantages of brokerage, and **social capital** theory integrates these ideas into a view of relationships as a productive asset. We close with the **commitment-trust theory** of relationship marketing, which brought the exchange tradition into the study of buyer-seller relationships (Table 9.1).

Table 9.1. Chapter at a glance: theories of social exchange and networks.

Theory	Core question	Level of analysis	Key source
Social exchange theory	Why do people and organizations sustain reciprocal relationships without contracts?	Dyad	Homans (1958), Blau (1964)
Network exchange theory	How does position in a network of exchange opportunities create power?	Network / position	Cook & Emerson (1978), Markovsky et al. (1988)
Strength of weak ties	Why do acquaintances provide novel information?	Dyad / network	Granovetter (1973)
Embeddedness	How do social relations shape economic action?	Dyad / interfirm network	Granovetter (1985), Uzzi (1997)
Structural holes	Why do brokers between groups gain information and control advantages?	Ego network	Burt (1992)
Social capital	How do relationships function as a resource for individuals and organizations?	Individual / group / firm	Coleman (1988), Nahapiet & Ghoshal (1998)

Table 9.1. Chapter at a glance: theories of social exchange and networks. *(continued)*

Theory	Core question	Level of analysis	Key source
Commitment-trust theory	What makes buyer-seller relationships durable and cooperative?	Dyad (relationship)	Morgan & Hunt (1994)

9.1 Social Exchange Theory

THEORY CARD

Core claim. Social behaviour is an exchange of valued resources, material and non-material. Relationships persist when they are rewarding relative to the alternatives, and they deepen when unreciprocated benefits create obligations that are later repaid, building trust and commitment over time.
Origins. Sociology and social psychology: Homans (1958), Blau (1964), J. W. Thibaut and Kelley (1959), and Emerson (1962, 1976). The moral foundation is the norm of reciprocity (Gouldner 1960). Reviewed for management by Cropanzano and Mitchell (2005).
Level of analysis. Primarily the dyad (employee-organization, supervisor- subordinate, buyer-seller, firm-firm), extended to groups and networks.
Key constructs. Rewards and costs, comparison level for alternatives, reciprocity, dependence and power, trust, social versus economic exchange, perceived organizational support.
Boundary conditions. Actors value the resources exchanged, interact repeatedly, and can observe (at least roughly) what each party contributes. Exchange logic weakens where behaviour is governed by strong role obligations, pure altruism, or one-shot interactions with no prospect of repayment.
Related theories. Network exchange theory, Commitment-trust theory, Resource dependence theory (Section 7.1), Equity theory (Section 19.3), Agency theory (Section 6.1).

9.1.1 Origins and intellectual history

Social exchange theory emerged in the late 1950s and early 1960s from several directions at once. Homans (1958) proposed that social behaviour could be analysed as an exchange of goods, both material and non-material, approval and prestige as well as money and services, and that the behavioural regularities of small groups could be derived from elementary principles of reinforcement and cost. At almost the same moment, J. W. Thibaut and Kelley (1959) developed an interdependence framework in which each party evaluates a relationship against a **comparison level** (what one feels one deserves) and a **comparison level for alternatives** (what one could get elsewhere). A relationship can be unsatisfying yet stable if the alternatives are worse, and satisfying yet fragile if better alternatives appear.

Blau (1964) gave the theory its macro-sociological reach. He distinguished **social exchange** from **economic exchange**. Economic exchange specifies in advance what will be exchanged and is enforced by contract. Social exchange entails *unspecified* future obligations whose timing and form are left to the discretion of the party who owes them. Because repayment cannot be enforced, social exchange requires trust, and because it requires trust it also generates trust as obligations are honoured. Blau also showed how exchange produces **power**: when one party provides benefits that

the other needs and cannot obtain elsewhere and cannot repay in kind, the recipient can discharge the obligation only by compliance.

Emerson (1962) formalized that last point in the **power-dependence** principle: the power of A over B equals the dependence of B on A, and B's dependence increases with the value B places on what A provides and decreases with the availability of those resources outside the relationship. Power, in this view, is a property of the *relation*, not of an actor. Emerson also described **balancing operations** through which weaker parties reduce their dependence: withdrawing motivational investment, cultivating alternatives, forming coalitions, or conferring status on the stronger party. In a later review, Emerson (1976) described social exchange less as a single theory than as a frame of reference within which many theories about social interaction can be stated: a characterization that remains apt.

The moral foundation of the framework is the **norm of reciprocity**, which Gouldner (1960) argued is a near-universal component of moral codes. In its generic form it makes two demands: people should help those who have helped them, and people should not injure those who have helped them. The norm stabilizes social systems because it gives actors a reason to initiate exchange with strangers, the recipient of a favour is expected to repay, and because it carries the relationship across periods in which obligations are outstanding.

DEFINITION

Definition 9.1. Social exchange. Voluntary actions of individuals that are motivated by the returns they are expected to bring and typically do bring from others (Blau 1964). Unlike economic exchange, the nature of the return is not specified in advance and cannot be bargained over or enforced.
Norm of reciprocity. A generalized moral norm prescribing that benefits be returned and that benefactors not be harmed (Gouldner 1960).
Dependence. The extent to which an actor's valued outcomes are contingent on another actor's behaviour, increasing in the value of the outcomes and decreasing in the availability of alternative sources (Emerson 1962).

9.1.2 Core logic and propositions

Cropanzano and Mitchell (2005) synthesize the management literature around three ideas: rules of exchange, the resources exchanged, and the relationships that emerge. The canonical sequence begins when one party (an *actor*) treats another favourably or unfavourably. The *target* reciprocates in kind, and if the exchange is successful the relationship moves from a sequence of transactions toward a high-quality relationship characterized by trust, loyalty, and mutual commitment. They also note an important conceptual ambiguity: "social exchange" has been used to refer both to a *type of transaction* (social rather than economic) and to a *type of relationship*, and research is clearer when the two meanings are kept apart.

The core propositions can be stated as follows:

- **P1 (Reciprocity).** Favourable treatment creates a felt obligation to reciprocate. Unfavourable treatment creates an inclination toward negative reciprocity (Gouldner 1960, Cropanzano and Mitchell 2005).
- **P2 (Trust formation).** Because social exchange entails unspecified and unenforceable obligations, trust is both a precondition for and an outcome of successful exchange (Blau 1964).

- **P3 (Dependence and power).** The power of one party over another increases with the other's dependence: the value of the resources received and the scarcity of alternatives (Emerson 1962).
- **P4 (Alternatives and stability).** A relationship is stable when its outcomes exceed the comparison level for alternatives, regardless of whether they meet the comparison level for satisfaction (J. W. Thibaut and Kelley 1959).
- **P5 (Form of exchange).** Reciprocal exchanges, in which benefits are given without negotiated terms, generate more trust and affective commitment than negotiated exchanges, precisely because they expose the giver to risk (Molm, Takahashi, and Peterson 2000, Lawler 2001).

Table 9.2 contrasts the two ideal types of exchange that Blau distinguished.

Table 9.2. Economic versus social exchange (after Blau, 1964).

Dimension	Economic exchange	Social exchange
Obligations	Specified in advance	Unspecified. Left to the discretion of the obligated party
Enforcement	Contract, law, third parties	Norm of reciprocity, reputation, trust
Time horizon	Discrete, often simultaneous	Open-ended, repayment deferred
Resources	Money, goods, defined services	Approval, support, information, loyalty, advice
What it builds	Little beyond the transaction	Trust, personal obligation, commitment
Organizational example	Spot purchase, transactional employment contract	Mentoring, discretionary effort, strategic supplier partnership

9.1.3 Mechanism

Figure 9.1 summarizes the mechanism. An initiating benefit, employer support, a supplier's extra effort, a colleague's favour, creates a felt obligation. Repayment shows that the partner can be trusted, which lowers the perceived risk of further giving and expands the relationship. The cycle can also run in reverse: mistreatment produces negative reciprocity, withdrawal, and eventually exit.

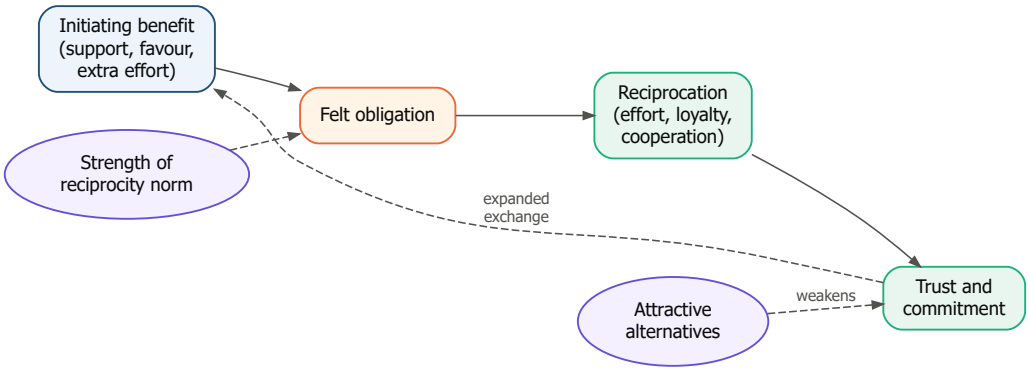


Figure 9.1. The social exchange cycle. Initiating benefits create felt obligation under the norm of reciprocity. Reciprocation builds trust, which encourages further exchange. Moderators shape the strength of each link.

9.1.4 Evidence and extensions

The most influential organizational application is organizational support theory. Eisenberger et al. (1986) proposed that employees form global beliefs about the extent to which the organization values their contributions and cares about their well-being, **perceived organizational support (POS)**, and that they reciprocate high support with greater effort and attendance. Employees personify the organization, attributing the actions of its agents to the organization itself. In a review of the subsequent literature, Rhoades and Eisenberger (2002) found that fairness, supervisor support, and favourable rewards and job conditions were associated with POS, which was in turn associated with affective commitment, performance, and reduced withdrawal. Settoon, Bennett, and Liden (1996) showed that the exchange with the organization (POS) and the exchange with the immediate supervisor (leader-member exchange (Section 15.5)) are distinct: the former was more closely tied to organizational commitment (Section 15.10), the latter to in-role performance and citizenship behaviour.

Experimental sociology has refined the theory's account of trust. Molm, Takahashi, and Peterson (2000) compared **negotiated** exchange, in which partners bargain jointly over binding terms, with **reciprocal** exchange, in which each gives unilaterally without knowing whether the other will return the favour. Trust was higher under reciprocal exchange: consistent with Blau's argument that the risk of non-reciprocation is what allows partners to demonstrate trustworthiness. Lawler (2001) added an affective layer: successful exchange produces positive emotions, and when actors attribute those emotions to the relationship or group, they develop affective commitment to it.

DEBATE AND EVIDENCE

Is social exchange theory too elastic? Almost any behaviour can be described after the fact as an "exchange" of some reward, which risks making the theory unfalsifiable. Cropanzano and Mitchell (2005) recommend precision: specify the *rule* of exchange (reciprocity, negotiation, or another rule), the *resource* exchanged, and whether the claim concerns a transaction or a relationship. A second debate pits **self-interest against norms**: Homans's version rests on individual reinforcement, Gouldner's on a moral norm, and the two diverge when repayment is unobserved or the giver cannot retaliate.

9.1.5 Applications in business

BUSINESS CASE

Employee reciprocity and perceived support. Organizational support theory gives HR leaders a concrete diagnostic. Actions that are *discretionary*, not required by law, contract, or union agreement, are especially informative to employees about the organization's regard for them, and thus especially likely to trigger reciprocation (Rhoades and Eisenberger 2002). Flexible arrangements offered during a family emergency, investment in an employee's development, or fair treatment in a restructuring signal that the organization values the individual. By contrast, benefits that employees attribute to external pressure generate less obligation. A practical implication is that the *framing and discretion* of support matter as much as its monetary value.

BUSINESS CASE

Illustrative scenario: dependence in a supplier relationship. A component maker that sells most of its output to one automotive assembler is highly dependent, so the assembler holds power (Emerson 1962). Emerson's balancing operations suggest remedies: cultivating customers in adjacent industries, joining an industry consortium, or developing process know-how the assembler cannot easily replace. The same logic underlies resource dependence theory (Section 7.1) at the level of the firm.

9.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Perceived organizational support is positively related to discretionary effort, and the relationship is stronger for employees high in exchange ideology.
- *H2.* Unilateral (non-negotiated) concessions by a supplier increase buyer trust more than equivalent concessions obtained through negotiation.
- *H3.* A partner's dependence on a focal firm increases the focal firm's ability to extract favourable terms, but this effect is weaker when the relationship has a long history of balanced reciprocity.

Designs. Multi-wave employee surveys with lagged outcomes. Supervisor-rated performance to reduce common-method bias. Laboratory exchange experiments that manipulate negotiated versus reciprocal exchange. Dyadic buyer-supplier data with both sides reporting.
Measures. Survey of Perceived Organizational Support (Eisenberger et al. 1986). Leader-member exchange scales. Dependence measured as sales share and switching costs. Trust and commitment scales from relationship-marketing research (R. M. Morgan and Hunt 1994).
Contexts. Employment relationships, gig and platform work, supplier partnerships, franchise systems, professional services.

COMMON PITFALL

Labelling every positive association "social exchange". A correlation between POS and performance does not by itself show reciprocation. The same pattern could arise because high performers receive more support. Establish temporal order, model the reciprocal path, and, where possible, measure felt obligation as the mediating mechanism rather than assuming it.

9.2 Network Exchange Theory

THEORY CARD

Core claim. Power in exchange relations depends not only on dyadic dependence but on the *structure* of the network of exchange opportunities: positions with many alternatives connected to partners with few alternatives can exclude others and extract favourable terms.
Origins. Experimental sociology: K. S. Cook and Emerson (1978) and K. S. Cook et al. (1983), building on Emerson (1962, 1976). The graph-theoretic approach of Markovsky, Willer, and Patton (1988) and colleagues (Willer 1999). K. S. Cook and Whitmeyer (1992) compare

exchange-theoretic and network-analytic approaches to social structure.
Level of analysis. Positions within exchange networks (individuals, firms, or organizational units).
Key constructs. Negatively versus positively connected exchange relations, exclusion, alternatives, graph-theoretic power index (GPI), strong- and weak-power structures, equity and commitment.
Boundary conditions. Actors seek favourable terms. Exchange relations are *negatively connected* (exchange with one partner reduces exchange with another). The number of exchanges per round is limited.
Related theories. Social exchange theory, Structural holes, Resource dependence theory (Section 7.1).

9.2.1 Core logic

Network exchange theory asks what happens when Emerson's alternatives are themselves actors embedded in a larger structure. K. S. Cook and Emerson (1978) created laboratory networks in which participants negotiated divisions of a fixed resource pool with the partners to whom they were connected. They found that positions with more and better alternatives obtained more favourable divisions, but also that concerns about **equity** and the development of **commitment** between particular partners dampened the use of power: participants who formed committed relationships often forwent better offers from others.

A key distinction is between **negatively connected** and **positively connected** relations. Two exchange relations are negatively connected if exchange in one reduces exchange in the other (a buyer who can purchase from only one of two suppliers in a period). They are positively connected if exchange in one increases exchange in the other (a supply chain in which a distributor's purchases enable its sales). Most power predictions concern negatively connected networks, where actors with alternatives can play partners off against one another.

K. S. Cook et al. (1983) showed that in negatively connected networks, standard **centrality** measures do not reliably predict power. The most central position in a network may be weak if its partners themselves have attractive alternatives. Markovsky, Willer, and Patton (1988) proposed a different approach: power derives from the capacity to **exclude** others from exchange. Their **graph-theoretic power index (GPI)** counts, for each position, the number of non-intersecting paths of each length that radiate from it, adding odd-length paths (which provide alternatives) and subtracting even-length paths (which indicate that one's alternatives have alternatives of their own). Positions with higher GPI than their partners are predicted to exercise power.

DEFINITION

Definition 9.2. Graph-theoretic power index. For position i, let m_k be the maximum number of non-intersecting paths of length k that begin at i. Then

$$\mathrm{GPI}_i = m_1 - m_2 + m_3 - m_4 + \cdots = \sum_{k \geq 1} (-1)^{k+1} m_k.$$

The index formalizes the intuition that a direct partner is an advantage, while a partner who has other partners is a source of competition (Markovsky, Willer, and Patton 1988).

9.2.2 Mechanism

Figure 9.2 computes the GPI for four simple networks in which each position can complete at most one exchange per round. In the **3-line**, B can always exclude either A or C, so B is strong. In the **5-line**, B and D are strong because each has a peripheral partner (A or E) with no alternative, whereas the *central* position C is weak: its partners B and D have better options. In the **4-line**, all positions have equal GPI, so the theory predicts roughly equal divisions. In the **star**, the hub can exclude any spoke and is predicted to extract most of the surplus.

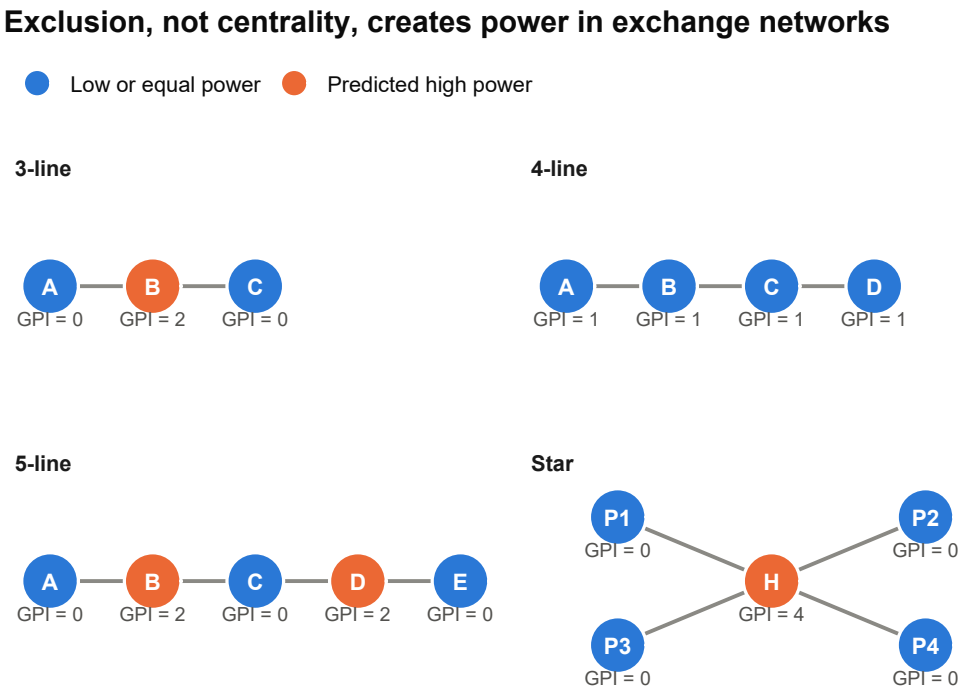


Figure 9.2. Power in four exchange networks, computed with the graph-theoretic power index (GPI) of Markovsky et al. (1988). Orange positions have higher GPI than their partners and are predicted to exercise power. Note that the central position C in the 5-line is weak.

Table 9.3 compares the GPI with two familiar centrality measures. Degree and betweenness centrality both rank the 5-line's middle position C as at least as important as B and D, yet the GPI, and the experimental evidence that motivated it, predicts that C is weak. The contrast illustrates the central lesson of the network exchange tradition: **power in negatively connected networks is about alternatives and exclusion, not connectedness per se** (K. S. Cook et al. 1983, Markovsky, Willer, and Patton 1988).

Table 9.3. Centrality versus the graph-theoretic power index in four exchange networks (computed).

Network	Position	Degree	Betweenness (normalized)	GPI
3-line	A	1	0.00	0
3-line	B	2	1.00	2
3-line	C	1	0.00	0
4-line	A	1	0.00	1

Table 9.3. Centrality versus the graph-theoretic power index in four exchange networks (computed). *(continued)*

Network	Position	Degree	Betweenness (normalized)	GPI
4-line	B	2	0.67	1
4-line	C	2	0.67	1
4-line	D	1	0.00	1
5-line	A	1	0.00	0
5-line	B	2	0.50	2
5-line	C	2	0.67	0
5-line	D	2	0.50	2
5-line	E	1	0.00	0
Star	H	4	1.00	4
Star	P1	1	0.00	0
Star	P2	1	0.00	0
Star	P3	1	0.00	0
Star	P4	1	0.00	0

9.2.3 Evidence and extensions

Subsequent work distinguished **strong-power** structures (such as the 3-line and the star), where some positions can be excluded without cost to the powerful, from **weak-power** structures, in which no position can be excluded for certain and power differences are smaller and more sensitive to actors' strategies (Willer 1999). K. S. Cook and Whitmeyer (1992) argue that exchange theory and network analysis are complementary: exchange theory supplies a behavioural model of why actors use ties, and network analysis supplies the structural description of which ties are available. Bonacich (1987) made a related point for centrality measures more generally: being connected to well-connected others can *reduce* power in bargaining situations, so a centrality index should allow the contribution of a neighbour's connections to be negative.

BUSINESS CASE

Sales networks and channel power. A distributor that is the only route between a group of small manufacturers and a group of small retailers occupies a strong-power position analogous to the hub of a star: it can play each manufacturer off against the others. If the manufacturers gain direct access to retailers through an online marketplace, the network shifts toward a weak-power structure, and the distributor's margins compress even though its *number* of connections has not changed. Network exchange theory directs managers to ask not "how many partners do we have?" but "how many alternatives do our partners have?"

9.3 Strength of Weak Ties and Embeddedness

THEORY CARD

Core claim. Weak ties (acquaintances) are more likely than strong ties (close friends) to bridge otherwise disconnected social circles, so they are disproportionately important sources of novel information. Strong, embedded ties, by contrast, are better at conveying trust,

fine-grained information, and cooperation.
Origins. Economic sociology: M. S. Granovetter (1973, 1974) on weak ties, M. Granovetter (1985) on embeddedness, Uzzi (1997) on the paradox of embeddedness.
Level of analysis. Dyadic ties and the network structures they form. Individuals and interfirm networks.
Key constructs. Tie strength, bridge and local bridge, triadic closure, novel versus redundant information, embeddedness, arm's-length ties.
Boundary conditions. Weak ties help when the valuable resource is *information* that is unevenly distributed across groups and can be transmitted cheaply. When the resource requires trust, influence, or tacit knowledge, strong ties may matter more.
Related theories. Structural holes, Social capital, Diffusion of innovations (Section 10.1).

9.3.1 Origins and core logic

M. S. Granovetter (1973) defined tie strength as a combination of the time spent together, emotional intensity, mutual confiding, and reciprocal services. His argument begins from a structural regularity. If A has strong ties to both B and C, then B and C are likely to know each other too, because they spend time with A, tend to be similar to A, and would experience strain if they disliked one another. Granovetter called the configuration in which A is strongly tied to both B and C but B and C have no tie at all the **forbidden triad**, because it should rarely occur.

The consequence is striking. A **bridge** is a tie that provides the only path between two parts of a network. If strong ties tend to close triads, then a strong tie can rarely be a bridge: the two people will typically share other contacts. Bridges are therefore almost always weak ties. Because weak ties connect us to people who move in different circles, they deliver information we would not otherwise receive. Our close friends, by contrast, tend to know what we already know.

- **P1 (Triadic closure).** The stronger the ties between A-B and A-C, the more likely a tie between B and C.
- **P2 (Bridging).** Local bridges are disproportionately weak ties.
- **P3 (Novel information).** Information obtained through weak ties is more likely to be nonredundant than information obtained through strong ties.
- **P4 (Opportunity).** Individuals with more weak ties have access to a wider set of opportunities (jobs, ideas, deals).

M. S. Granovetter (1974) provided the canonical evidence in *Getting a Job*: among professional, technical, and managerial workers who found jobs through personal contacts, those contacts were more often people they saw occasionally or rarely than people they saw often. The nuance matters: weak ties are valuable not because weakness is good in itself, but because weak ties are more likely to be *bridges*.

Table 9.4 summarizes the complementary strengths of strong and weak ties.

Table 9.4. Strong and weak ties: complementary strengths.

Dimension	Strong ties	Weak ties
Typical relationship	Close friends, family, long-standing partners	Acquaintances, former colleagues, occasional contacts
Network position	Embedded in dense clusters (closed triads)	More often bridges between clusters
Information	Redundant but rich, trusted, tacit	Novel but thin, codified

Table 9.4. Strong and weak ties: complementary strengths. *(continued)*

Dimension	Strong ties	Weak ties
Best for	Trust, influence, transfer of complex knowledge, support	Search, discovery, job leads, new ideas
Business example	Key-account relationship, trusted supplier	Referral to a new market, alumni network lead
Key evidence	Uzzi (1997), Hansen (1999), Bian (1997)	Granovetter (1973, 1974), Rajkumar et al. (2022)

9.3.2 Evidence and extensions

The most rigorous test to date exploited randomized variation in LinkedIn's "People You May Know" recommendation algorithm. Rajkumar et al. (2022) analysed large-scale experiments in which the algorithm's variants changed the mix of strong and weak ties that users formed, and linked those ties to subsequent job mobility. Because the variation in network formation was experimentally induced, the study could make causal claims that earlier observational studies could not. The headline result refines Granovetter's thesis: the relationship between tie strength and job transmission was **non-linear (an inverted U)**. Moderately weak ties, for example, contacts with a modest number of mutual connections, were the most helpful for job mobility, while the very weakest ties (with few or no mutual connections) and strong ties were less helpful. The value of weak ties was also greater in more digital industries. Figure 9.3 sketches the qualitative shape.

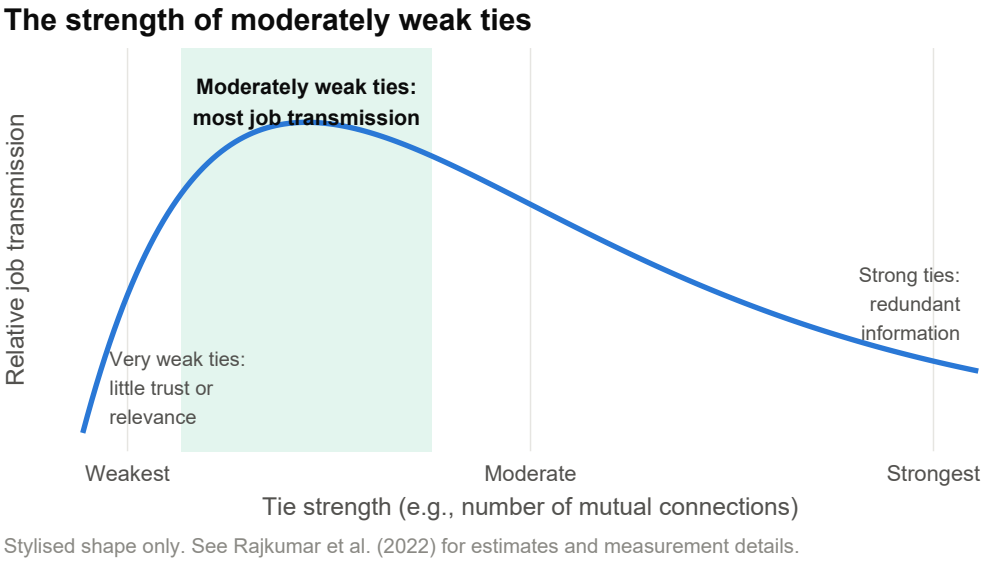


Figure 9.3. Stylised representation of the inverted-U relationship between tie strength and job transmission reported by Rajkumar et al. (2022). The curve is illustrative and does not reproduce the study's estimates.

Several other lines of evidence qualify the weak-ties thesis. Bian (1997) found that in a Chinese labour market where jobs were allocated through personal influence, strong ties, including indirect ties reached *through* strong ties, were more useful, because what job seekers needed was influence rather than information. Hansen (1999) identified a **search-transfer problem** inside a multinational firm: weak inter-unit ties helped project teams *search* for useful knowledge but impeded the

transfer of complex, tacit knowledge, which required strong ties. And McPherson, Smith-Lovin, and Cook (2001) documented the pervasive **homophily (Section @ref(homophily (Section 30.2)))** of social networks: because people tend to connect with similar others, even weak ties may carry less diversity than the theory assumes, and access to valuable weak ties is unequally distributed.

BUSINESS CASE

Referral hiring. Employee referral programmes institutionalize tie-based job matching. Studying a bank's telephone customer-service centre, Fernandez, Castilla, and Moore (2000) found that referral hiring produced economic returns for the employer, partly because referrals enriched the applicant pool and improved the match between new hires and jobs. Burks et al. (2015) analysed personnel data from firms in call-centre, trucking, and high-tech industries and found that referred workers were less likely to quit and were, on balance, more profitable to the employer than non-referred workers. Referral programmes therefore exploit *both* tie properties: weak ties spread information about openings, while the referrer's strong tie to the employer provides a screening and mentoring channel. A practical caution follows from homophily: referral hiring tends to reproduce the demographic composition of the existing workforce, which can conflict with diversity objectives unless it is complemented by other channels.

TRY IT YOURSELF

List the last three significant pieces of professional information you received: a job lead, a client introduction, an idea you adopted. For each, rate the tie strength to the person who gave it to you and note whether that person belongs to a different social circle from your closest colleagues. Does your experience fit the bridging argument, Bian's influence argument, or Hansen's search-transfer argument best?

9.3.3 Embeddedness and its paradox

M. Granovetter (1985) broadened the argument into a critique of economic theory. Economic action is neither the atomized behaviour of neoclassical models (an *under-socialized* view) nor the mechanical enactment of internalized norms (an *over-socialized* view). It is **embedded** in concrete, ongoing networks of personal relations. Those relations generate trust and discourage malfeasance, because actors prefer partners of known reputation and because ongoing relationships carry expectations beyond the contract. The argument was directed in part at transaction cost economics (Section 6.3), which attributed order in exchange primarily to hierarchical governance.

Uzzi (1997) studied the networks of manufacturers and contractors in the New York apparel industry and identified three features of **embedded ties**: trust, fine-grained information transfer, and joint problem-solving arrangements. Embedded ties let firms respond quickly to changing fashions and coordinate without costly contracting. But Uzzi also identified a **paradox of embeddedness**: networks made up *only* of embedded ties insulated firms from new information and left them vulnerable if a key partner failed. Firms that *combined* embedded and arm's-length ties were more likely to survive than those at either extreme: an inverted-U relationship.

BUSINESS CASE

Alliance portfolios. Uzzi's paradox has a direct analogue in alliance strategy. A firm whose

partnerships are all deep, exclusive, and long-standing benefits from trust and joint problem-solving but may miss technological shifts that originate outside its circle. A firm whose partnerships are all transactional enjoys flexibility but cannot coordinate complex joint work. The relational view of strategy argues that interfirm advantage arises from relation-specific assets, knowledge-sharing routines, complementary resources, and effective governance (Dyer and Singh 1998), and that these typically require embedded ties. Portfolio design is therefore a matter of balance: a core of embedded partnerships for execution, surrounded by weaker, more diverse ties for search (Gulati 1998).

COMMON PITFALL

Equating "weak" with "good". The weak-ties argument concerns *bridging*. A weak tie inside one's own clique adds little, and a strong tie that happens to span two groups can be extremely valuable. Measure the structural position of ties (bridging, redundancy) rather than relying on tie strength as a proxy, and specify which resource, information, influence, trust, or tacit knowledge, the tie is supposed to carry.

9.4 Structural Holes and Brokerage

THEORY CARD

Core claim. People and organizations whose contacts are disconnected from one another, who span **structural holes**, gain information and control advantages, because they are positioned to see and combine non-redundant information and to broker between parties who cannot coordinate directly.
Origins. Sociology of competition: Burt (1992, 2004, 2005), building on M. S. Granovetter (1973) and on the sociology of brokerage.
Level of analysis. The ego network of an individual, team, or organization.
Key constructs. Structural hole, non-redundant contacts, network constraint, effective size, brokerage, betweenness, *tertius gaudens* (the third who benefits) and *tertius iungens* (the third who joins).
Boundary conditions. Advantages accrue when information is unevenly distributed across groups, when the broker can act on it, and when the context rewards novelty or arbitrage. In settings that require coordination, trust, or norm enforcement, closure may be more valuable.
Related theories. Weak ties, Social capital, Network exchange theory, Absorptive capacity (Section 5.2).

9.4.1 Core logic

Burt (1992) shifted attention from the strength of a tie to the **structure** of a person's network. A **structural hole** is the absence of a tie between two of a person's contacts: a gap between non-redundant sources of information. Contacts are redundant when they are connected to one another (cohesion) or when they connect to the same third parties (structural equivalence). A network rich in structural holes provides two kinds of benefit:

- **Information benefits.** Access to more diverse information, earlier (timing), and through contacts who refer the broker to opportunities (referrals).

- **Control benefits.** The ability to broker between parties who cannot coordinate directly: the *tertius gaudens*, or "third who benefits", who plays one party against another.

Burt connects the argument to Granovetter's: weak ties often span structural holes, but it is the hole, not the weakness of the tie, that matters.

DEFINITION

Definition 9.3. Network constraint. Let p_{ij} be the proportion of i's network time and energy invested in contact j. Burt's constraint index is

$$c_i = \sum_{j \neq i} \left(p_{ij} + \sum_{q \neq i,j} p_{iq}\, p_{qj} \right)^2 ,$$

which is high when i's contacts are few and interconnected (little brokerage) and low when they are many and disconnected (much brokerage) (Burt 1992). **Betweenness centrality** measures the share of shortest paths between other pairs that pass through i (L. C. Freeman 1977).

9.4.2 Mechanism

Figure 9.4 illustrates the logic with a stylised network of three tightly knit groups. Robert holds one tie into each group, and those groups are otherwise disconnected. James is embedded in one group and connected only to people who are connected to each other: in the spirit of the contrast Burt draws between brokered and closed networks (Burt 2005). Both indices are computed directly from the graph with `igraph`.

Table 9.5. Brokerage indices for selected nodes in the stylised network (computed with igraph).

Node	Degree	Betweenness (normalized)	Constraint
Robert	3	0.718	0.333
A1	4	0.462	0.484
B1	4	0.385	0.583
C1	4	0.385	0.583
A2	4	0.043	0.701
James	3	0.000	0.750
B2	3	0.000	0.866

Table 9.5 shows the numbers behind the figure. Robert has fewer contacts than James's neighbours but the highest betweenness and the lowest constraint: every path between the three groups runs through him. James has as many contacts as Robert, but because they are all tied to one another, his constraint is high and his betweenness is close to zero. Degree alone would not distinguish their positions.

9.4.3 Evidence and extensions

Burt (2004) offered one of the most cited tests. Studying supply-chain managers in a large American electronics company, he found that managers whose networks spanned structural holes were more likely to express ideas that senior managers judged valuable, and that brokerage was associ-

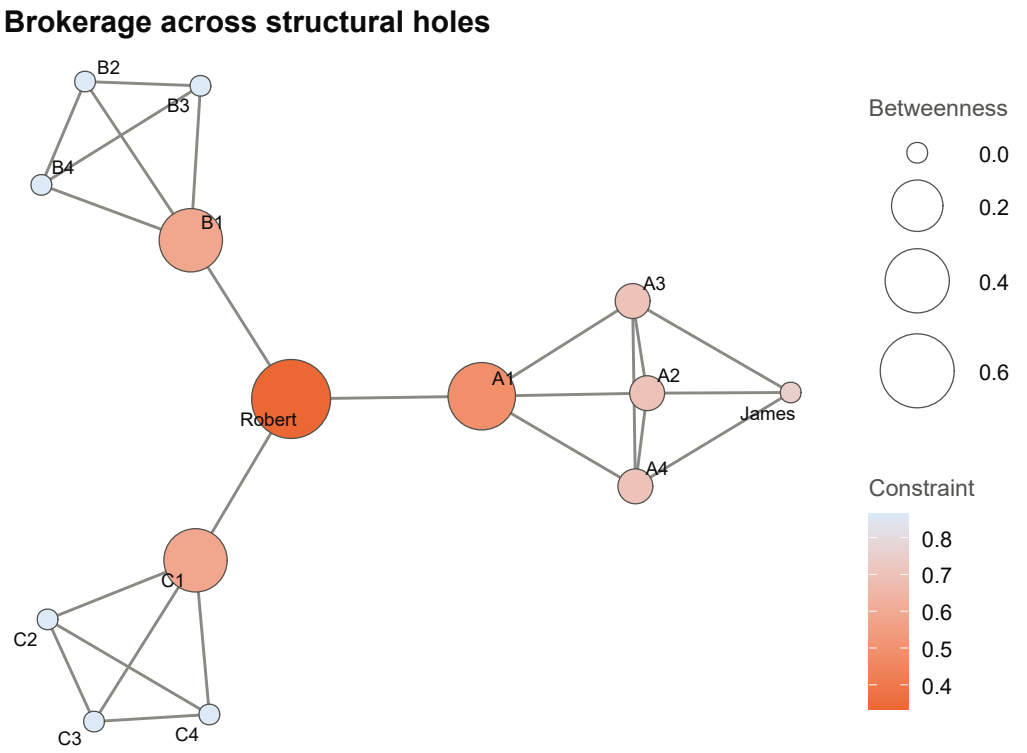


Figure 9.4. A broker spanning structural holes (stylised network). Node size shows betweenness centrality and colour shows Burt's network constraint, both computed with igraph. Robert, who bridges three otherwise disconnected groups, has high betweenness and low constraint. James, embedded in a single closed group, has the reverse.

ated with more favourable compensation, evaluations, and promotion. The mechanism is a **vision advantage**: people at the boundaries between groups are exposed to alternative ways of thinking and can translate practices from one group to another.

The brokerage argument has generated important counter-evidence and boundary conditions:

- **Closure for innovation.** In a longitudinal study of collaboration networks in the chemicals industry, Ahuja (2000) found that direct and indirect ties increased firms' innovation output, but structural holes *reduced* it: consistent with the view that collaborative innovation requires the trust and norms that closure provides.
- **Cohesion and range.** Reagans and McEvily (2003) found that both network cohesion around a tie and the range of a person's network eased knowledge transfer: cohesion increases willingness to help, while range increases the ability to convey knowledge to diverse audiences.
- **Joining versus separating.** Obstfeld (2005) proposed the *tertius iungens* orientation, brokers who introduce disconnected parties to one another rather than exploiting the gap, and found it associated with involvement in innovation in an engineering setting.
- **Diversity-bandwidth trade-off.** Aral and Van Alstyne (2011) showed that structurally diverse networks tend to have lower-bandwidth ties, so the novelty that brokerage provides comes at the cost of less information flow through each tie.

DEBATE AND EVIDENCE

Brokerage or closure? The structural-holes and closure arguments make opposing predictions about the same network feature, and both have empirical support. Burt (2005) proposed a synthesis: closure *within* a group creates the trust and coordination needed to deliver value, while brokerage *across* groups creates the novelty that defines what value is. Performance is highest when a group is internally cohesive and externally well bridged. A second debate concerns **selection**: people with distinctive skills or personalities may both create brokerage positions and perform well, so causal claims require longitudinal or quasi-experimental designs (Borgatti et al. 2009).

BUSINESS CASE

Illustrative scenario: the cross-functional product manager. A product manager sits between engineering, sales, and customer-success groups that rarely talk to one another. She hears customer complaints before engineering does and can frame engineering constraints in commercial terms. Structural-holes theory predicts that she will generate more valuable ideas than single-function peers, and that her advantage would erode if the groups began to communicate directly. The firm must choose between preserving the brokerage role and closing the hole with formal coordination.

9.4.4 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Employees with lower network constraint will generate ideas rated as more valuable by independent evaluators.
- *H2.* The positive effect of brokerage on idea quality will be stronger in units facing high environmental turbulence.
- *H3.* Firms whose alliance portfolios combine a cohesive core of partners with bridging ties to distant industries will produce more novel patents than firms with portfolios that are uniformly dense or uniformly sparse.

Designs. Sociometric surveys ("whom do you go to for advice?"). Network reconstruction from email, collaboration platform, co-authorship, patent co-invention, or alliance databases. Panel models with ego fixed effects to address selection. Randomized introductions or team assignments where feasible.

Measures. Constraint, effective size, and betweenness (Burt 1992, L. C. Freeman 1977). Brokerage roles, tie strength as frequency or duration. Outcomes such as idea ratings, promotion, patent citations, or deal flow (Wasserman and Faust 1994). In R, `igraph` computes all of these indices.

9.5 Social Capital

THEORY CARD

Core claim. Relationships are a resource. The structure and content of an actor's social relations provide access to information, influence, and solidarity that facilitate action and produce returns for individuals, groups, and organizations.
Origins. Sociology (Coleman 1988, Portes 1998, Lin 2001), political science (Putnam 2000), and management (Nahapiet and Ghoshal 1998, Adler and Kwon 2002).
Level of analysis. Individuals, groups, organizations, and communities.
Key constructs. Closure, brokerage, bonding and bridging, structural, relational, and cognitive dimensions, obligations and expectations, norms and sanctions, information channels.
Boundary conditions. Social capital is context-specific and not fully fungible: ties that help in one task may hinder in another. It has costs, and its benefits may be appropriated by some at the expense of others.
Related theories. Weak ties, Structural holes, Resource-based theory (Section 4.1), Knowledge-based view (Section 4.3).

9.5.1 Two views: closure and brokerage

Coleman (1988) defined social capital by its function: aspects of social structure that facilitate certain actions of actors within the structure. He identified three forms, **obligations and expectations** (which depend on the trustworthiness of the social environment), **information channels**, and **norms with effective sanctions**, and argued that **closure**, a network in which one's contacts are connected to one another, strengthens all three. In a closed network, a broken promise becomes known to everyone and can be collectively sanctioned. Coleman illustrated the point with the wholesale diamond market in New York, where merchants could hand bags of stones to one another for inspection without formal insurance because the community was dense, and with evidence that *intergenerational* closure among parents was associated with lower dropout rates from school.

Burt's argument, by contrast, locates social capital in the *absence* of ties among one's contacts. Once the resource is specified, the views are compatible: closure produces trust and norm enforcement, brokerage produces novel information. Putnam's distinction between **bonding** social capital (within homogeneous groups) and **bridging** social capital (across groups) captures the same duality at the community level (Putnam 2000). Table 9.6 compares the two views.

Table 9.6. Two views of social capital: closure versus brokerage.

Dimension	Closure (bonding)	Brokerage (bridging)
Key proponents	Coleman (1988), Putnam (2000), bonding	Burt (1992, 2004), Granovetter (1973)
Network structure	Dense, contacts connected to one another	Sparse, contacts disconnected (structural holes)
Primary benefit	Trust, norm enforcement, coordination, support	Novel information, early access, control
Mechanism	Reputation and collective sanctions	Non-redundant information and arbitrage
Typical outcome	Cooperation, reliable execution, reduced opportunism	Creativity, innovation, career advancement
Risks	Insularity, conformity, lock-in	Distrust, role strain, value capture by the broker

Table 9.6. Two views of social capital: closure versus brokerage. *(continued)*

Dimension	Closure (bonding)	Brokerage (bridging)
Business example	Close-knit supplier community, tight project team	Boundary-spanning manager, diverse alliance portfolio

9.5.2 Three dimensions of organizational social capital

Nahapiet and Ghoshal (1998) brought social capital into strategic management, arguing that organizations have an advantage over markets in creating and sharing **intellectual capital** because they are dense with social capital. They distinguished three dimensions (Table 9.7): the **structural** dimension (the pattern of connections), the **relational** dimension (the quality of relationships: trust, norms, obligations, identification), and the **cognitive** dimension (shared codes, language, and narratives). Social capital facilitates the combination and exchange of knowledge, which in turn creates new intellectual capital (Figure 9.5).

Table 9.7. Three dimensions of social capital.

Dimension	What it captures	Facets (Nahapiet & Ghoshal, 1998)	Example measure
Structural	Who is connected to whom, and how	Network ties, network configuration, appropriable organization	Density, centrality, constraint
Relational	The quality of the relationships	Trust, norms, obligations and expectations, identification	Trust scales, reciprocity norms
Cognitive	Shared understanding that enables communication	Shared codes and language, shared narratives	Shared vision, common terminology

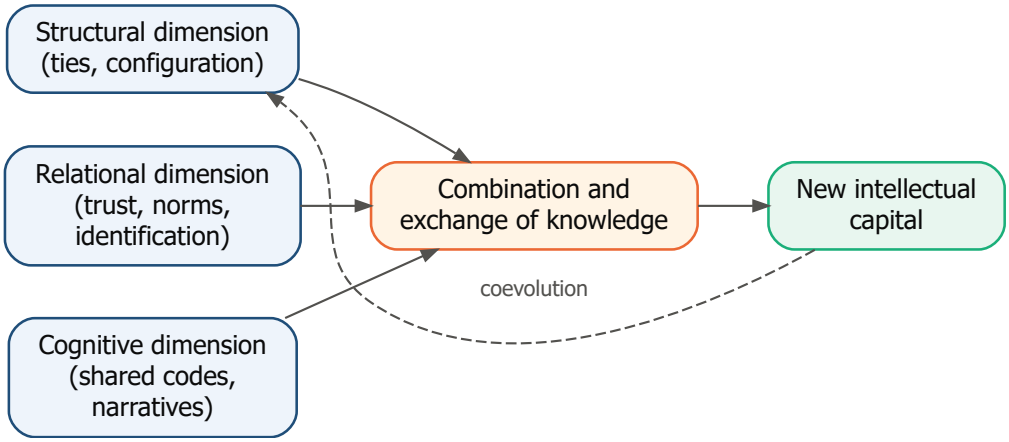


Figure 9.5. Social capital and the creation of intellectual capital, following Nahapiet and Ghoshal (1998). The three dimensions of social capital enable the combination and exchange of knowledge, which creates new intellectual capital.

Adler and Kwon (2002) define social capital as the goodwill available to individuals or groups, located in the structure and content of their social relations. They organize its sources into **opportunity** (the network ties), **motivation** (the willingness of contacts to help), and **ability** (the competencies and resources of contacts), and they emphasize that social capital involves costs and risks

as well as benefits. Portes (1998) catalogued its negative consequences: exclusion of outsiders, excessive claims on successful group members, restrictions on individual freedom, and downward-levelling norms.

DEBATE AND EVIDENCE

Concept or metaphor? Because the term has been stretched to cover networks, trust, norms, and their outcomes, defining social capital by what it achieves risks tautology. Portes (1998) and Adler and Kwon (2002) recommend separating its *sources* (the structure and content of ties) from its *benefits* and *risks*. Lin (2001) makes the separation explicit by defining social capital as resources embedded in a network and accessed or mobilized through ties.

BUSINESS CASE

B2B key-account relationships. Key-account management converts individual relationships into organizational social capital. A key-account team typically deliberately builds all three dimensions: *structural*, multiple contact points across the client's functions, so that the relationship does not depend on a single salesperson. *Relational*, trust built through joint planning and problem-solving. *Cognitive*, a shared vocabulary about the client's business and metrics. Closure around the account protects the relationship against the departure of any one person, while bridging ties to the supplier's engineering groups bring the client's problems to those who can solve them.

9.6 Commitment-Trust Theory of Relationship Marketing

THEORY CARD

Core claim. Successful relational exchange requires **commitment** and **trust**. They are the key mediating variables between a relationship's antecedents (benefits, termination costs, shared values, communication, opportunism) and its outcomes (cooperation, acquiescence, reduced propensity to leave).
Origins. Marketing: R. M. Morgan and Hunt (1994), building on social exchange theory and on models of buyer-seller relationship development (Dwyer, Schurr, and Oh 1987).
Level of analysis. The relationship between exchange partners: suppliers, customers, channel members, and internal partners.
Key constructs. Relationship commitment, trust, relationship termination costs, relationship benefits, shared values, communication, opportunistic behaviour, cooperation, acquiescence, propensity to leave, functional conflict, decision-making uncertainty.
Boundary conditions. Most applicable to relational (rather than discrete) exchange, where future interaction is expected and partners are vulnerable to one another. Its predictions are weaker for low-involvement, transactional customers.
Related theories. Social exchange theory, Social capital, Transaction cost economics (Section 6.3).

9.6.1 Core logic and the KMV model

R. M. Morgan and Hunt (1994) argued that relationship marketing, establishing, developing, and maintaining successful relational exchanges, had become central to marketing, and that its success

depended on two constructs. **Relationship commitment** is an exchange partner's belief that an ongoing relationship is so important that it warrants maximum effort to maintain it. **Trust** exists when one party has confidence in an exchange partner's reliability and integrity. They proposed that commitment and trust are **key mediating variables (KMV)**: rather than relying on power to condition others, firms build relationships in which partners choose to cooperate. They identified ten forms of relational exchange spanning supplier, lateral (competitors, non-profit organizations, government), buyer, and internal (business units, employees, functional departments) partnerships, and tested the model with independent automobile tyre retailers' relationships with their suppliers.

Figure 9.6 reproduces the structure of the model.

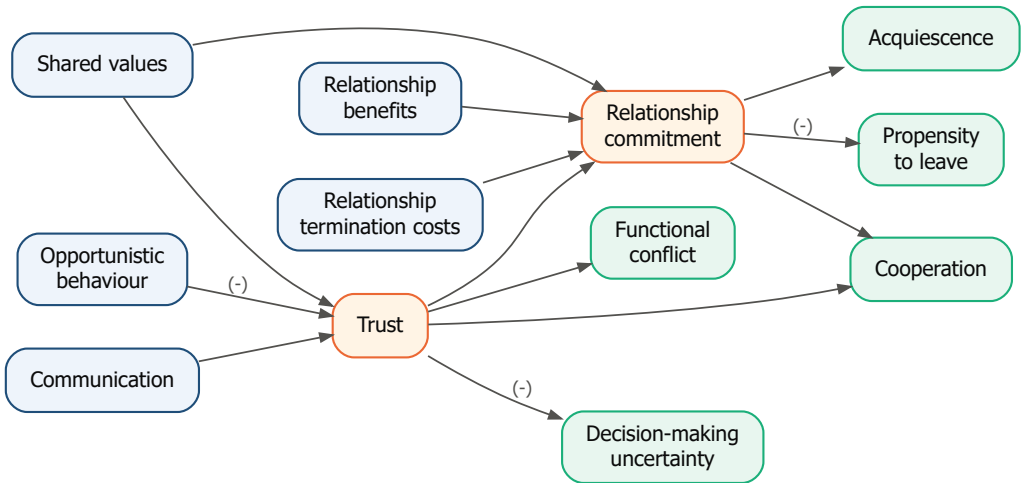


Figure 9.6. The key mediating variable (KMV) model of relationship marketing, after Morgan and Hunt (1994). Commitment and trust mediate the effects of five antecedents on five relational outcomes. Negative paths are marked with a minus sign.

The propositions embedded in the model are:

- **P1.** Relationship termination costs and relationship benefits increase commitment.
- **P2.** Shared values increase both commitment and trust.
- **P3.** Communication increases trust, opportunistic behaviour decreases it.
- **P4.** Trust increases commitment.
- **P5.** Commitment increases acquiescence and cooperation and decreases the propensity to leave. Trust increases cooperation and functional conflict (disagreements resolved amicably) and decreases decision-making uncertainty.

9.6.2 Evidence and extensions

The KMV model became one of the most cited frameworks in marketing, and later work refined it. Doney and Cannon (1997) analysed the processes through which industrial buyers come to trust suppliers and their salespeople, and found that trust influenced buyers' anticipated future interaction with a supplier but did not directly determine the current supplier choice, which depended more on price and performance. Garbarino and Johnson (1999) found that for customers with weak relational orientation, overall satisfaction drove future intentions, whereas for highly relational customers, trust and commitment did: a segmentation insight with direct implications for

loyalty programmes. In a meta-analysis, Palmatier et al. (2006) found that relationship-marketing efforts had larger effects on outcomes when relationships were more critical to customers, as in service offerings, channel exchanges, and business markets, and when relationships were built with an individual (such as a salesperson) rather than with the selling firm as a whole.

DEBATE AND EVIDENCE

How much does trust explain? Commitment and trust are conceptually close to the outcomes they predict, so survey-based tests risk measuring one underlying evaluation several times. Many consumers, moreover, do not want "relationships" with brands at all (Garbarino and Johnson 1999). The theory also sits in tension with transaction cost economics (Section 6.3), which emphasizes safeguards against opportunism rather than trust. Contemporary research often treats contracts and relational governance as complements.

BUSINESS CASE

Toyota's supplier network. Toyota's relationships with its suppliers are a widely studied example of trust-based relational exchange in practice. The company organized supplier associations, provided consulting on production methods, and encouraged knowledge sharing among suppliers, creating a network in which members expected that sharing valuable know-how would be reciprocated rather than exploited (Dyer and Nobeoka 2000). The case shows all elements of the KMV model: high relationship benefits and termination costs, shared values around the Toyota Production System, frequent communication, and norms against opportunism, yielding cooperation that arm's-length contracting would be unlikely to produce.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Supplier communication quality increases buyer trust, which mediates the effect of communication on buyer commitment.
- *H2.* The effect of trust on renewal is stronger for customers with high relational orientation than for customers with low relational orientation.
- *H3.* Salesperson turnover reduces customer commitment more when the relationship is anchored in a single individual than when it is anchored in multiple contacts.

Designs. Dyadic surveys with both buyer and supplier reports. CRM and contract-renewal data linked to survey measures. Panel data on salesperson departures as quasi-experimental shocks to relationships.

Measures. Commitment and trust scales (R. M. Morgan and Hunt 1994), relationship benefits and termination costs, objective outcomes such as share of wallet, retention, and contract renewals.

COMMON PITFALL

Measuring trust and commitment as one construct. Because the KMV model specifies distinct antecedents and outcomes for trust and commitment, collapsing them into a single "relationship quality" index loses the theory's discriminating predictions. Demonstrate discriminant validity and test the mediating structure, including the direct effects the model does *not* predict.

9.7 Chapter Summary

KEY TAKEAWAYS

- **Social exchange theory** explains relationships as the exchange of valued resources sustained by the norm of reciprocity. Unspecified obligations require and build trust (Blau 1964, Gouldner 1960, Cropanzano and Mitchell 2005).
- **Power is relational**: it equals the other party's dependence, which rises with the value of what is provided and falls with the availability of alternatives (Emerson 1962).
- **Network exchange theory** shows that in negatively connected networks power comes from the ability to exclude, which centrality measures can miss (K. S. Cook et al. 1983, Markovsky, Willer, and Patton 1988).
- **Weak ties** matter because they bridge. The best causal evidence shows that *moderately* weak ties are the most valuable for job mobility (M. S. Granovetter 1973, Rajkumar et al. 2022).
- **Embeddedness** brings trust and joint problem-solving, but networks composed only of embedded ties become insular: the paradox of embeddedness (M. Granovetter 1985, Uzzi 1997).
- **Brokerage** across structural holes yields information and control benefits, while **closure** yields trust and coordination. High performance typically combines both (Burt 1992, 2005, Coleman 1988).
- **Social capital** has structural, relational, and cognitive dimensions and carries costs as well as benefits (Nahapiet and Ghoshal 1998, Adler and Kwon 2002, Portes 1998).
- **Commitment and trust** are the key mediating variables of successful relational exchange in marketing (R. M. Morgan and Hunt 1994, Palmatier et al. 2006).

Table 9.8 presents the comparison of the chapter's theories.

Table 9.8. Comparison of the chapter's theories.

Theory	Unit of analysis	Key mechanism	Signature prediction	Typical business use
Social exchange	Dyad	Reciprocity, dependence	Support is repaid with effort and loyalty	Employee engagement, supplier relations
Network exchange	Position in network	Exclusion via alternatives	Positions whose partners lack alternatives gain power	Channel power, platform intermediation
Weak ties	Tie / network	Bridging to novel information	Moderately weak ties carry job and idea leads	Referral hiring, networking strategy
Embeddedness	Dyad / interfirm network	Trust, joint problem-solving	Mixed embedded and arm's-length ties perform best	Supplier and alliance portfolios
Structural holes	Ego network	Non-redundant information, control	Brokers generate better ideas and advance faster	Boundary-spanning roles, innovation
Social capital	Individual to community	Closure and brokerage	Relationships yield returns and costs	Knowledge sharing, key accounts
Commitment-trust	Relationship	Commitment and trust as mediators	Trust and commitment drive cooperation and retention	Relationship marketing, B2B loyalty

9.8 Discussion Questions

1. Blau distinguished social from economic exchange. Is the employment relationship in a gig-economy platform an economic exchange, a social exchange, or both? What evidence would help you decide?
2. Using Figure 9.2, explain why the most central position in the 5-line is predicted to be weak. Identify a real market in which a "central" intermediary has little bargaining power.
3. Rajkumar et al. (2022) found that moderately weak ties were most useful for job mobility. Propose two mechanisms that could explain why the very weakest ties were less useful, and design a study to distinguish them.
4. Uzzi's paradox of embeddedness and Burt's structural holes both imply that networks can be "too closed". How do the two arguments differ in mechanism?
5. A company wants to increase innovation by creating more boundary-spanning roles. Drawing on Burt (2004), Ahuja (2000), and Obstfeld (2005), what risks should it anticipate, and how should it measure success?
6. For a B2B firm of your choice, map the antecedents of the KMV model (R. M. Morgan and Hunt 1994). Which antecedent would you invest in first, and why?

9.9 Further Reading

- Blau (1964): the classic statement of social exchange, power, and the difference between social and economic exchange.
- Cropanzano and Mitchell (2005): the most widely used management review of social exchange theory, with a clear account of its conceptual ambiguities.
- K. S. Cook and Whitmeyer (1992): a concise comparison of exchange-theoretic and network-analytic approaches to social structure.
- M. S. Granovetter (1973): the original weak-ties argument, short, elegant, and still generative.
- Rajkumar et al. (2022): a large-scale causal test of the weak-ties thesis that refines rather than overturns it.
- Uzzi (1997): a rich field study of embeddedness and its paradox in an interfirm network.
- Burt (2005): an accessible synthesis of brokerage and closure as forms of social capital.
- R. M. Morgan and Hunt (1994): the foundational commitment-trust model of relationship marketing.

Chapter 10

Innovation Diffusion and Technology Adoption

Why did hybrid seed corn, which offered American farmers markedly higher yields, take more than a decade to spread through communities in which its benefits were plain to see? Why do some technology products sell briskly to enthusiasts and then stall, while others glide into the mainstream? Why does a firm that spends heavily on a new enterprise system often find, a year later, that employees still keep their real work in spreadsheets? And how can a company forecast the sales of a product that has never existed before?

These questions concern the **adoption** of new ideas, products, and technologies by individuals and organizations, and the **diffusion** of those innovations through a population over time. This chapter brings together two traditions that address them at different levels. The *macro* tradition, **diffusion of innovations** theory and the **Bass model**, explains the aggregate shape of adoption over time: why cumulative adoption follows an S-curve and how communication among adopters drives it. The *micro* tradition, the **theory of reasoned action**, the **theory of planned behavior**, the **technology acceptance model**, and the **unified theory of acceptance and use of technology**, explains why a given individual decides to adopt and use a technology. Each tradition supplies what the other lacks: diffusion models describe the curve but say little about individual decisions, while acceptance models explain individual decisions but say little about social dynamics (Table 10.1).

Table 10.1. Chapter at a glance: theories of innovation diffusion and technology adoption.

Theory	Core question	Level of analysis	Key source
Diffusion of innovations	How and why do innovations spread through social systems?	Population / social system	Rogers (2003)
Bass diffusion model	How can the time path of first purchases of a new product be forecast?	Market	Bass (1969)
Theory of reasoned action / planned behavior	What determines intentions and behaviour?	Individual	Fishbein & Ajzen (1975), Ajzen (1991)
Technology acceptance model (TAM)	Why do people accept or reject information technology?	Individual	Davis (1989)
UTAUT / UTAUT2	What unified set of factors predicts technology intention and use?	Individual (in organizations / consumer markets)	Venkatesh et al. (2003, 2012)

10.1 Diffusion of Innovations

THEORY CARD

Core claim. Innovations spread through a social system over time via communication channels. The rate of adoption depends on how potential adopters *perceive* the innovation's attributes, and adoption over time typically follows an S-shaped cumulative curve as early adopters influence later ones.
Origins. Rural sociology (B. Ryan and Gross 1943), agricultural economics (Griliches 1957), and medical sociology (Coleman, Katz, and Menzel 1966). Synthesized by Rogers (2003), first published in 1962.
Level of analysis. The social system (community, market, industry, or organization), with individual adoption decisions as the micro-foundation.
Key constructs. Innovation, communication channels, time, social system, perceived attributes (relative advantage, compatibility, complexity, trialability, observability), adopter categories, opinion leaders, change agents, innovation-decision process, re-invention, critical mass.
Boundary conditions. Adoption is voluntary (or partly so), information about the innovation spreads through observable channels, and potential adopters are connected. Mandated adoption, network-externality goods, and fads follow different dynamics.
Related theories. Bass model, Strength of weak ties (Section 9.3), Technology acceptance model, Institutional theory (Section 7.2), Social influence (Chapter 22).

10.1.1 Origins and intellectual history

Diffusion research has unusually concrete origins. B. Ryan and Gross (1943) interviewed farmers in two Iowa communities about their adoption of hybrid seed corn, an innovation that offered clear yield advantages. Adoption nonetheless unfolded over many years: a few farmers planted hybrid seed early, most followed only after seeing it work on neighbours' farms, and a minority held out much longer. Salespeople were an important source of first knowledge about the seed, but neighbours were more influential in the decision to adopt: an early demonstration that *awareness* and *persuasion* run through different channels. Farmers also typically tried hybrid seed on part of their acreage before converting fully, illustrating the importance of trialability.

Griliches (1957) analysed the same innovation from an economic perspective. Fitting logistic curves to adoption across American states and crop-reporting districts, he showed that differences in when adoption began, how fast it proceeded, and where it levelled off could be explained largely by economic factors: the profitability of hybrid corn in each area and the availability of hybrids adapted to local conditions. The Ryan-Gross and Griliches studies are often read as rival sociological and economic accounts, but together they establish that both *perceived profitability* and *social communication* shape diffusion.

Coleman, Katz, and Menzel (1966) extended the paradigm to medicine with their *Medical Innovation* study of physicians' adoption of a new antibiotic in four Illinois cities. Doctors who were more integrated into professional networks adopted earlier, and the authors interpreted the pattern as evidence of social contagion. The study became the classic demonstration of network effects (Section 17.6) in diffusion, and, decades later, a cautionary tale (see the debate box below).

Rogers (2003) synthesized these and thousands of later studies into a general theory. In his account,

diffusion is the process through which an innovation is communicated, via particular channels and over time, among the members of a social system, a definition that identifies four main elements: the innovation, communication channels, time, and the social system.

10.1.2 Core logic: perceived attributes and the decision process

Rogers argued that the rate of adoption is shaped above all by how potential adopters perceive an innovation. Five perceived attributes recur across studies (Table 10.2). A meta-analysis by Tornatzky and Klein (1982) found that relative advantage, compatibility, and complexity were the attributes most consistently related to adoption, and G. C. Moore and Benbasat (1991) developed a widely used instrument for measuring perceived innovation characteristics in information systems research.

Table 10.2. Rogers's five perceived attributes of innovations (definitions after Rogers, 2003).

Attribute	Definition	Effect on rate of adoption	Business lever
Relative advantage	Degree to which the innovation is perceived as better than the idea it supersedes	Positive	Quantify savings, price below value, demonstrate ROI
Compatibility	Consistency with existing values, past experience, and needs	Positive	Integrations, familiar interfaces, position as an upgrade
Complexity	Degree to which the innovation is perceived as difficult to understand and use	Negative	Simplify onboarding, default settings, training
Trialability	Degree to which it can be experimented with on a limited basis	Positive	Free trials, pilots, freemium tiers, test drives
Observability	Degree to which results are visible to others	Positive	Case studies, visible design, referral programmes, usage dashboards

At the individual level, Rogers described an **innovation-decision process** with five stages: *knowledge* (exposure and understanding), *persuasion* (forming a favourable or unfavourable attitude), *decision* (adoption or rejection), *implementation* (putting the innovation to use, often with **re-invention**, or modification by the user), and *confirmation* (seeking reinforcement and possibly reversing the decision). Mass-media channels matter most at the knowledge stage. Interpersonal channels, especially with **near-peers**, matter most at the persuasion stage.

The core propositions can be summarized as follows:

- **P1 (Perceived attributes).** Innovations perceived as having greater relative advantage, compatibility, trialability, and observability, and less complexity, are adopted more rapidly.
- **P2 (Interpersonal influence).** Interpersonal communication with prior adopters, especially similar others, is the main driver of the persuasion stage. Diffusion therefore accelerates as the number of adopters grows.
- **P3 (S-curve).** Cumulative adoption follows an S-shaped curve: slow at first, rapid once interpersonal networks are activated, and slowing as the pool of non-adopters is exhausted.
- **P4 (Opinion leadership).** Adoption by opinion leaders, members whose views others seek, accelerates adoption by their followers.
- **P5 (Adopter heterogeneity).** Earlier adopters differ systematically from later adopters in socio-economic status, personality, and communication behaviour.

10.1.3 Mechanism: adopter categories and the S-curve

Rogers partitioned adopters by the timing of their adoption relative to the average, using standard deviations of a normal distribution of adoption times. The resulting categories (Figure 10.1) are **innovators** (the first 2.5 per cent), **early adopters** (the next 13.5 per cent), the **early majority** (34 per cent), the **late majority** (34 per cent), and **laggards** (the final 16 per cent). The percentages follow by construction from the normal curve. They are an ideal type, not an empirical regularity.

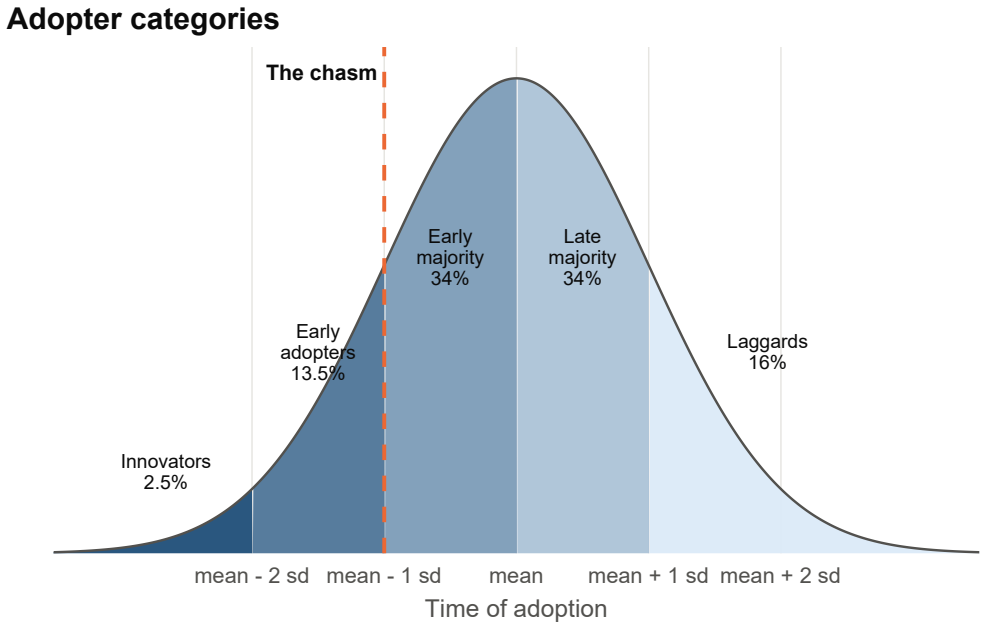


Figure 10.1. Rogers's adopter categories, defined by standard deviations from the mean adoption time of a normal distribution. The dashed line marks where Moore (1991) locates the 'chasm' between early adopters and the early majority. The percentages follow from the normal curve by construction.

The same distribution, accumulated over time, produces the S-curve in Figure 10.2. Adoption is slow while only innovators and early adopters are involved, accelerates as the early majority joins, and decelerates as the remaining non-adopters are increasingly late majority and laggards.

Table 10.3 summarizes the characteristics Rogers associated with each category. Opinion leadership is concentrated among early adopters, which is why they are often the most strategically important segment for marketers: they are close enough to the mainstream to be emulated, yet venturesome enough to adopt before the evidence is in.

Table 10.3. Adopter categories and their characteristics (after Rogers, 2003. Seller needs follow Moore, 1991).

Category	Share (ideal type)	Typical orientation	What they need from the seller
Innovators	2.5%	Venturesome, tolerate uncertainty, cosmopolitan networks	Access to the technology itself
Early adopters	13.5%	Respected locally, high opinion leadership	A vision of strategic advantage
Early majority	34%	Deliberate, adopt once others have	Proven, complete solutions and references from peers

Table 10.3. Adopter categories and their characteristics (after Rogers, 2003. Seller needs follow Moore, 1991). *(continued)*

Category	Share (ideal type)	Typical orientation	What they need from the seller
Late majority	34%	Sceptical, adopt from necessity or peer pressure	Low cost, low risk, standardization
Laggards	16%	Traditional, oriented to the past	Compatibility with existing practices, or no choice

10.1.4 Evidence and extensions

The diffusion paradigm has been applied to agricultural practices, medical technologies, consumer durables, management practices, policies, and social movements (Strang and Soule 1998). Four extensions are particularly important for business research.

- **Organizational adoption.** Organizations face knowledge barriers that individuals do not. Attewell (1992) argued that the diffusion of business computing was paced less by awareness than by the burden of organizational learning, and that intermediaries such as service bureaus lowered that burden. Fichman and Kemerer (1999) documented **assimilation gaps**: many firms that *acquired* software process innovations had not *deployed* them, so acquisition-based diffusion curves overstated real adoption.
- **Management fashions.** Abrahamson (1991) showed that administrative innovations can diffuse, and be rejected, for reasons unrelated to their efficiency, through fad and fashion processes driven by imitation and by fashion-setting organizations such as consultancies.
- **Network structure.** Thresholds matter: people adopt once enough of their contacts have (M. Granovetter 1978). In an online experiment, Centola (2010) found that a health behaviour spread farther and faster in clustered networks with redundant ties than in random networks, suggesting that behaviours requiring social reinforcement are **complex contagions** that travel through dense clusters rather than the long, weak bridges that carry information (compare the weak-ties argument (Section 9.3)).
- **Contagion versus homophily (Section @ref(homophily (Section 30.2))).** Correlated adoption among connected people need not mean that one influenced the other. Aral, Muchnik, and Sundararajan (2009) showed that a large share of apparent contagion in the adoption of a mobile service could be attributed to homophily: similar people adopting for similar reasons.

DEBATE AND EVIDENCE

Was *Medical Innovation* really about contagion? Van den Bulte and Lilien (2001) reanalysed the Coleman, Katz, and Menzel data together with the volume of pharmaceutical advertising during the diffusion period. Once marketing effort was controlled, the evidence for social contagion largely disappeared: doctors in the same networks adopted together because they were exposed to the same marketing, not because they influenced one another. Later work using richer data on a different drug found evidence of contagion over network ties even after controlling for marketing, with influence depending on who was influencing whom (R. Iyengar, Van den Bulte, and Valente 2011). The lesson is methodological: distinguishing contagion from common exposure and homophily requires designs that measure all three.

Pro-innovation bias. Rogers himself noted that diffusion research tends to assume that inno-

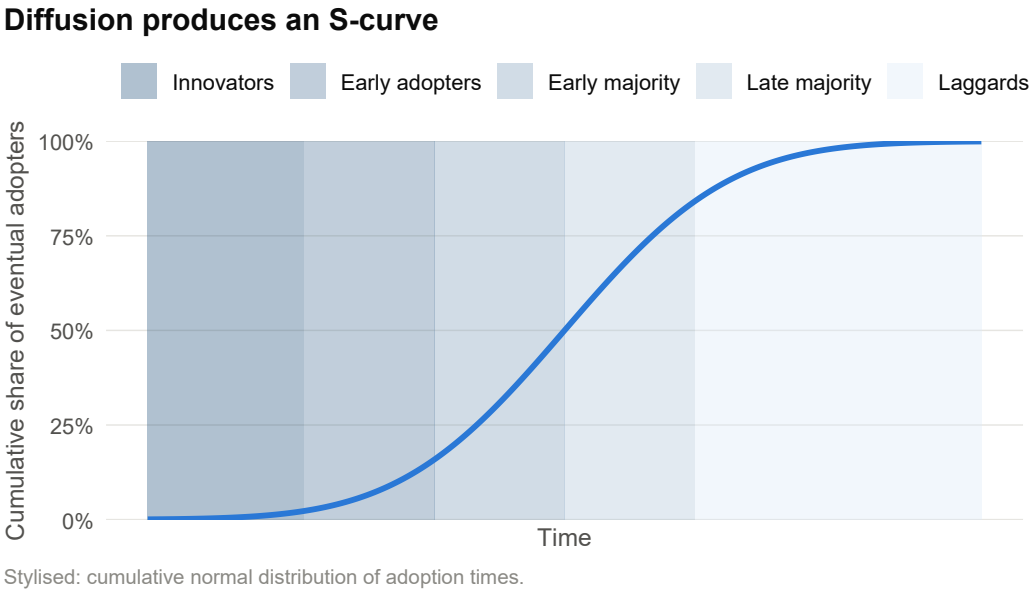


Figure 10.2. Stylised S-shaped cumulative adoption curve implied by normally distributed adoption times, with the period in which each adopter category adopts shaded.

vations *should* be adopted, that non-adopters are deficient, and that faster diffusion is better. This bias leads researchers to neglect rejection, discontinuance, and re-invention, and to study successful innovations whose diffusion was traceable after the fact.

10.1.5 Moore's chasm: a practitioner extension

G. A. Moore (1991) argued, from experience marketing high-technology products, that the adoption curve is not continuous. Early adopters, "visionaries" seeking breakthrough advantage, buy for different reasons than the early majority: "pragmatists" seeking proven, incremental improvement and references from peers like themselves. Because pragmatists do not regard visionaries as credible references, products can win the early market and then stall. Moore's prescription is to cross this **chasm** by concentrating on a narrow beachhead segment of pragmatic customers and delivering a **whole product**, the core technology plus the complementary services, integrations, and support that make it usable for that segment, before expanding to adjacent segments.

Academic research has partly vindicated the idea. 1 documented a **saddle** in the sales of many consumer electronics products: an initial peak, a trough, and a later recovery that eventually exceeds the initial peak. They explained it with a model in which the early market and the main market communicate weakly with one another, so that the early market's adoption does not fully transfer to the mainstream: essentially a formalization of Moore's chasm in network terms. Golder and Tellis (1997) showed that the **takeoff** of really new consumer durables is closely associated with price declines, underscoring the role of relative advantage.

BUSINESS CASE

Electric-vehicle diffusion (conceptual analysis). Rogers's attributes offer a structured di-

agnosis of electric-vehicle adoption. *Relative advantage* is mixed: lower running costs and strong acceleration against higher purchase prices in many markets. *Compatibility* depends on whether a household has home charging and how its driving patterns match vehicle range. *Complexity* includes learning about charging networks and battery health. *Trialability* is limited by the cost of the product but can be increased through test drives, rentals, and car-sharing. *Observability* is high: electric vehicles are visible on the road, and neighbourhood clustering of adoption is consistent with interpersonal influence (though homophily is an alternative explanation). Policy levers map onto the attributes: purchase incentives raise relative advantage, public charging raises compatibility, and fleet programmes increase observability. Norway, where sustained tax incentives preceded electric vehicles becoming the majority of new passenger-car sales in the early 2020s, is the most frequently cited example of policy accelerating diffusion.

COMMON PITFALL

Treating adopter percentages as data. The 2.5/13.5/34/34/16 split is a definition based on the normal curve, not an empirical finding. Real adoption distributions are often skewed, and many innovations never reach the "late majority" at all. Use the categories as a segmentation heuristic and estimate the actual distribution of adoption times from data.

10.2 The Bass Diffusion Model

THEORY CARD

Core claim. The probability that a not-yet-adopter adopts at time t is the sum of an **innovation** effect (external influence, independent of prior adopters) and an **imitation** effect proportional to the share of the market that has already adopted. This simple hazard produces the S-shaped diffusion curve and the bell-shaped curve of new adoptions.
Origins. Marketing science. F. M. Bass (1969), building on epidemic models and on Rogers's adopter categories. Reviewed by Mahajan, Muller, and Bass (1990) and Peres, Muller, and Mahajan (2010).
Level of analysis. The market for a new product or technology (first purchases).
Key constructs. Market potential (m), coefficient of innovation (p), coefficient of imitation (q), hazard rate, time to peak, takeoff.
Boundary conditions. Applies to first-time adoption of a new product category with a fixed market potential and no repeat purchase. Extensions are needed for repeat purchases, multiple product generations, and strong supply constraints or marketing-mix changes.
Related theories. Diffusion of innovations, Wisdom of the crowd (Section 30.5).

10.2.1 Core logic

F. M. Bass (1969) formalized Rogers's insight that early adopters and later adopters respond to different influences. Let $F(t)$ be the cumulative fraction of the eventual adopter population that has adopted by time t, and $f(t)$ its derivative. The Bass model specifies the hazard of adoption, the probability of adopting at t given no adoption before t, as

$$\frac{f(t)}{1-F(t)} = p + q\,F(t),$$

where p is the **coefficient of innovation** (external influence, such as advertising and media) and q is the **coefficient of imitation** (internal influence, such as word of mouth (Section 26.3) from previous adopters). With market potential m, the number of new adopters at t is $n(t) = m\,f(t)$. Solving the differential equation with $F(0) = 0$ gives the closed form

$$F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p}\,e^{-(p+q)t}}.$$

Three implications are especially useful for managers. When $q > p$, new adoptions rise to a peak and then decline. The peak occurs at $t^* = \ln(q/p)/(p+q)$, when cumulative penetration equals $\frac{1}{2} - \frac{p}{2q}$, and the peak rate of adoption is $m\,(p+q)^2/(4q)$. A larger q relative to p produces a later, sharper peak. A larger p produces faster early sales and an earlier, flatter peak.

DEFINITION

Definition 10.1. Coefficient of innovation (p). The baseline probability of adoption that does not depend on how many others have adopted: the influence of external communication. **Coefficient of imitation (q).** The increase in the probability of adoption per unit increase in the cumulative adopted share: the influence of prior adopters (F. M. Bass 1969, Mahajan, Muller, and Bass 1990).

10.2.2 Mechanism

Figure 10.3 simulates three stylised products. The *benchmark* uses parameters of the order reported as averages across many published applications in the meta-analysis of Sultan, Farley, and Lehmann (1990) ($p \approx 0.03$, $q \approx 0.38$). The *word-of-mouth-driven* product has a very small p and a large q. The *advertising-driven* product has a large p and a small q.

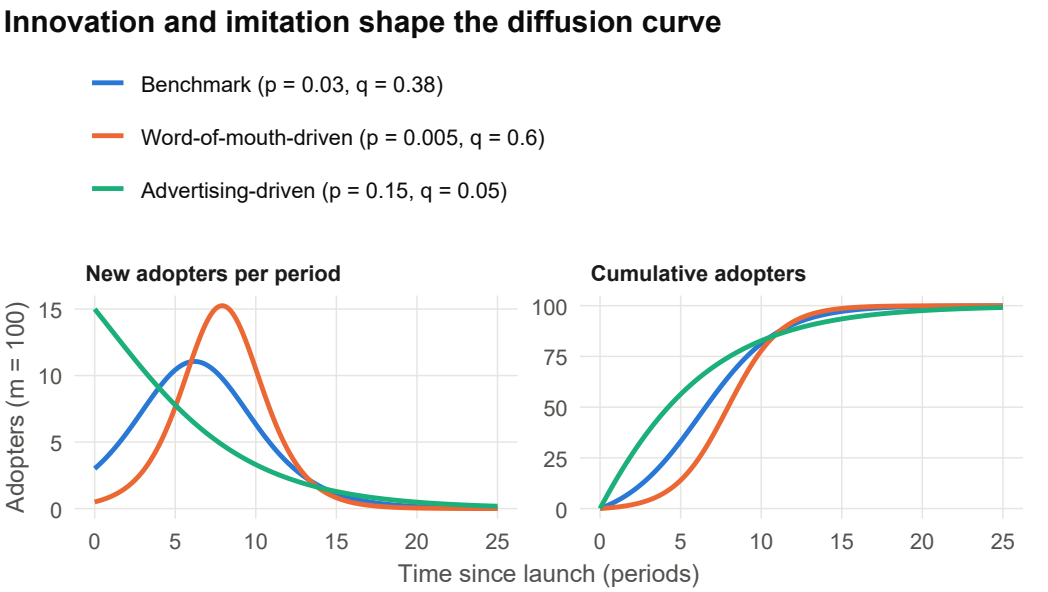
Table 10.4 computes the managerial quantities implied by each scenario directly from the formulas above.

Table 10.4. Managerial metrics implied by the Bass model for the three illustrative scenarios (computed). When q is not greater than p, new adoptions are highest at launch.

Scenario	q / p	Time to peak	Peak adopters per period (m = 100)	Penetration at peak
Benchmark (p = 0.03, q = 0.38)	12.7	6.2	11.1	46%
Word-of-mouth-driven (p = 0.005, q = 0.6)	120.0	7.9	15.3	50%
Advertising-driven (p = 0.15, q = 0.05)	0.3	0.0	15.0	0%

10.2.3 Estimating the model in R

The Bass model can be estimated from early sales data by nonlinear least squares on per-period adoptions (Srinivasan and Mason 1986). The code below is self-contained and can be adapted by



Illustrative simulation of the Bass (1969) model. Benchmark values are of the order reported by Sultan et al. (1990).

Figure 10.3. Illustrative Bass-model simulations for three stylised products with market potential m = 100. The left panel shows new adopters per period. The right panel shows cumulative adoption. Imitation-driven diffusion produces a late, sharp peak. Innovation-driven diffusion produces an early, flat one.

replacing the simulated sales vector with real data. It simulates nine periods of noisy sales for a product with known parameters, then recovers them.

```
# Bass model: cumulative fraction adopted and per-period adoptions
bass_F <- function(t, p, q) (1 - exp(-(p + q) * t)) / (1 + (q / p) * exp(-(p + q) * t))
bass_sales <- function(t, m, p, q) m * (bass_F(t, p, q) - bass_F(t - 1, p, q))

# Illustrative data: true m = 1,000,000, p = 0.02, q = 0.45, plus noise
set.seed(1969)
period <- 1:9
sales <- round(bass_sales(period, m = 1e6, p = 0.02, q = 0.45) *
                 exp(rnorm(length(period), 0, 0.06)))

# Nonlinear least squares with bounds that keep m, p, q positive
fit <- nls(sales ~ bass_sales(period, m, p, q),
           start = list(m = 8e5, p = 0.03, q = 0.3),
           algorithm = "port", lower = c(1, 1e-4, 1e-4))
round(coef(fit), 3)
```

```
#>           m           p           q
#> 929385.905       0.019       0.485
```

Figure 10.4 uses the fitted model to forecast the remainder of the life cycle.

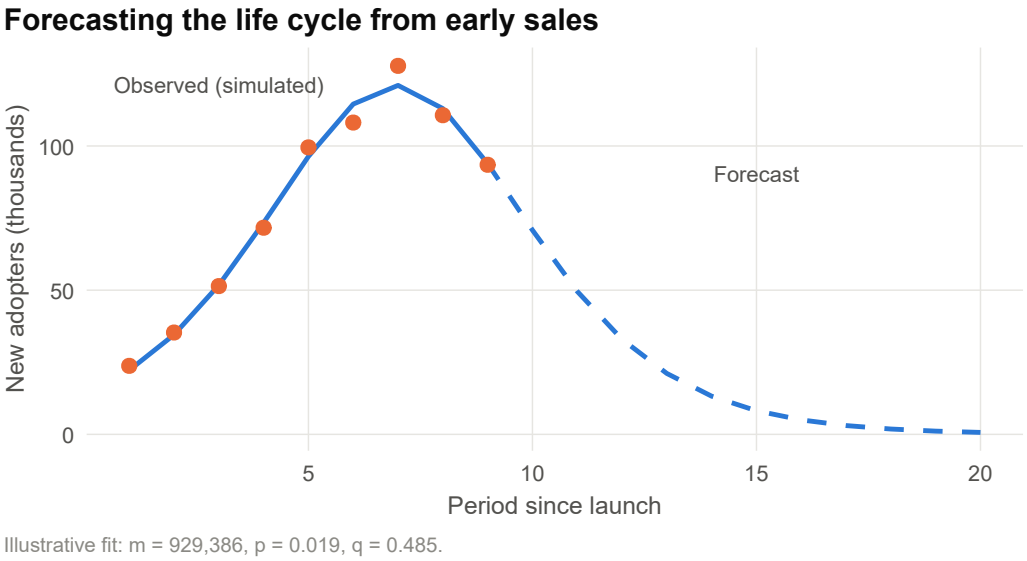


Figure 10.4. Illustrative Bass-model forecast. Points are simulated sales for the first nine periods. The line is the fitted model, extended as a forecast. Simulated data, not an actual product.

10.2.4 Evidence and extensions

The Bass model has become one of the most widely applied models in marketing science (Mahajan, Muller, and Bass 1990, Peres, Muller, and Mahajan 2010). Several findings and extensions are particularly relevant:

- **Parameter regularities.** Across published applications, q is typically much larger than p: word of mouth dominates external influence in shaping the curve's form (Sultan, Farley, and Lehmann 1990). In the absence of sales data, analysts often forecast **by analogy**, borrowing parameters from similar past products.
- **Marketing-mix variables.** The original model contains no price or advertising terms. F. M. Bass, Krishnan, and Jain (1994) developed a **Generalized Bass Model** that includes current marketing effort and showed why the original model fits well anyway: when marketing variables change at roughly constant rates, their effect is absorbed into the estimated parameters.
- **Alternative explanations of the S-curve.** An S-curve can arise from heterogeneity in adopters' thresholds or incomes rather than from contagion. Van den Bulte and Stremersch (2004) found that the ratio q/p was higher in countries with greater income inequality, consistent with an income-heterogeneity interpretation of what the Bass model labels "imitation".
- **Beyond first purchase.** Extensions cover repeat purchases, successive technology generations, cross-country diffusion, competition among brands, and network externalities (Peres, Muller, and Mahajan 2010).

DEBATE AND EVIDENCE

Is q really "imitation"? The Bass model is a reduced form: it reproduces the *shape* of diffusion without identifying the *mechanism*. A large estimated q is consistent with word of mouth, but also with falling prices, expanding distribution, learning by the firm, or heterogeneous

adoption thresholds (F. M. Bass, Krishnan, and Jain 1994, Van den Bulte and Stremersch 2004). Estimates from early data are also notoriously unstable, parameters can change substantially as data points beyond the inflection point are added, so forecasts made before the peak should be presented with wide uncertainty bands and cross-checked against analogies.

BUSINESS CASE

Illustrative scenario: forecasting a new product launch. A consumer electronics firm is launching a home device in a new category. Before launch, it has no sales history, so it forecasts by analogy: it collects estimated p and q values for comparable past categories, adjusts market potential m using purchase-intention surveys and household penetration of complementary products, and simulates the life cycle. The analysis yields three uses. First, **capacity planning**: the forecast peak sales rate determines manufacturing and inventory commitments. Second, **marketing allocation**: if diffusion is imitation-driven (high q/p), early spending on seeding, referral incentives, and reviews may yield more than heavy launch advertising. Third, **re-estimation**: once several periods of sales are observed, the model is refit (as in the code above), and the forecast is updated, with explicit uncertainty, because pre-peak estimates are unstable.

TRY IT YOURSELF

Modify the code above to fit the model using only the first five periods of simulated sales. How much do the estimates of m, p, and q change? What does this imply for forecasting a product's peak from its launch year alone?

10.3 Theories of Reasoned Action and Planned Behavior

THEORY CARD

Core claim. Behaviour is determined by the intention to perform it. Intentions are in turn determined by attitudes toward the behaviour, subjective norms, and, in the theory of planned behavior, perceived behavioral control, each of which rests on underlying beliefs.
Origins. Social psychology: the theory of reasoned action (TRA) of Fishbein and Ajzen (1975) and Ajzen and Fishbein (1980). The theory of planned behavior (TPB) of Ajzen (1991).
Level of analysis. Individual.
Key constructs. Behavioural, normative, and control beliefs, attitude toward the behaviour, subjective norm, perceived behavioral control, intention, behaviour.
Boundary conditions. Best suited to deliberate behaviours under volitional control, measured at a matching level of specificity (the principle of compatibility: same action, target, context, and time). Weaker for habitual, impulsive, or highly constrained behaviours.
Related theories. Technology acceptance model, UTAUT, Social norms (Section 22.5), Self-regulation (Chapter 20).

The theory of reasoned action provides the micro-foundation on which most technology-acceptance models are built. Fishbein and Ajzen (1975) proposed that the proximal determinant of a volitional behaviour is the person's **intention** to perform it. Intention depends on two factors: the person's **attitude toward the behaviour** (a summary evaluation based on beliefs about its

consequences, weighted by the evaluation of those consequences) and the **subjective norm** (the perceived expectations of important others, weighted by the motivation to comply with them). In a meta-analysis, Sheppard, Hartwick, and Warshaw (1988) found that the theory predicted intentions and behaviour well, even in some situations that fell outside its original boundary conditions.

Ajzen (1991) extended the TRA to behaviours that are not fully under volitional control by adding **perceived behavioral control (PBC)**: the perceived ease or difficulty of performing the behaviour, based on control beliefs about the presence of resources and obstacles. PBC influences intention and, to the extent that it reflects actual control, also predicts behaviour directly (Figure 10.5). A meta-analysis by Armitage and Conner (2001) found that the TPB accounted for substantial variance in intentions and somewhat less in behaviour, and that PBC added explanatory power beyond attitudes and subjective norms, while subjective norm was the weakest predictor of intention.

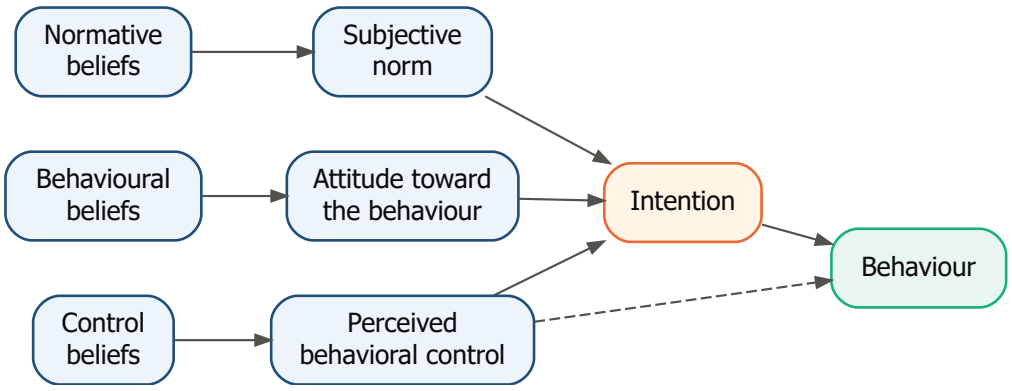


Figure 10.5. The theory of planned behavior (Ajzen, 1991). The theory of reasoned action (Fishbein & Ajzen, 1975) is the same model without perceived behavioral control. The dashed path from control to behaviour reflects the extent to which perceived control approximates actual control.

DEBATE AND EVIDENCE

The intention-behaviour gap. Correlational support for the TPB is strong, but experimental evidence is more sobering. Webb and Sheeran (2006) meta-analysed experiments that changed intentions and found that the resulting change in behaviour was considerably smaller than the change in intention. Sniehotta, Presseau, and Araújo-Soares (2014) went further, arguing that the TPB's limited predictive validity and its weak record as a guide to behaviour-change interventions warrant retiring it. Defenders reply that the theory was never designed as an intervention manual, and that the gap reflects habits, implementation problems, and measurement mismatches that the theory's boundary conditions already acknowledge.

10.4 Technology Acceptance Model

THEORY CARD

Core claim. People's intention to use an information technology is determined mainly by two beliefs: **perceived usefulness** (the degree to which they believe the system would improve their job performance) and **perceived ease of use** (the degree to which they believe using it would be free of effort). Ease of use also increases perceived usefulness.
Origins. Information systems: F. D. Davis (1989) and F. D. Davis, Bagozzi, and Warshaw (1989), adapting the theory of reasoned action. Extended in TAM2 (Venkatesh and Davis 2000) and TAM3 (Venkatesh and Bala 2008).
Level of analysis. Individual users, typically within organizations.
Key constructs. Perceived usefulness (PU), perceived ease of use (PEOU), attitude toward using, behavioural intention, system use, external variables (in TAM2: subjective norm, image, job relevance, output quality, result demonstrability, experience, voluntariness).
Boundary conditions. Developed for workplace productivity technologies, assumes individual choice over use. Explanatory power falls when use is mandated, when adoption is collective, or when hedonic or social motives dominate.
Related theories. Theory of planned behavior, UTAUT, Diffusion of innovations.

10.4.1 Origins and core logic

F. D. Davis (1989) developed and validated scales for two beliefs that he argued were fundamental determinants of user acceptance. **Perceived usefulness** captured the instrumental benefits of a system. **Perceived ease of use** captured its effort cost. In his studies, usefulness was more strongly linked to usage than ease of use, and ease of use appeared to operate partly *through* usefulness: a system that is easier to use leaves more effort available for productive work. F. D. Davis, Bagozzi, and Warshaw (1989) compared TAM with the theory of reasoned action in a longitudinal study of MBA students' use of a word-processing system. Perceived usefulness strongly influenced intentions at both measurement points. Ease of use had a smaller effect that faded as users gained experience, and TAM, despite its parsimony, predicted intentions at least as well as the TRA.

TAM simplified the TRA in two ways that proved decisive for its popularity. It replaced the elicitation of salient beliefs for each new behaviour with two general beliefs that apply to almost any technology, and it dropped the subjective norm, which Davis and colleagues found to have little effect on intentions in their setting. The result was a short, reliable survey that could be deployed in virtually any IT implementation (Figure 10.6).

10.4.2 Evidence and extensions

TAM became one of the most cited models in information systems. Venkatesh and Davis (2000) proposed **TAM2**, which explained perceived usefulness and intention in terms of **social influence processes** (subjective norm, image, voluntariness) and **cognitive instrumental processes** (job relevance, output quality, result demonstrability, and perceived ease of use). In four longitudinal field studies, subjective norm affected intention directly in mandatory settings, and its effects weakened as users gained experience. Venkatesh and Bala (2008) then proposed **TAM3**, adding determinants of perceived ease of use, such as computer self-efficacy (Section 21.5), perceptions of external control, computer anxiety, and perceived enjoyment, and an agenda for designing interventions before

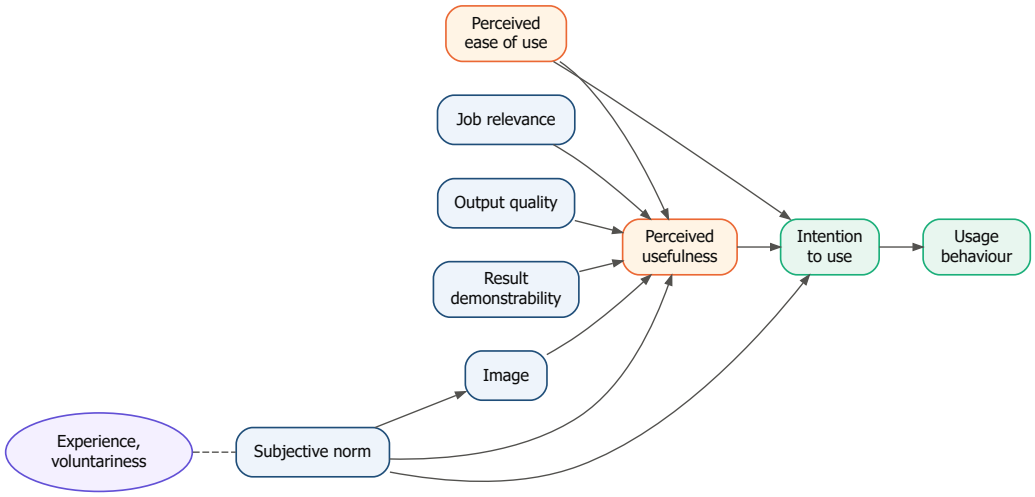


Figure 10.6. The technology acceptance model (Davis, 1989, Davis et al., 1989) with the TAM2 extensions of Venkatesh and Davis (2000). Social-influence and cognitive-instrumental processes shape perceived usefulness. Experience and voluntariness moderate the effect of subjective norm.

and after implementation.

Meta-analytic evidence supports the model's core relationships. W. R. King and He (2006) found the TAM measures to be highly reliable and the path from perceived usefulness to intention to be strong across many studies, with ease of use operating largely through usefulness. S. Taylor and Todd (1995) compared TAM with the TPB and a "decomposed" TPB: TAM predicted usage nearly as well as the richer models, though the decomposed TPB offered a more complete understanding of the determinants of intention. Legris, Ingham, and Collerette (2003), reviewing the literature, noted the model's predictive usefulness but also its reliance on self-reported use and student samples, and argued that it should be integrated into broader models that include organizational and social factors.

Post-adoption research addresses what happens after initial acceptance. Bhattacherjee (2001) proposed an **expectation-confirmation model** of continuance, in which satisfaction, driven by the confirmation of pre-adoption expectations, and post-adoption perceived usefulness determine whether users keep using a system. Continuance, not initial adoption, is often what determines the return on a technology investment.

BUSINESS CASE

Enterprise software rollout and user acceptance. Consider a firm replacing its legacy customer-relationship management system with a cloud platform. Licences and integration are costly, but the investment pays off only if sales staff actually use the system. TAM and TAM2 translate into a practical rollout plan. *Perceived usefulness* is raised by configuring the system around salespeople's own goals (pipeline visibility, faster quoting) rather than management's reporting needs alone, and by demonstrating results (TAM2's *result demonstrability*) through pilot teams. *Perceived ease of use* is raised through training, sensible defaults, and integration with email and calendar. Because use is often mandatory, *subjective norm* matters early, managers modelling use and reviewing pipelines in the system, but its influence fades with

experience, so long-run use must rest on usefulness. Finally, the assimilation-gap warning of Fichman and Kemerer (1999) applies inside the firm: licences deployed are not the same as the system used.

10.5 Unified Theory of Acceptance and Use of Technology

THEORY CARD

Core claim. Four constructs, **performance expectancy**, **effort expectancy**, **social influence**, and **facilitating conditions**, are the main determinants of technology intention and use, with their effects moderated by gender, age, experience, and voluntariness of use. UTAUT2 adapts the model to consumers by adding **hedonic motivation**, **price value**, and **habit**.
Origins. Information systems: Venkatesh et al. (2003) integrated eight prior models. Venkatesh, Thong, and Xu (2012) extended the theory to consumer contexts (UTAUT2).
Level of analysis. Individual users in organizations (UTAUT) and consumers (UTAUT2).
Key constructs. Performance expectancy, effort expectancy, social influence, facilitating conditions, behavioural intention, use behaviour, hedonic motivation, price value, habit, moderators.
Boundary conditions. Strongest for individual adoption of discrete technologies. The many moderators make the full model hard to test in small samples, and the constructs are general rather than specific to particular system features.
Related theories. Technology acceptance model, Theory of planned behavior, Diffusion of innovations.

10.5.1 Core logic

By the early 2000s, researchers could choose among many overlapping models of technology acceptance. Venkatesh et al. (2003) reviewed eight of them, the theory of reasoned action, TAM, the motivational model, the theory of planned behavior, a model combining TAM and the TPB, the model of PC utilization, innovation diffusion theory, and social cognitive theory, and compared them empirically in longitudinal studies in organizations. They then formulated a unified model in which:

- **Performance expectancy** (the belief that using the system will help attain gains in job performance. It subsumes perceived usefulness and relative advantage) affects intention, more strongly for men and younger workers.
- **Effort expectancy** (the ease associated with use. It subsumes perceived ease of use and complexity) affects intention, more strongly for women, older workers, and in early stages of experience.
- **Social influence** (the perception that important others believe one should use the system. It subsumes subjective norm and image) affects intention, mainly in mandatory settings and with complex moderation by gender, age, and experience.
- **Facilitating conditions** (the belief that organizational and technical infrastructure exists to support use. It subsumes perceived behavioral control and compatibility) affect *use* directly rather than intention, more strongly for older workers with more experience.

In their validation data, the unified model explained more variance in usage intention than any of the eight original models, which helped make UTAUT a standard framework. Figure 10.7 shows its

structure.

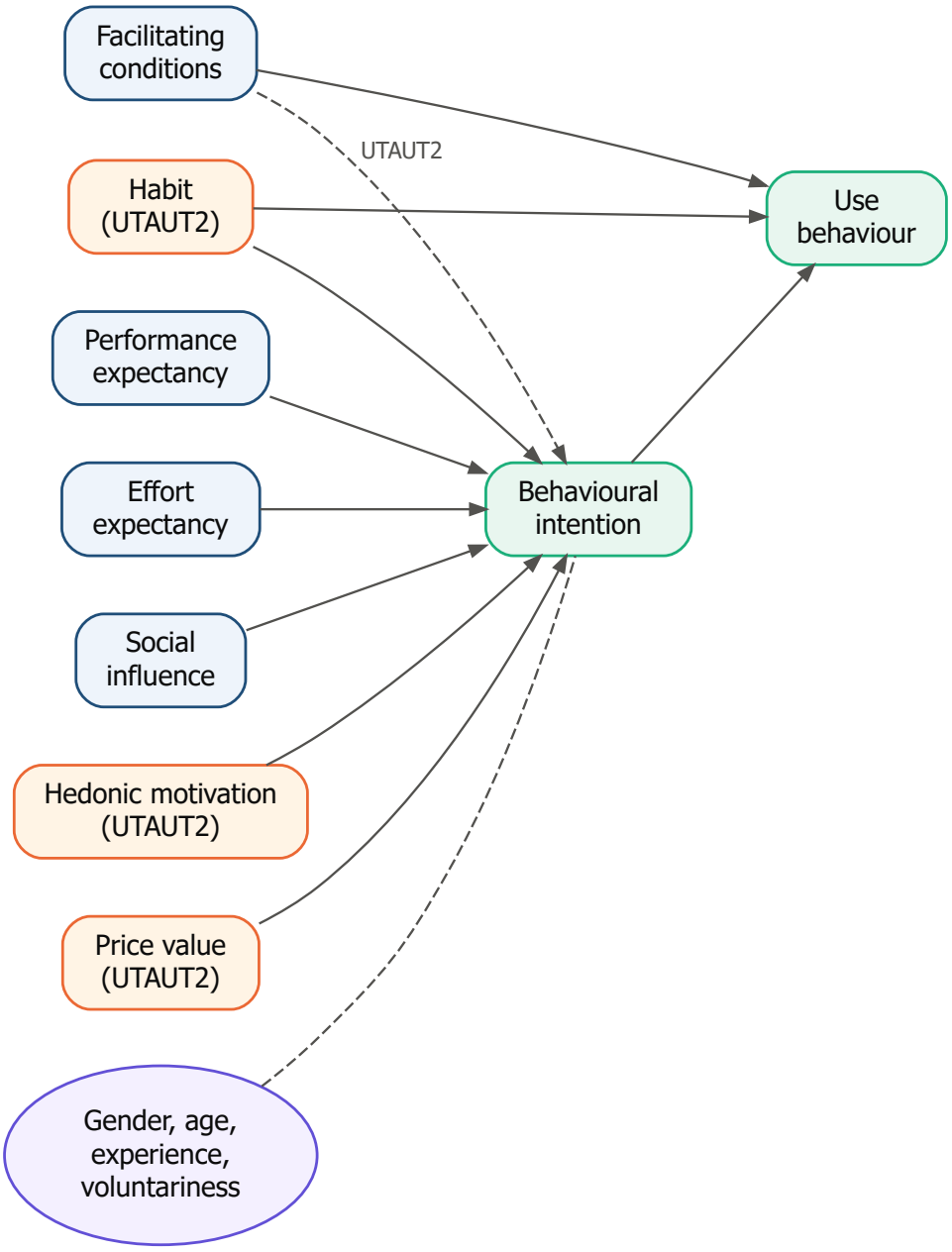


Figure 10.7. The unified theory of acceptance and use of technology (UTAUT. Venkatesh et al., 2003), with the constructs added in UTAUT2 for consumer contexts (Venkatesh et al., 2012) shown in orange. Moderators: gender, age, experience, and (in UTAUT) voluntariness of use.

10.5.2 UTAUT2: from employees to consumers

Venkatesh, Thong, and Xu (2012) adapted UTAUT for consumer technology use, studying mobile Internet use. Three constructs were added. **Hedonic motivation** (the fun or pleasure derived from using a technology) reflects the fact that consumers often use technology for enjoyment, not job per-

formance. **Price value** (the consumer's trade-off between perceived benefits and monetary cost) matters because consumers, unlike employees, usually pay for the technology themselves. **Habit** (the extent to which people perform behaviours automatically because of learning) affects both intention and use directly, capturing the fact that much technology use becomes routine. Voluntariness was dropped as a moderator because consumer use is voluntary, and facilitating conditions were allowed to affect intention as well as use. Table 10.5 traces how the constructs evolved across the model family.

Table 10.5. Evolution of constructs in the technology acceptance model family (IDT = innovation diffusion theory. SCT = social cognitive theory).

Construct	TAM	TAM2	UTAUT	UTAUT2	Root constructs in earlier models
Usefulness / performance expectancy	Perceived usefulness	Perceived usefulness	Performance expectancy	Performance expectancy	Relative advantage (IDT), outcome expectations (SCT)
Ease / effort expectancy	Perceived ease of use	Perceived ease of use	Effort expectancy	Effort expectancy	Complexity (IDT)
Social influence	-	Subjective norm, image	Social influence	Social influence	Subjective norm (TRA, TPB)
Facilitating conditions	-	-	Facilitating conditions	Facilitating conditions	Perceived behavioral control (TPB), compatibility (IDT)
Hedonic motivation	-	-	-	Hedonic motivation	Intrinsic motivation (motivational model)
Price value	-	-	-	Price value	Consumer value trade-offs
Habit	-	-	-	Habit	Habit and automaticity research
Key moderators	-	Experience, voluntariness	Gender, age, experience, voluntariness	Gender, age, experience	-

10.5.3 Critiques and the future of acceptance research

DEBATE AND EVIDENCE

A paradigm in need of a shift? In a special issue of the *Journal of the Association for Information Systems* devoted to TAM, Bagozzi (2007) argued that the model's parsimony had become its weakness. He identified several shortcomings: the neglect of the gap between intention and actual behaviour. The treatment of adoption as a terminal goal rather than a means to users' broader goals. The absence of group, social, and cultural aspects of decision making, and the proliferation of ad hoc extensions that fragmented the field. He proposed grounding acceptance research in a richer theory of goal-directed decision making and self-regulation. In the same issue, Benbasat and Barki (2007) argued that TAM's popularity had diverted research away from the **IT artifact** itself: because perceived usefulness and ease of use are treated as black boxes, the research offers little guidance on *which design features* make a system useful or easy, and the stream of incremental extensions had produced dimin-

ishing returns. They called for research on the antecedents of usefulness in system design and on a wider range of post-adoption behaviours such as adaptation and learning.
UTAUT has drawn related criticisms: its many moderators make the full model difficult to replicate, and its general constructs can be applied to any technology without engaging with what is distinctive about it. Taken together, the critiques do not deny that usefulness and ease of use matter. They question whether further variations on the same constructs advance understanding.

BUSINESS CASE

Generative-AI tool adoption in firms (conceptual analysis). When a firm makes a generative-AI writing or coding assistant available to employees, the acceptance models suggest where adoption will succeed or stall. *Performance expectancy* depends on task fit: employees whose work involves drafting, summarizing, or routine code see clearer gains than those whose work is mostly relational. Field evidence from customer-support work found that access to a generative-AI assistant raised productivity on average, with the largest gains for less experienced workers (Brynjolfsson, Li, and Raymond 2025): suggesting that perceived usefulness may be highest, and acceptance fastest, among newer employees. *Effort expectancy* includes learning to write effective prompts and to verify outputs. *Social influence* operates through managers and peers who model use, or who signal, implicitly, that using the tool is a form of shortcut. *Facilitating conditions* include data-governance rules, integration with existing software, and clarity about what information may be shared with the tool. The critiques of TAM apply with force: concerns about accuracy, accountability, and job security fall outside the core constructs, and the most important question is often not *whether* employees use the tool but *how*, the post-adoption behaviours that Benbasat and Barki (2007) urged researchers to study. Diffusion theory adds that observability matters: firms that make successful uses visible (shared prompt libraries, internal case studies) are applying Rogers's attribute directly.

10.5.4 Using the acceptance models in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Performance expectancy will have a stronger effect on intention to use a generative-AI assistant for employees in roles with a high share of writing tasks than for employees in roles with a low share.
- *H2.* The effect of social influence on intention will be stronger when use is mandated, and will weaken with experience.
- *H3.* Habit will predict continued use of a consumer app over and above intention after six months of use.

Designs. Longitudinal surveys at pre-implementation, early use, and established use. Linkage of survey constructs to *logged* system use rather than self-reports. Field experiments that vary training, defaults, or social information. Multilevel models for employees nested in teams.
Measures. Validated scales for TAM, UTAUT, and UTAUT2 constructs (F. D. Davis 1989, Venkatesh et al. 2003, Venkatesh, Thong, and Xu 2012), perceived innovation characteristics (G. C. Moore and Benbasat 1991), objective usage logs, continuance (Bhattacherjee 2001).

COMMON PITFALL

Predicting intention and calling it adoption. Many studies measure the predictors and the intention in the same survey and stop there. Common-method variance inflates the relationships, and intentions translate imperfectly into use (Webb and Sheeran 2006). Measure actual use, preferably from system logs and at a later time, and test whether the model explains *behaviour*, not just stated plans.

10.6 Information Systems Success Model

THEORY CARD

Core claim. The success of an information system is multidimensional. **System quality**, **information quality**, and **service quality** shape **use** and **user satisfaction**, which in turn produce **net benefits** for individuals, groups, and organizations. Net benefits feed back into further use and satisfaction.
Origins. Information systems. DeLone and McLean (1992) synthesized the fragmented measures of IS success into six interrelated categories. DeLone and McLean (2003) updated the model in response to a decade of tests and critiques, notably Seddon (1997). Petter, DeLone, and McLean (2008) reviewed the evidence.
Level of analysis. Individual users, workgroups, and organizations.
Key constructs. System quality, information quality, service quality, intention to use and use, user satisfaction, net benefits (in the 1992 model, individual impact and organizational impact).
Boundary conditions. Use is an informative success measure only when it is at least partly voluntary. When use is mandated, satisfaction, the quality of use, and net benefits carry more information. The relative weight of each quality dimension depends on the system: service quality matters most when users depend on an IT function or provider for support.
Related theories. Technology acceptance model, UTAUT, Task-technology fit, Service quality gap model (Section 28.3), Expectancy disconfirmation (Section 27.1).

10.6.1 Origins and core logic

The acceptance models in this chapter explain whether people *use* a technology. Managers who fund systems also need to know whether the systems *work*: whether they deliver accurate information, satisfy users, and improve performance. By the late 1980s, IS researchers had measured "success" in dozens of incompatible ways, from system response time to user satisfaction to return on investment. DeLone and McLean (1992) called this the search for the **dependent variable** of IS research. Drawing on communication theory's distinction between the technical level (is the message transmitted accurately?), the semantic level (does it convey the intended meaning?), and the effectiveness level (does it influence the receiver?), they reviewed the empirical literature and organized its measures into six categories: system quality, information quality, use, user satisfaction, individual impact, and organizational impact. Their central argument was that these categories are *interdependent* and that studies should measure several of them rather than choosing one.

The original model was criticized for mixing two kinds of model. Seddon (1997) argued that it combined a process model (a system is built, then used, then has impacts) with a causal or variance model (more quality causes more use and more benefits), and that "use" was ambiguous: more use

can reflect a valuable system or an inefficient one that consumes users' time. DeLone and McLean (2003) responded with an updated model that made three changes. They added **service quality** to capture the support users receive from the IS function or provider. They distinguished **intention to use** (an attitude) from **use** (a behaviour). And they combined individual and organizational impacts into a single **net benefits** construct whose referent (individual, group, firm, society) should be chosen to fit the research question. Figure 10.8 shows the updated model.

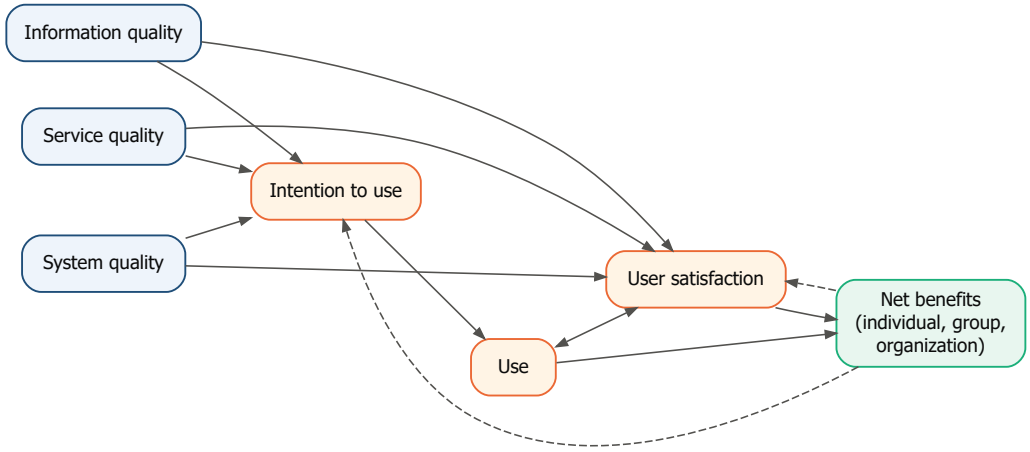


Figure 10.8. The updated DeLone and McLean model of information systems success (DeLone and McLean, 2003). Three quality dimensions shape intention to use, use, and user satisfaction, which jointly produce net benefits. Dashed arrows show the feedback of net benefits to subsequent use and satisfaction.

Table 10.6 lists the dimensions with typical measures. Because each dimension can be measured objectively (from logs and audits) or perceptually (from user surveys), the model also serves as a checklist for system evaluation.

Table 10.6. Dimensions of the updated DeLone and McLean IS success model with example measures.

Dimension	What it captures	Example measures
System quality	Desirable technical characteristics of the system	Reliability, response time, availability, ease of use, flexibility
Information quality	Desirable characteristics of the system's outputs	Accuracy, completeness, timeliness, relevance, format
Service quality	Quality of support from the IS function or provider	Responsiveness, assurance, empathy of support staff, resolution time
Intention to use / use	Extent and nature of system utilization	Frequency, duration, breadth of features used, logged transactions
User satisfaction	Users' overall evaluation of the system	Multi-item satisfaction scales, net promoter style ratings, complaints
Net benefits	Contribution to individuals, groups, organizations	Decision quality, productivity, cost savings, revenue, customer outcomes

10.6.2 Evidence and extensions

Petter, DeLone, and McLean (2008) reviewed tests of each relationship in the model. Support was strongest for links involving user satisfaction, for example from system quality and information quality to satisfaction and from satisfaction to net benefits at the individual level. Evidence was

weaker and more mixed for links involving system use, for service quality, and for net benefits at the organizational level, where many other factors intervene between a system and firm performance. The model is widely used as an evaluation framework for enterprise systems, e-commerce sites, e-government services, and knowledge management systems.

BUSINESS CASE

The launch of HealthCare.gov. When the US federal health-insurance marketplace opened on 1 October 2013, many visitors were unable to create accounts or complete applications because the site suffered from outages, slow responses, and errors. In the language of the model, deficiencies in **system quality** blocked **use** and destroyed **user satisfaction**, so that the intended **net benefits** (enrolment in coverage) could not materialize regardless of the quality of the plan information on the site. The recovery effort that followed concentrated on reliability and capacity, the system quality dimension, before the site could deliver its intended benefits. The episode illustrates why the model insists on measuring several dimensions: a system can be well specified at the information level and still fail at the technical level.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Service quality will have a stronger effect on user satisfaction for cloud software purchased by small firms (which lack internal IT support) than for large firms. *H2.* The effect of use on net benefits will be weaker when use is mandated. *H3.* Information quality will mediate the effect of a data-integration project on managers' decision quality.
Designs. Multi-wave surveys linked to usage logs and performance data, comparisons of business units before and after system implementation, structural equation models that test the full set of paths.
Measures. Validated scales for each quality dimension and satisfaction (Petter, DeLone, and McLean 2008), logged use, objective performance outcomes.

COMMON PITFALL

Measuring one dimension and calling it success. Many evaluations report only satisfaction or only usage. The model's main lesson is that the dimensions are distinct and can diverge: a heavily used system may be unsatisfying, and a satisfying system may produce no net benefit. Measure several dimensions and state which level of net benefits is the target.

10.7 Task-Technology Fit

THEORY CARD

Core claim. A technology improves individual performance when its capabilities fit the tasks its users must perform. Use alone is not enough: use of a poorly fitting technology adds little or nothing to performance.
Origins. Information systems. Goodhue and Thompson (1995) proposed the **technology-to-performance chain**. Goodhue (1995) examined user evaluations of fit. Extensions to group support systems (Zigurs and Buckland 1998) and integration with TAM (Dishaw and Strong 1999).

Level of analysis. Individual user and task. Extended to groups.
Key constructs. Task characteristics, technology characteristics, individual characteristics, task-technology fit, utilization, performance impacts.
Boundary conditions. Fit matters most when tasks are well defined and technology is a significant input to performance. Fit is judged relative to a specific task, so the same system may fit one task and not another.
Related theories. Technology acceptance model, UTAUT, IS success model, Contingency theory (Section 7.6).

10.7.1 Core logic

Goodhue and Thompson (1995) observed that acceptance models focus on utilization and that IS success models assume that more use produces more benefit. They argued that both overlook the *match* between technology and task. A spreadsheet is an excellent tool for financial modelling and a poor tool for managing a customer database. Their **technology-to-performance chain** combines two links. First, the fit between the requirements of a task and the functionality of a technology, shaped by the characteristics of tasks, technologies, and individuals, determines how much a technology can contribute to performance. Second, utilization depends on beliefs about the consequences of use, social norms, habit, and facilitating conditions, and these beliefs are themselves informed by fit. Performance impacts follow from fit and utilization together. The core propositions are:

- **P1.** Task-technology fit is a function of task characteristics, technology characteristics, and individual characteristics.
- **P2.** Task-technology fit influences beliefs about the consequences of use and therefore utilization.
- **P3.** Performance impacts depend on both utilization and task-technology fit. Utilization of a technology with low fit produces little benefit.

Figure 10.9 shows the chain.

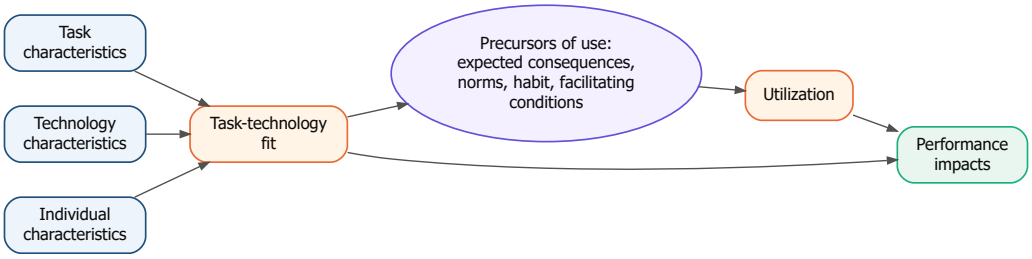


Figure 10.9. The technology-to-performance chain (after Goodhue and Thompson, 1995). Task-technology fit arises from task, technology, and individual characteristics. It shapes utilization through beliefs about the consequences of use, and it affects performance both directly and through utilization.

To measure fit, Goodhue and Thompson (1995) surveyed users across many departments and technologies and identified eight dimensions of user-evaluated fit, listed in Table 10.7.

Table 10.7. Eight dimensions of user-evaluated task-technology fit identified by Goodhue and Thompson (1995).

Dimension	What users evaluate
Quality	Currency, right level of detail, and accuracy of data for the task
Locatability	Ease of finding data and understanding what it means
Authorization	Having access to the data needed for the job
Compatibility	Consistency of data from different sources
Ease of use and training	Ease of using the system and adequacy of training
Production timeliness	Whether scheduled operations and reports arrive on time
Systems reliability	Dependability and availability of the system
Relationship with users	Understanding, interest, and responsiveness of IS staff

10.7.2 Evidence and extensions

In their survey, Goodhue and Thompson (1995) found that task-technology fit, together with utilization, predicted users' perceptions of the performance impacts of their systems, while support for the link from fit to utilization was weaker. The result supported their argument that fit is central to performance even when it does not strongly drive use. Dishaw and Strong (1999) combined TAM and TTF constructs in a study of software maintenance tools and found that the integrated model explained use better than either model alone. Zigurs and Buckland (1998) extended the logic to group support systems by classifying group tasks by complexity and specifying which combinations of communication support, process structuring, and information processing would fit each task type. Figure 10.10 illustrates the central implication: utilization pays off only when fit is high.

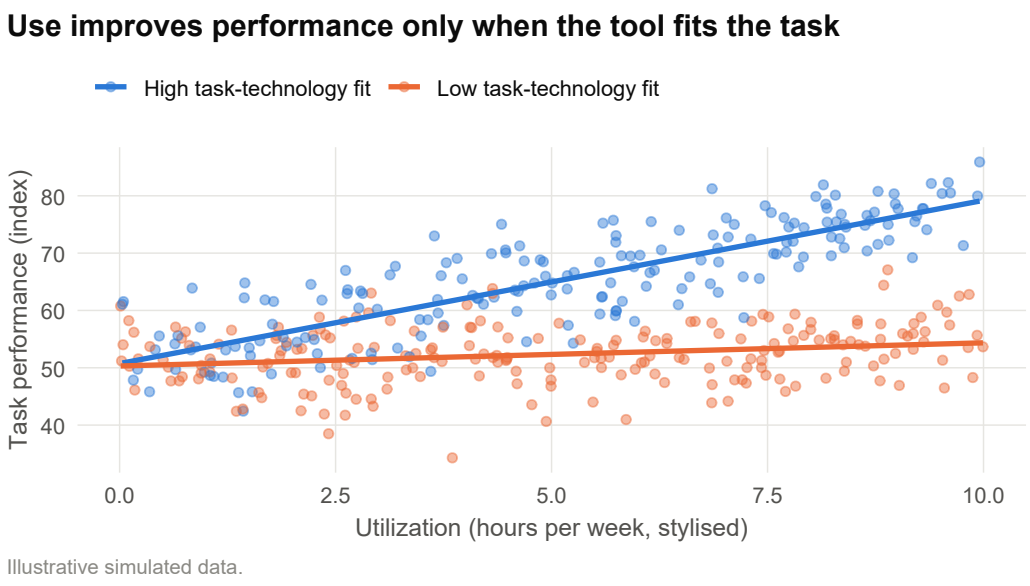


Figure 10.10. Illustrative simulation of proposition P3. For workers whose tasks fit the technology well, more utilization raises performance. For workers with low fit, more utilization adds little. Simulated data (n = 400), not results from a published study.

BUSINESS CASE

Illustrative scenario: deploying a generative-AI assistant. A professional services firm gives all consultants access to a generative-AI writing assistant and tracks adoption. Usage is high, but measured productivity gains are uneven. A task-technology fit analysis explains why. For drafting summaries and first versions of standard documents, task requirements (fluent text, quick turnaround) fit the tool's capabilities. For tasks that require verified client data or precise calculations, fit is low on the quality and authorization dimensions, and heavy use may even reduce performance if outputs are not checked. The managerial implication is to map tasks before mandating tools: encourage use where fit is high, redesign or restrict use where it is low, and invest in integrations (for example, secure access to internal data) that raise fit on the weakest dimensions.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* The effect of CRM utilization on salesperson performance will be stronger for salespeople whose tasks involve managing many accounts than for those managing a few strategic accounts. *H2.* Perceived task-technology fit will predict perceived usefulness, which will mediate its effect on intention to use. *H3.* Task-technology fit will moderate the effect of mandated use on performance.
Designs. Surveys of users across multiple tasks and technologies, multilevel models with tasks nested in individuals, field experiments that vary tool functionality for different task types, log data linked to output quality.
Measures. Goodhue and Thompson's fit dimensions adapted to the context, task characteristics (non-routineness, interdependence), utilization from logs, objective and perceived performance.

10.8 Chapter Summary

KEY TAKEAWAYS

- Diffusion is the communication of an innovation through channels over time among members of a social system. Cumulative adoption typically follows an **S-curve** (Rogers 2003).
- **Perceived attributes**, relative advantage, compatibility, complexity, trialability, and observability, shape the rate of adoption (Rogers 2003, Tornatzky and Klein 1982).
- Classic studies of hybrid corn and medical innovation established both the **economic** and the **social** drivers of diffusion, but apparent contagion can reflect common exposure or homophily (Griliches 1957, Van den Bulte and Lilien 2001, Aral, Muchnik, and Sundararajan 2009).
- **Moore's chasm** and the empirical **saddle** describe discontinuities between early and mainstream markets (G. A. Moore 1991, Goldenberg, Libai, and Muller 2002).
- The **Bass model** decomposes adoption into innovation (p) and imitation (q) and is a practical tool for forecasting from early sales or by analogy (F. M. Bass 1969, Mahajan, Muller, and Bass 1990).
- The **TRA and TPB** explain behaviour through intentions formed from attitudes, norms, and perceived control, but intentions translate imperfectly into behaviour (Ajzen 1991,

Webb and Sheeran 2006).

- **TAM**, **TAM2**, **UTAUT**, and **UTAUT2** identify usefulness, ease of use, social influence, facilitating conditions, hedonic motivation, price value, and habit as determinants of technology acceptance (F. D. Davis 1989, Venkatesh et al. 2003, Venkatesh, Thong, and Xu 2012).
- Critics argue that acceptance research should open the black box of the IT artifact and study post-adoption behaviour (Bagozzi 2007, Benbasat and Barki 2007).

Table 10.8 presents the comparison of the chapter's theories.

Table 10.8. Comparison of the chapter's theories.

Theory	Level	Key mechanism	Signature prediction	Typical business use
Diffusion of innovations	Social system	Perceived attributes, interpersonal communication	S-shaped cumulative adoption, early adopters influence the majority	Launch strategy, segmentation, policy design
Bass model	Market	Innovation (p) and imitation (q)	Bell-shaped new adoptions, peak at ln(q/p)/(p+q)	Sales and capacity forecasting
TRA / TPB	Individual	Attitudes, norms, perceived control -> intention	Intention predicts volitional behaviour	Behaviour-change campaigns, consumer research
TAM / TAM2	Individual	Perceived usefulness and ease of use	Usefulness dominates intention, ease works partly through usefulness	Enterprise system rollout, user-centred design
UTAUT / UTAUT2	Individual	Four core determinants plus hedonic, price, habit	Moderated effects by age, gender, experience	Workplace and consumer technology adoption
IS success model	Individual / organization	System, information, service quality -> use and satisfaction -> net benefits	Quality drives satisfaction and net benefits, with feedback to use	Evaluating enterprise systems and digital services
Task-technology fit	Individual / task	Fit between task requirements and technology functionality	Use improves performance only when fit is high	Tool selection, system design, AI deployment

10.9 Discussion Questions

1. Ryan and Gross emphasized social communication, Griliches emphasized profitability. Design a study that could estimate the relative contribution of each to the diffusion of a modern agricultural or business technology.
2. For a product you know well, rate each of Rogers's five attributes. Which attribute would you change first to accelerate adoption, and how?
3. Using the code in this chapter, explain why Bass-model forecasts made before the sales peak are unstable. What practical safeguards would you recommend to a product manager?
4. Van den Bulte and Lilien (2001) showed that marketing effort could explain apparent contagion. What data would you need to separate contagion, homophily, and common exposure in the adoption of a new software tool within a firm?

5. Bagozzi and Benbasat & Barki criticize TAM's focus on two general beliefs. Propose a research design that examines how a *specific design feature* of a generative-AI tool affects perceived usefulness and actual use.
6. Is Moore's chasm a distinct phenomenon or a consequence of heterogeneity that the Bass model already captures? Defend your answer with reference to
    1.

10.10 Further Reading

- Rogers (2003): the definitive synthesis of diffusion research. The starting point for any study of adoption.
- Griliches (1957): a classic economic analysis of hybrid corn that shows how profitability shapes the diffusion curve.
- Van den Bulte and Lilien (2001): a model reanalysis that every student of social contagion should read.
- F. M. Bass (1969): the original Bass model, brief and still widely used.
- Peres, Muller, and Mahajan (2010): a modern review of diffusion models, including network effects and cross-country diffusion.
- Ajzen (1991): the theory of planned behavior in the author's own words.
- Venkatesh et al. (2003): the unified theory, with a careful comparison of the eight models it integrates.
- Benbasat and Barki (2007): a short, influential critique of the TAM research programme.

Chapter 11

Strategy, Competition, and Industries

Why do pharmaceutical companies and soft-drink makers earn persistently higher returns than airlines and steel producers? Why did U.S. airlines that met each other on many routes charge higher fares than airlines that met only occasionally, and why did the rest of the U.S. wireless industry follow when T-Mobile abandoned two-year service contracts in 2013? Why do some pioneers, such as eBay in online auctions, keep their lead for decades while others, such as Friendster and MySpace in social networking, are overtaken by later entrants? And why do managers invest in small pilot projects that, judged one at a time, fail the standard net-present-value test?

Each question concerns **competitive strategy**: how firms position themselves within industries, anticipate rivals, and decide when and how much to commit. We begin with the industrial-organization tradition, the **structure-conduct-performance paradigm** and the theory of **strategic groups**. We then examine rivalry through the empirical lens of **competitive dynamics** and the formal lens of **game theory**. Two theories address the timing and staging of commitments: **first-mover advantage** and **real options theory**. The chapter closes with two behavioural theories of the firm's decision process, the **behavioral theory of the firm** and the **attention-based view**. Together they span the classic tension in strategy between models of rational rivals and models of boundedly rational organizations (Table 11.1).

Table 11.1. Chapter at a glance: theories of strategy, competition, and industries.

Theory	Core question	Level of analysis	Key source
Structure-conduct-performance	Why do some industries earn higher profits than others?	Industry	Mason (1939), Bain (1956)
Strategic groups	Why do clusters of firms within one industry follow similar strategies and earn different returns?	Group within industry	Caves & Porter (1977)
Competitive dynamics	Who attacks whom, and who responds, how fast, and how strongly?	Firm dyad	Chen (1996)
Game theory	What should a firm do when its payoff depends on what rivals do?	Interacting firms	von Neumann & Morgenstern (1944), Nash (1950)
First-mover advantage	When does entering a market first pay off?	Firm within market	Lieberman & Montgomery (1988)

Table 11.1. Chapter at a glance: theories of strategy, competition, and industries. *(continued)*

Theory	Core question	Level of analysis	Key source
Real options theory	How should firms invest when the future is uncertain and investments are irreversible?	Investment project, firm	Myers (1977), Bowman & Hurry (1993)
Behavioral theory of the firm	How do performance relative to aspirations, search, and slack shape organizational decisions?	Organization	Cyert & March (1963)
Attention-based view	How does the allocation of managerial attention shape what firms do?	Organization, decision maker	Ocasio (1997)

11.1 Structure-Conduct-Performance

THEORY CARD

Core claim. The structure of an industry (concentration, barriers to entry, product differentiation) shapes the conduct of the firms within it (pricing, advertising, investment), and conduct determines industry performance (profitability and efficiency). Industry is the primary unit for explaining why profits differ.

Origins. Industrial organization economics. Edward Mason's Harvard program of industry studies (E. S. Mason 1939) and Joe Bain's work on concentration and entry barriers (Bain 1951, 1956). Brought into strategic management by Michael E. Porter (1980b) and Michael E. Porter (1981).

Level of analysis. Industry (market), with firms treated as largely similar within it.

Key constructs. Seller concentration, barriers to entry (economies of scale, product differentiation, absolute cost advantages, capital requirements), the condition of entry, conduct, industry profitability.

Boundary conditions. Industry boundaries must be definable. Structure must be relatively stable and exogenous to any single firm's conduct. The theory is weakest when firm-level efficiency differences are large or when industries are reshaped quickly by technology.

Related theories. Five forces (Section 31.4), Strategic groups, Resource-based theory (Section 4.1), Organizational ecology (Section 7.5).

11.1.1 Origins and intellectual history

The structure-conduct-performance (SCP) paradigm grew out of the Harvard industrial organization program of the 1930s. E. S. Mason (1939) argued that the policies of large firms could not be understood through perfect competition alone and called for industry-by-industry study of how market structure conditions business behaviour. Joe Bain, trained in the same Harvard tradition, turned the program into a testable agenda: in U.S. manufacturing of the late 1930s, profit rates were higher in the most concentrated industries (Bain 1951).

Concentration alone, however, cannot sustain high profits if outsiders can enter freely. Bain (1956) therefore introduced the concept of the **condition of entry**: the extent to which established firms

can raise prices above the competitive level without attracting new entrants. He identified the structural sources of **barriers to entry**, which Table 11.2 summarises alongside later additions. The welfare logic of the paradigm was clear: high barriers and high concentration enable collusive or tacitly coordinated conduct, which produces high profits at the expense of consumers. SCP research therefore became the intellectual foundation of mid-century antitrust policy.

Table 11.2. Barriers to entry in the SCP tradition. The first four follow Bain (1956). The last two are later additions emphasised in strategy and platform research.

Barrier	Why it deters entry	Business example
Economies of scale	An entrant must either enter at large scale (and depress prices) or accept a cost disadvantage	Commercial aircraft manufacturing, semiconductor fabrication
Product differentiation	Buyers prefer established brands, so entrants must spend more on advertising or accept lower prices	Carbonated soft drinks, cigarettes, breakfast cereals
Absolute cost advantages	Incumbents control superior inputs, technology, or locations at any scale	Patented drugs, scarce airport slots, mineral deposits
Capital requirements	Large up-front investment raises financing costs and risk for entrants	Wireless spectrum and networks, steel mills
Switching costs (later addition)	Buyers incur costs to change suppliers, locking in installed bases	Enterprise software, banking relationships
Network effects (later addition)	The value of joining rises with the number of existing users	Payment networks, marketplaces, social platforms

11.1.2 Core logic and propositions

The paradigm runs from structure to conduct to performance, with basic conditions of supply and demand (technology, demand elasticity, growth) at the start of the chain. Figure 11.1 shows the logic, including the feedback loops that later critics emphasised.

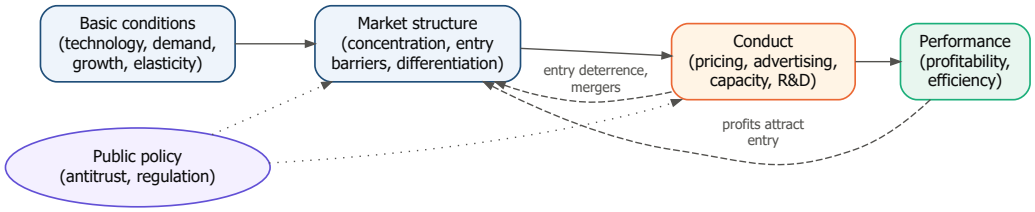


Figure 11.1. The structure-conduct-performance paradigm. Basic conditions shape market structure, which shapes firm conduct and, in turn, industry performance. Dashed arrows show the feedback effects (conduct reshaping structure, and performance attracting entry) stressed by later work.

The paradigm can be stated as four propositions:

- **P1 (Concentration).** Industries with higher seller concentration earn higher average profits, because fewer sellers find it easier to coordinate, tacitly or explicitly (Bain 1951).
- **P2 (Entry barriers).** High concentration translates into persistent profits only when barriers to entry are high (Bain 1956).
- **P3 (Industry primacy).** Because structure constrains conduct, industry membership explains a substantial share of the variance in firm profitability.
- **P4 (Strategic inversion).** For a firm, the attractive course is to enter industries with

favourable structure or to reshape structure in its favour, the inversion of the paradigm's welfare logic that Michael E. Porter (1981) made the basis of competitive strategy.

That inversion is the paradigm's most consequential legacy for business. Where the antitrust economist saw high barriers as a social cost, Michael E. Porter (1980b) saw them as a source of protected returns, and organized industry analysis into the five forces (Section 31.4) framework: rivalry, threat of entry, threat of substitutes, and the bargaining power of buyers and suppliers (Michael E. Porter 1979).

11.1.3 Evidence and extensions

The SCP paradigm met two major challenges. The first came from the Chicago school. Demsetz (1973) argued that the correlation between concentration and profits might reflect efficiency rather than collusion: if a few firms are genuinely better, they will grow large, the industry will become concentrated, and their superior profits will raise the industry average. Market power predicts that *all* firms in a concentrated industry gain. Efficiency predicts that mainly the large, efficient firms do. Demsetz reported evidence more consistent with the second view, opening a debate about whether profits reflect market power or superior capability that continues in resource-based theory (Section 4.1).

The second challenge came from variance decomposition. If structure drives profitability, industry membership should explain much of the variation in business-unit returns. Schmalensee (1985), using a single year of U.S. Federal Trade Commission line-of-business data, found that industry effects accounted for about a fifth of that variance, whereas corporate and market share effects were negligible or small. Rumelt (1991) used several years of the same data, which allowed him to separate stable business-unit effects from industry effects. Business-unit effects were several times larger than stable industry effects, a result widely read as evidence that firm-specific resources and strategies matter more than industry. McGahan and Porter (1997) and Hawawini, Subramanian, and Verdin (2003) qualified the conclusion, as Table 11.3 summarises, and Figure 11.2 stylises the contrast between the two research designs.

Table 11.3. Key variance-decomposition studies of firm profitability.

Study	Data	Headline finding	Interpretation
Schmalensee (1985)	FTC line-of-business data, single year	Industry effects explain about a fifth of the variance, corporate effects negligible	Supports the industry focus of SCP
Rumelt (1991)	FTC line-of-business data, several years	Stable business-unit effects are far larger than stable industry effects	Firm-specific strategy and resources matter most
McGahan & Porter (1997)	Compustat business segments, 1981-1994	Industry about 19 percent, business-specific about 32 percent, corporate about 4 percent	Industry still matters, and its importance varies by sector
Hawawini, Subramanian & Verdin (2003)	Value-based performance measures such as economic profit	Firm effects dominate for leaders and laggards, industry effects matter more for average firms	The answer depends on where a firm sits in its industry

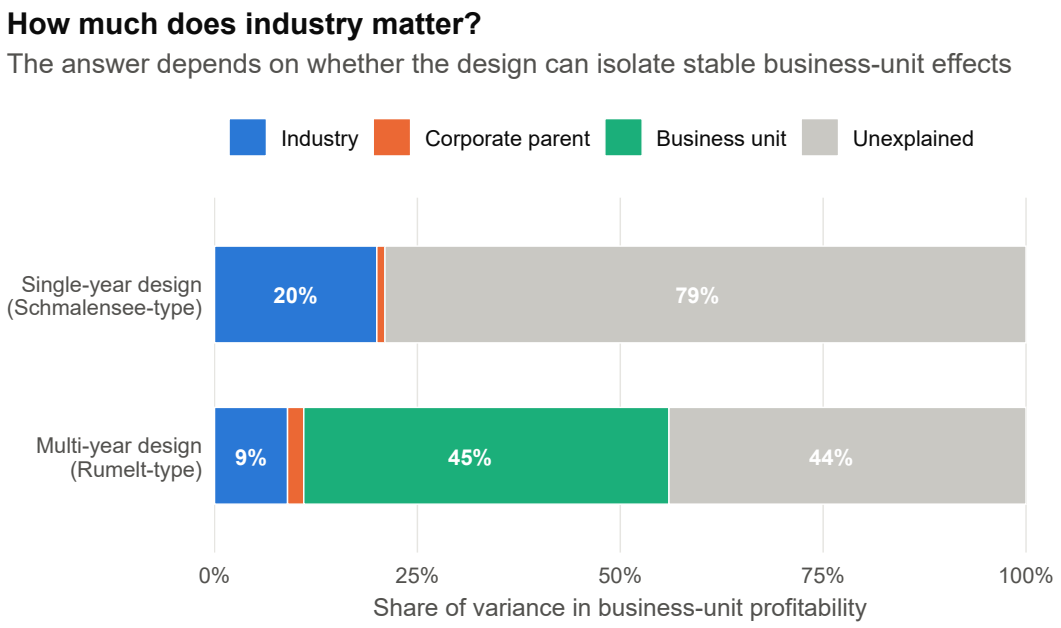


Figure 11.2. Stylised variance decomposition of business-unit profitability. A single-year design cannot separate business-unit effects from noise, so industry appears to be the main systematic source of profit differences. A multi-year design isolates stable business-unit effects, which then dominate. Shares are illustrative and are not the published estimates of any one study.

DEBATE AND EVIDENCE

Market power or efficiency? The SCP paradigm treats structure as exogenous, but efficient firms grow and make industries concentrated (Demsetz 1973). Variance-decomposition results depend on the sample, the performance measure, and whether stable and transient effects are separated (Rumelt 1991, McGahan and Porter 1997, Hawawini, Subramanian, and Verdin 2003). Moreover, decomposing variance says nothing about *causal mechanisms*: a large business-unit effect is a residual that could reflect resources, skill, luck, or a favourable niche. Most scholars now treat industry and firm effects as complementary explanations.

11.1.4 Applications in business

BUSINESS CASE

Why airlines struggle and soft drinks thrive. The U.S. airline industry is a classic case of unfavourable structure. Aircraft and airport access can be leased, so capital barriers are lower than they appear, fixed costs are high and marginal costs per seat are low, products are hard to differentiate, and buyers compare fares instantly. The industry's long history of bankruptcies and consolidation is what SCP would predict. Carbonated soft drinks show the opposite configuration: strong brands built over a century, bottler networks, and shelf and fountain contracts create formidable barriers, and the industry is dominated by two firms. The same paradigm explains why airlines have pursued consolidation, loyalty programmes, and control of hub airports: each is an attempt to change structure rather than merely to compete within

it.

11.1.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Industry concentration will be positively associated with price-cost margins, and the association will be stronger when entry barriers (measured by minimum efficient scale relative to market size) are high.
- *H2.* The share of variance in business-unit returns explained by industry effects will be larger in regulated and capital-intensive industries than in fragmented service industries.

Measures. Concentration ratios (CR4, CR8) and the Herfindahl-Hirschman Index, minimum efficient scale, advertising and R&D intensity, price-cost margins, return on assets, Tobin's q.
Designs. Variance decomposition with random-effects or multilevel models (Compustat segments, census concentration data), and within-industry natural experiments such as deregulation, patent expiries, and entry shocks.

COMMON PITFALL

Treating industry definitions as given. Industry codes are administrative categories, not markets. A four-digit code may lump together firms that do not compete (luxury and economy hotels) and separate firms that do (streaming services and cable television). Concentration measured at the wrong boundary is meaningless. Define the relevant market by substitution patterns before computing structure, and test the sensitivity of results to alternative definitions.

11.2 Strategic Groups

THEORY CARD

Core claim. Within an industry, firms cluster into **strategic groups** that follow similar strategies along key dimensions. Groups are protected from one another by **mobility barriers**, which can sustain profit differences between groups within the same industry.
Origins. Industrial organization and strategic management. The term originated in doctoral research on the U.S. home-appliance industry in the early 1970s. The theory of mobility barriers was developed by Caves and Porter (1977) and Michael E. Porter (1980b). Empirical and cognitive extensions by K. O. Cool and Schendel (1987), Porac, Thomas, and Baden-Fuller (1989), and Reger and Huff (1993).
Level of analysis. Groups of firms within an industry.
Key constructs. Strategic dimensions (scope, resource commitments), strategic group, mobility barriers, within-group rivalry, between-group rivalry, cognitive groups.
Boundary conditions. Firms must differ systematically along a small number of costly-to-change strategic dimensions. If strategies are easy to imitate, groups dissolve and mobility barriers vanish.

Related theories. Structure-conduct-performance, Competitive dynamics, Resource-based theory (Section 4.1), Institutional theory (Section 7.2).

11.2.1 Origins and core logic

The SCP paradigm treated firms in an industry as similar, but in home appliances full-line national brands, private-label suppliers, and narrow specialists faced the same structure yet behaved and performed differently. Caves and Porter (1977) generalised Bain's entry barriers into **mobility barriers**, which deter firms from moving between strategic groups as well as from entering the industry. A national brand, a dealer network, or a patent portfolio protects its owners from both outsiders and insiders in other groups. Because firms invest deliberately in such barriers, they are partly a product of strategy rather than a fixed feature of structure.

The theory yields four propositions:

- **P1 (Clustering).** Firms cluster along a few strategic dimensions, such as product scope, vertical integration, brand investment, and geographic reach.
- **P2 (Mobility barriers).** Moving between groups requires costly, slow, or uncertain investments, so membership persists (Caves and Porter 1977).
- **P3 (Performance differences).** Groups protected by higher mobility barriers earn higher returns.
- **P4 (Rivalry structure).** Rivalry is usually most intense within groups, and rivalry between groups depends on how much their markets overlap.

Figure 11.3 shows an illustrative strategic group map for a stylised airline industry.

11.2.2 Evidence and extensions

Strategic groups are easy to *find* with cluster analysis but harder to *validate*. K. O. Cool and Schendel (1987) studied the U.S. pharmaceutical industry from 1963 to 1982, defining groups by scope and resource commitments and identifying periods of stable group structure. Their results, like much subsequent work, suggested that performance differences between groups were less systematic than the mobility-barrier argument implied, and they later examined performance differences *within* groups (K. Cool and Schendel 1988). J. B. Barney and Hoskisson (1990) argued that the literature rested on two untested assertions: that groups exist as more than statistical artefacts, and that performance depends on group membership.

A productive response came from the cognitive tradition. Porac, Thomas, and Baden-Fuller (1989) showed that Scottish knitwear manufacturers defined their rivals narrowly, as other high-end classic knitwear producers in their region, and largely ignored the mass-market producers who dominated global volume. Reger and Huff (1993) found that bank executives perceived consistent cognitive groupings of their rivals. Cognitive groups explain why firms watch some rivals closely and ignore others, linking the theory to the attention-based view.

DEBATE AND EVIDENCE

Real structures or analyst constructs? Cluster algorithms will always produce clusters, so the existence of groups cannot be inferred from clustering alone (J. B. Barney and Hoskisson 1990). Critics also note that groups are defined on strategic choices and then used to explain

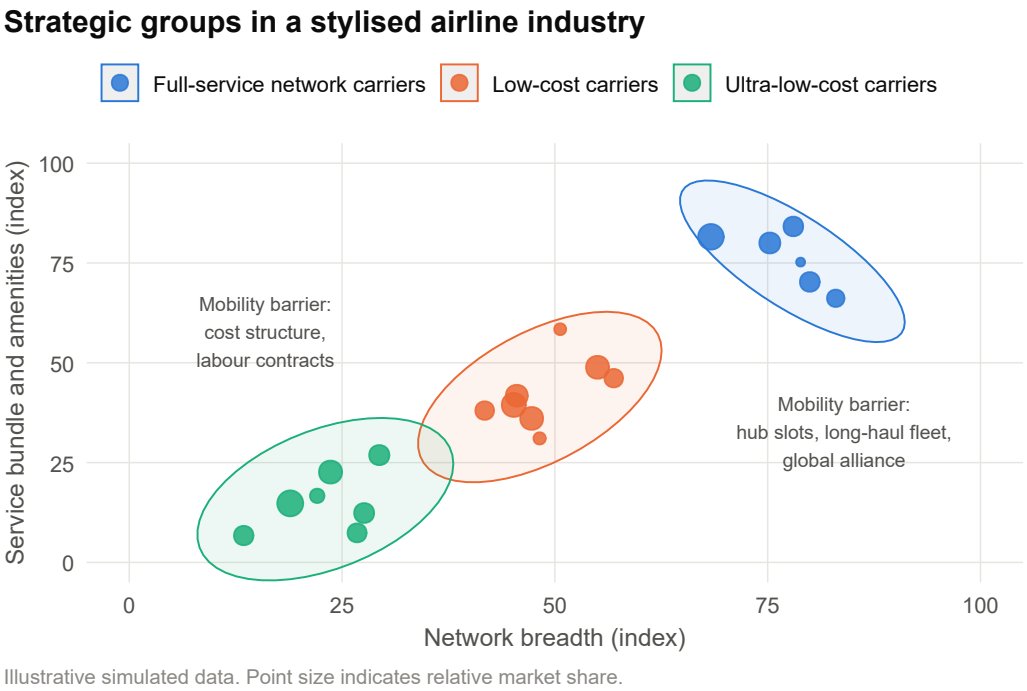


Figure 11.3. Illustrative strategic group map of a stylised airline industry. Firms cluster on network breadth and service bundle. Mobility barriers, such as hub slots and long-haul fleets, separate full-service network carriers from low-cost and ultra-low-cost carriers. Simulated data.

performance, which invites circularity when the choices themselves respond to performance. Defenders reply that groups are useful when they are defined *ex ante* on theoretically motivated dimensions, validated against managers' own perceptions (Reger and Huff 1993), and used to predict behaviour such as rivalry and response patterns rather than only profits.

11.2.3 Applications in business

BUSINESS CASE

Airline strategic groups. U.S. aviation offers a well-documented example of strategic groups. Legacy network carriers operate hub-and-spoke systems, long-haul international fleets, global alliances, and multi-class cabins. Low-cost carriers historically offered point-to-point service, a single aircraft type, and a simpler product, while ultra-low-cost carriers unbundle almost everything beyond the seat. Mobility barriers run in both directions: low-cost carriers cannot easily acquire congested hub slots or long-haul networks, and network carriers cannot easily replicate low-cost labour contracts and simplified operations. The network carriers' introduction of "basic economy" fares in the 2010s can be read as an attempt to compete with the lower groups without crossing the mobility barrier.

11.2.4 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms will respond more quickly to competitive moves by members of their own strategic group than to moves by firms in other groups.
- *H2.* The performance gap between strategic groups will be larger in industries where the investments that define groups are less reversible.
- *H3.* Managers' cognitive groupings of rivals will predict competitive responses better than groupings derived from cluster analysis of financial data.

Measures. Strategic dimensions (product scope, advertising intensity, R&D intensity, vertical integration, geographic scope), mobility barrier proxies (sunk investments, regulatory licences), cognitive maps elicited from managers through repertory grids or rival-naming tasks.
Designs. Cluster analysis with *ex ante* dimensions and stability tests, multilevel models of performance (firm within group within industry), event studies of group entry and exit, and text analysis of annual reports to recover self-defined competitor sets.

11.3 Competitive Dynamics

THEORY CARD

Core claim. Competition is a sequence of actions and responses between specific rivals. Whether a firm attacks a rival, and whether and how a rival responds, depends on the pair's **market commonality** and **resource similarity**, which shape each firm's **awareness**, **motivation**, and **capability** to act. Firms that meet in many markets often compete less aggressively because of **mutual forbearance**.
Origins. Strategic management, building on Austrian economics and industrial organization. M.-J. Chen, Smith, and Grimm (1992), M.-J. Chen (1996), and the review by M.-J. Chen and Miller (2012). Multimarket contact and mutual forbearance: Karnani and Wernerfelt (1985), Bernheim and Whinston (1990), Gimeno and Woo (1996).
Level of analysis. The competitive dyad (a pair of rivals), and the actions and responses between them.
Key constructs. Competitive action and response, market commonality, resource similarity, competitor asymmetry, awareness, motivation, capability (AMC), multimarket contact, mutual forbearance, spheres of influence.
Boundary conditions. Rivals must be able to observe each other's actions and to respond. The theory is most informative in industries with frequent, identifiable moves (price changes, product launches, route entries).
Related theories. Game theory, Strategic groups, Attention-based view, Resource-based theory (Section 4.1).

11.3.1 Origins and intellectual history

Classic industry analysis treats competition as a property of markets. Competitive dynamics research, which emerged in the late 1980s, treats it as a series of concrete moves (a price cut, a new route, a product launch) and the replies they provoke, often coded from trade publications on U.S.

airlines. M.-J. Chen, Smith, and Grimm (1992) found that actions affecting many competitors drew more responses, whereas **strategic actions** (large, hard-to-reverse commitments) drew fewer and slower responses than **tactical actions** such as price changes. D. Miller and Chen (1994) linked past success to **competitive inertia**, a narrower and less changeable competitive repertoire.

11.3.2 Core logic and propositions

M.-J. Chen (1996) integrated this work into a theory of **competitor analysis** built on two dyad-level constructs. **Market commonality** is the degree to which a competitor is present in the markets that matter to the focal firm. **Resource similarity** is the extent to which a competitor's strategic endowments resemble the focal firm's in type and amount. Because each firm's market portfolio differs, the two constructs are **asymmetric**: Firm A may be Firm B's main rival while B is a minor concern for A. These constructs operate through three behavioural drivers, known as the **awareness-motivation-capability** (AMC) perspective (M.-J. Chen 1996, M.-J. Chen and Miller 2012). A firm attacks or responds only if it is *aware* of the rival or the move, *motivated* to act (because the stakes are high), and *capable* of acting (because it has the resources).

Figure 11.4 summarises the mechanism.

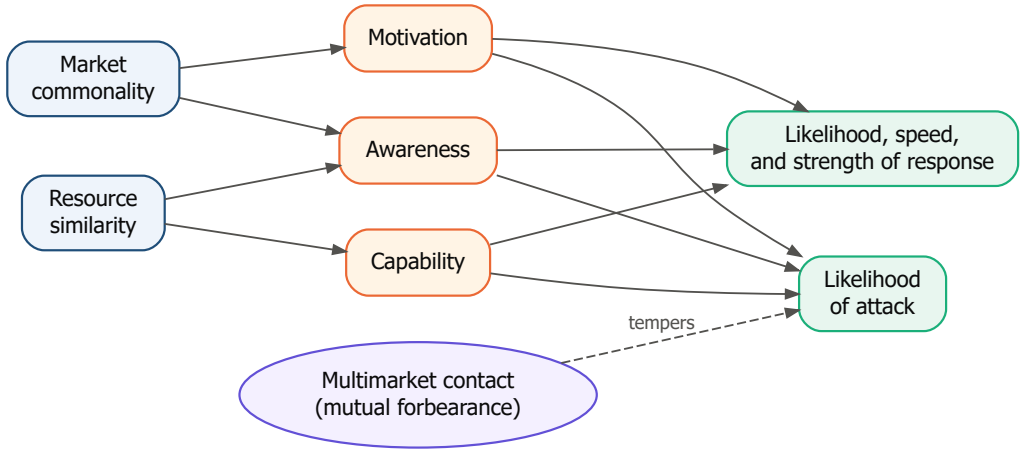


Figure 11.4. Competitive dynamics mechanism. Market commonality and resource similarity shape a firm's awareness of, motivation toward, and capability against a rival. These drivers determine the likelihood of attack and the likelihood, speed, and strength of response. Multimarket contact can temper rivalry through mutual forbearance.

The central propositions are:

- **P1 (Attack deterrence).** A firm is less likely to attack a rival with high market commonality, because the rival is highly motivated to retaliate across the many markets they share (M.-J. Chen 1996).
- **P2 (Response likelihood).** A firm is more likely to respond to an attack by a rival with high market commonality, because the attack threatens markets that matter to it.
- **P3 (Capability).** Resource similarity raises the rival's capability to respond in kind, which makes attack less attractive and response more feasible.
- **P4 (Action characteristics).** Responses are more likely and faster for visible, easily reversed, tactical actions than for complex, irreversible, strategic actions (M.-J. Chen, Smith, and Grimm

1992).

- **P5 (Mutual forbearance).** Firms that meet in many markets compete less aggressively in each, because aggression in one market invites retaliation in others (Karnani and Wernerfelt 1985, Bernheim and Whinston 1990).

Table 11.4 translates the constructs into measures.

Table 11.4. Key constructs in competitive dynamics research and typical operationalizations.

Construct	Definition	Typical measure
Market commonality	Overlap of a rival in markets that matter to the focal firm, weighted by importance	Share of focal firm's revenue earned in markets where the rival is present
Resource similarity	Comparability of strategic endowments in type and amount	Distance in resource profiles (fleet, assets, capabilities)
Awareness	Degree to which a firm notices a rival or its move	Visibility of the action, analyst or media coverage, managerial mentions
Motivation	Perceived stakes of acting or responding	Revenue at risk in the attacked market
Capability	Resources available to act or respond	Slack, scale, product breadth
Multimarket contact	Number of markets in which two firms meet	Count or share of overlapping routes, regions, or product markets
Response	A counter-move aimed at the attacker	Coded from trade press, response lag in days, number of responders

11.3.3 Evidence and extensions

The theory has strong empirical support in industries where moves can be observed. In airlines, W. N. Evans and Kessides (1994) found that fares were higher on routes served by carriers that met each other in many other markets, consistent with mutual forbearance. Gimeno (1999) showed that airlines with dominant positions in each other's hubs tended to respect one another's **spheres of influence**, and Gimeno and Woo (1996) found that multimarket contact de-escalated rivalry, whereas strategic similarity intensified it. Baum and Korn (1996) showed that market domain overlap and multimarket contact jointly shaped entry and exit among California commuter airlines. The formal foundation came from Bernheim and Whinston (1990), who showed that multimarket contact helps sustain collusion by pooling incentive constraints across markets: when markets or firms differ, slack in one market's incentive to cooperate can be transferred to another. When all markets and firms are identical, multimarket contact adds nothing. That result is an important boundary condition.

DEBATE AND EVIDENCE

Forbearance or collusion? Mutual forbearance raises a normative question that matters for merger review: is it efficient coexistence or tacit collusion that harms customers? Measurement is also contested, because actions coded from trade publications over-represent visible, newsworthy moves. Finally, the dyadic focus can neglect the wider network of rivals, who also observe every response (M.-J. Chen and Miller 2012).

11.3.4 Applications in business

BUSINESS CASE

Fare wars and spheres of influence in airlines. Airline competition offers a textbook record of action and response. In the spring of 1992, American Airlines introduced a simplified "Value Pricing" fare structure. Rivals answered with promotions of their own, and by the summer the industry was locked in a broad fare war that deepened losses across the sector. The episode illustrates P2 and P4: fare changes are visible, easy to reverse, and threaten rivals in many shared markets, so responses come quickly. By contrast, carriers have generally avoided head-on attacks on each other's fortress hubs, consistent with the spheres of influence documented by Gimeno (1999).

BUSINESS CASE

The Un-carrier and wireless rivalry. In 2013, T-Mobile US, then the fourth-largest U.S. wireless carrier, launched its "Un-carrier" strategy, beginning with the elimination of two-year service contracts and continuing with a sequence of publicly announced moves on pricing and plan features. The AMC perspective explains both the attack and the responses. As the smallest national carrier, T-Mobile had little to lose from disrupting industry pricing norms. Its moves were highly visible and threatened larger rivals' core subscriber bases, and those rivals also moved away from traditional contracts over the following years. Verizon reintroduced an unlimited data plan in early 2017.

11.3.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The likelihood that a firm responds to a rival's price cut will increase with the rival's market commonality with the focal firm.
- *H2.* Multimarket contact between two firms will be negatively associated with the frequency of price cuts in their shared markets, and the association will be stronger when the firms differ in market positions (asymmetric spheres of influence).

Measures. Structured content analysis of trade press for actions and responses, response lag, number of responders, market commonality and resource similarity indices, multimarket contact counts at the dyad level.

Designs. Dyad-level panel models, hazard models of response timing, and natural experiments from mergers that change multimarket overlap, using sources such as airline route and fare data, wireless plan histories, and retail price scrapes.

COMMON PITFALL

Ignoring asymmetry. A frequent error is to treat rivalry as symmetric, for example by computing a single overlap measure for a pair of firms. Market commonality and resource similarity are directional: what matters for the focal firm's behaviour is how much of *its* business the rival touches, not the reverse. Use directed dyads and test whether predictions hold separately for each direction.

11.4 Game Theory

THEORY CARD

Core claim. When the payoff of each actor depends on the choices of others, rational strategy requires anticipating rivals' responses. Outcomes can be predicted as **equilibria**, combinations of strategies from which no player gains by deviating unilaterally.
Origins. Mathematics and economics. Neumann and Morgenstern (1944) founded the field. J. F. Nash (1950) and J. Nash (1951) defined the equilibrium concept that bears his name. Applications to business strategy include Schelling (1960) on commitment, A. Dixit (1980) on entry deterrence, Tirole (1988) on industrial organization, and Brandenburger and Nalebuff (1996) on co-opetition.
Level of analysis. Interacting decision makers: firms in an oligopoly, negotiating parties, platform and complementors.
Key constructs. Players, strategies, payoffs, information, Nash equilibrium, dominant strategy, prisoner's dilemma, Cournot and Bertrand competition, sequential moves, commitment, credibility, repeated games, reputation, added value.
Boundary conditions. The number of relevant players is small enough that interdependence is recognised. Players are assumed to understand the game and to reason strategically. Predictions are sensitive to the timing of moves, the information structure, and the equilibrium concept chosen.
Related theories. Competitive dynamics, Collective action (Section 17.2), Auction theory (Section 17.8), Two-sided markets (Section 17.7), Signaling theory (Section 8.2).

11.4.1 Origins and intellectual history

Neumann and Morgenstern (1944) established game theory as a formal discipline, solving two-person zero-sum games. Most business situations, however, are not zero-sum: firms can both gain from restraint or both lose from a price war. J. F. Nash (1950) proved that every finite game has at least one equilibrium (possibly in mixed strategies) in which each player's strategy is a best response to the others', opening the analysis of non-cooperative games (J. Nash 1951). Older oligopoly models, Cournot's (1838) firms choosing quantities and Bertrand's (1883) firms undercutting each other's prices down to marginal cost, were recast as Nash equilibria and extended to sequential moves, incomplete information, and repeated interaction (Tirole 1988, Shapiro 1989).

11.4.2 Core logic: the prisoner's dilemma

The **prisoner's dilemma** is the canonical structure behind many competitive traps. Each firm has a dominant strategy to defect (cut price, increase advertising), yet both would be better off if both cooperated. Table 11.5 presents an illustrative pricing game between two soft-drink rivals.

Table 11.5. Illustrative scenario: a pricing prisoner's dilemma between two cola rivals. Cells show (Coca-Cola payoff, Pepsi payoff) in stylised profit units. Discounting is each firm's dominant strategy, so the one-shot equilibrium is mutual discounting even though mutual restraint pays both more.

	Pepsi holds price	Pepsi discounts
Coca-Cola holds price	10, 10 (joint best)	2, 14
Coca-Cola discounts	14, 2	5, 5 (Nash equilibrium)

Whatever Pepsi does, Coca-Cola earns more by discounting (14 > 10 and 5 > 2), and the same holds for Pepsi. The unique Nash equilibrium is mutual discounting, with payoffs of 5 each, although both would earn 10 under mutual restraint. The game explains why industries with few rivals do not automatically enjoy monopoly profits and why price wars, advertising escalation, and capacity races can persist even when every participant regrets them.

11.4.3 Core logic: from one-shot to repeated games

Real rivals meet repeatedly. In a repeated game, cooperation can be sustained if each firm credibly threatens to punish defection and if the future matters enough. With a "grim trigger" strategy (cooperate until the rival defects, then defect forever), cooperation is an equilibrium when the discount factor δ satisfies $\delta \geq (T - R)/(T - P)$, where T is the temptation payoff, R the reward for mutual cooperation, and P the punishment payoff. With the payoffs in Table 11.5, this threshold is $(14 - 10)/(14 - 5) \approx 0.44$. This logic underlies the **folk theorems** of repeated games and the forbearance results of the previous section.

Axelrod (1984) ran computer tournaments in which strategies submitted by researchers played a repeated prisoner's dilemma. The simplest entry, **tit for tat** (cooperate first, then copy the rival's last move), won. It was nice (never defected first), retaliatory, forgiving, and clear. In noisy environments, where moves are sometimes misread or mistakes occur, strict retaliation can trigger long echoes of punishment. Figure 11.5 simulates this problem.

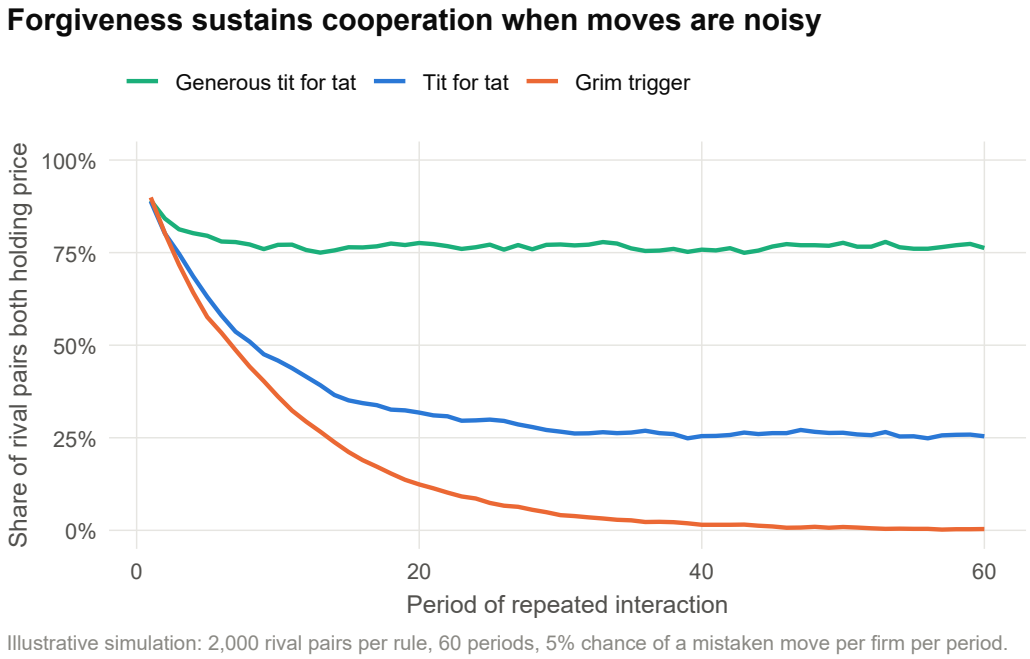


Figure 11.5. Illustrative simulation of a repeated pricing game with noise. Each period, each firm's intended move is reversed with probability 0.05 (a mistake or a misread price). Grim trigger collapses into permanent price war after the first error. Tit for tat sustains less cooperation than it does without noise because errors echo back and forth. Generous tit for tat, which forgives a defection one-third of the time, recovers cooperation quickly.

11.4.4 Core logic: commitment and the timing of moves

When moves are sequential, the first mover can shape the rival's choices by **commitment**: an irreversible action that changes its own future incentives. Schelling (1960) showed that limiting one's own options can pay if it makes threats credible. In A. Dixit (1980), an incumbent that invests in capacity before entry lowers its marginal cost of expanding output, which makes aggressive post-entry pricing credible and can deter entry. Ghemawat (1991) argued that strategy turns on a small number of such **sticky, lumpy, irreversible** commitments. **Reputation** works similarly: even a small probability that an incumbent is a "tough" type can make it rational to fight early entrants and so deter later ones (Kreps and Wilson 1982, P. Milgrom and Roberts 1982).

Table 11.6 maps the main models to business situations.

Table 11.6. Game-theoretic models and the business situations they illuminate.

Model	Strategic variable	Key prediction	Business situation
Prisoner's dilemma	Cooperate or defect	Mutual defection in one-shot play	Price wars, advertising escalation, promotions
Cournot oligopoly	Quantities or capacity	Prices above cost, falling as rivals increase	Commodity chemicals, memory chips, oil output
Bertrand competition	Prices for similar products	Prices driven toward marginal cost	Online retail of identical goods, generic drugs
Stackelberg leadership	Sequential quantity choice	The leader commits to large output and gains	Capacity pre-emption by incumbents
Entry deterrence	Capacity, pricing, reputation	Credible commitments deter entrants	Incumbent investment ahead of demand
Repeated games	Strategies over time	Cooperation if the future matters enough	Tacit price coordination, mutual forbearance
Co-opetition (value net)	Players, added values, rules, tactics, scope	Change the game rather than play it better	Complementors in platforms, joint standards

11.4.5 Co-opetition and changing the game

Brandenburger and Nalebuff (1996) translated game theory for managers through two ideas. The **value net** maps a firm's customers, suppliers, competitors, and **complementors** (players whose products make the firm's products more valuable), and the same firm can be both competitor and complementor. **Added value** is the amount by which total value would fall if a player left the game, and no player can capture more than its added value. The lesson is that firms often gain more by *changing the game* (its players, added values, rules, tactics, and scope) than by playing it more cleverly.

DEBATE AND EVIDENCE

Is game theory useful for strategy? Saloner (1991) defended formal models as disciplined thought experiments even when they yield no precise prescription. C. F. Camerer (1991) noted that predictions hinge on assumptions about rationality and common knowledge that experiments often contradict. A further criticism is that almost any outcome can be rationalised by some game, so the game must be specified before the data are examined. Defenders reply that the theory's value lies in diagnosing the *structure* of a situation, which then guides action.

11.4.6 Applications in business

BUSINESS CASE

The cola wars as a repeated game. Coca-Cola and PepsiCo have competed in carbonated soft drinks for more than a century, a textbook repeated game between two dominant players. Their rivalry has been conducted mostly through advertising, brand campaigns, and contracts with restaurant chains and bottlers rather than through sustained list-price wars. Pepsi's blind-taste "Pepsi Challenge," launched in the 1970s, was a highly visible attack. Coca-Cola's decision in April 1985 to replace its flagship formula with "New Coke," widely interpreted as a reaction to Pepsi's gains in taste tests, provoked a consumer backlash, and within about three months the original formula returned as "Coca-Cola Classic." In game-theoretic terms, heavy advertising has some prisoner's-dilemma features (each firm advertises partly to neutralise the other), while restraint in price competition is consistent with the repeated-game logic of Figure 11.5.

BUSINESS CASE

Illustrative scenario: credible capacity commitment. A regional cement producer learns that a rival plans to build a plant in its market. If the incumbent announces a price war, the threat is not credible: once the entrant has sunk its investment, fighting would hurt the incumbent too. If the incumbent instead completes a capacity expansion *before* the entrant decides, the calculation changes. With low marginal cost on its new capacity, the incumbent will rationally price aggressively after entry, and the entrant, anticipating this, may stay out (A. Dixit 1980). The commitment works only because it is irreversible and observable.

11.4.7 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Price cooperation among rivals will be greater in markets where firms expect to interact for longer (for example, lower probability of exit).
- *H2.* Incumbents will invest more in capacity ahead of demand in markets where the threat of entry is high and capacity is irreversible.
- *H3.* In laboratory pricing games, the introduction of noise in observed rival prices will reduce cooperation more under strict tit for tat than under forgiving strategies.

Designs. Formal modelling to derive predictions, structural estimation of oligopoly conduct, laboratory and online experiments with repeated games, and quasi-experiments from changes in information (price transparency regulation, public price comparison sites).
Measures. Price-cost margins, price dispersion, frequency and depth of price cuts, capacity announcements, entry and exit, and experimentally elicited cooperation rates.

COMMON PITFALL

Reasoning from the wrong game. The most common misuse is to apply a familiar game (usually the prisoner's dilemma) without checking that its assumptions hold. Are moves simultaneous or sequential? Is the interaction one-shot or repeated? Are payoffs known? Can

commitments be observed? A Bertrand model predicts cut-throat pricing, a Cournot model moderate margins, and a repeated game possible cooperation, all for the same industry. Specify the game from institutional facts before deriving a prediction.

11.5 First-Mover Advantage

THEORY CARD

Core claim. Pioneering firms can earn positive economic profits because early entry allows them to gain technological leadership, pre-empt scarce assets, and create buyer switching costs. Early entry also carries disadvantages, so the net advantage depends on the firm's resources and on how fast technology and markets evolve.
Origins. Strategic management and marketing. Lieberman and Montgomery (1988) and Lieberman and Montgomery (1998). Marketing evidence from W. T. Robinson and Fornell (1985), G. S. Carpenter and Nakamoto (1989), and the critique by Golder and Tellis (1993). Contingency theory (Section 7.6) of entry timing by Suarez and Lanzolla (2007).
Level of analysis. Firm within a new product market.
Key constructs. Order of entry, pioneer, early follower, technological leadership, pre-emption of assets, switching costs, free-rider effects, resolution of uncertainty, incumbent inertia, pace of technological and market evolution.
Boundary conditions. Advantages require mechanisms that let the pioneer appropriate value (learning that stays proprietary, scarce assets, switching costs, network effects (Section 17.6)). They weaken when technology or customer needs shift quickly or when imitation is cheap.
Related theories. Resource-based theory (Section 4.1), Network effects (Section 17.6), Dominant design (Section 12.5), Diffusion of innovations (Section 10.1), Liability of newness (Section 12.10).

11.5.1 Origins and core logic

The idea that being first pays is intuitive and was popular with managers long before it was a theory. Lieberman and Montgomery (1988) gave it theoretical structure by asking *how* a pioneer could earn positive economic profits and why the advantage would not be competed away. They identified three families of mechanisms and four countervailing disadvantages. Figure 11.6 presents the mechanism, and Table 11.7 gives examples.

Table 11.7. Sources of first-mover advantages and disadvantages, following Lieberman and Montgomery (1988).

Mechanism	Direction	Logic	Illustration
Technological leadership	Advantage	Learning-curve cost reductions and patents that stay proprietary	Early leaders in semiconductors riding cumulative output down the cost curve
Pre-emption of scarce assets	Advantage	Control of inputs, locations, shelf space, or capacity before rivals	Retail chains securing the best sites in a new region
Switching costs and network effects	Advantage	Buyers who adopt early are costly to lure away, and installed bases attract more users	eBay's early lead in online auctions in the United States

Table 11.7. Sources of first-mover advantages and disadvantages, following Lieberman and Montgomery (1988). *(continued)*

Mechanism	Direction	Logic	Illustration
Free-rider effects	Disadvantage	Followers imitate at lower cost and learn from the pioneer's market education	Platform owners entering categories that complementors have proven
Resolution of uncertainty	Disadvantage	Followers enter after technology and demand become clear	Late entrants building on a settled dominant design
Shifts in technology or needs	Disadvantage	The pioneer's assets fit the old technology or needs	Early social networks overtaken by later entrants
Incumbent inertia	Disadvantage	The pioneer becomes locked into early routines and investments	Early leaders slow to adopt the next technology generation

The core propositions are:

- **P1 (Appropriation).** First movers gain an advantage only when an isolating mechanism lets them appropriate the returns of early entry (Lieberman and Montgomery 1988).
- **P2 (Consumer learning).** In consumer markets, pioneers can shape buyer preferences and become the prototype against which later entrants are judged (G. S. Carpenter and Nakamoto 1989).
- **P3 (Resources and timing).** The optimal entry time differs across firms because firms differ in resources: a firm with strong commercialisation capabilities may profit from waiting (Lieberman and Montgomery 1998).
- **P4 (Environmental dynamics).** First-mover advantages are most likely when both technology and the market evolve smoothly, and least likely when technology changes abruptly and overturns the pioneer's position (Suarez and Lanzolla 2007).

11.5.2 Evidence and extensions

Early evidence was striking. Studies using the Profit Impact of Market Strategy (PIMS) database found that pioneers held larger market shares than later entrants (W. T. Robinson and Fornell 1985), and reviews of order-of-entry research found an inverse relationship between order of entry and market share in many consumer product categories (Kalyanaram, Robinson, and Urban 1995). G. S. Carpenter and Nakamoto (1989) explained why in psychological terms: the pioneer can define the attributes that consumers consider important.

Golder and Tellis (1993) challenged this record. PIMS and survey studies rely on surviving firms that classify themselves as pioneers, so failed pioneers vanish from the data and successful early followers may claim the label. Using historical records to identify the true pioneers in a set of product categories, Golder and Tellis found that nearly half had failed and that survivors' average share was much lower than earlier studies suggested. Early market leaders, who typically entered after the pioneer, performed better. A meta-analysis by VanderWerf and Mahon (1997) confirmed that methodological choices, including the use of market share as the outcome and the treatment of survivors, shape the observed size of first-mover effects.

Lieberman and Montgomery (1998) responded by connecting the theory to resource-based theory (Section 4.1). Entry timing is not an exogenous treatment but a strategic choice that reflects resources. Firms with strong technological capabilities may rationally enter early, and firms with strong marketing and distribution capabilities may rationally wait. Comparing early and late en-

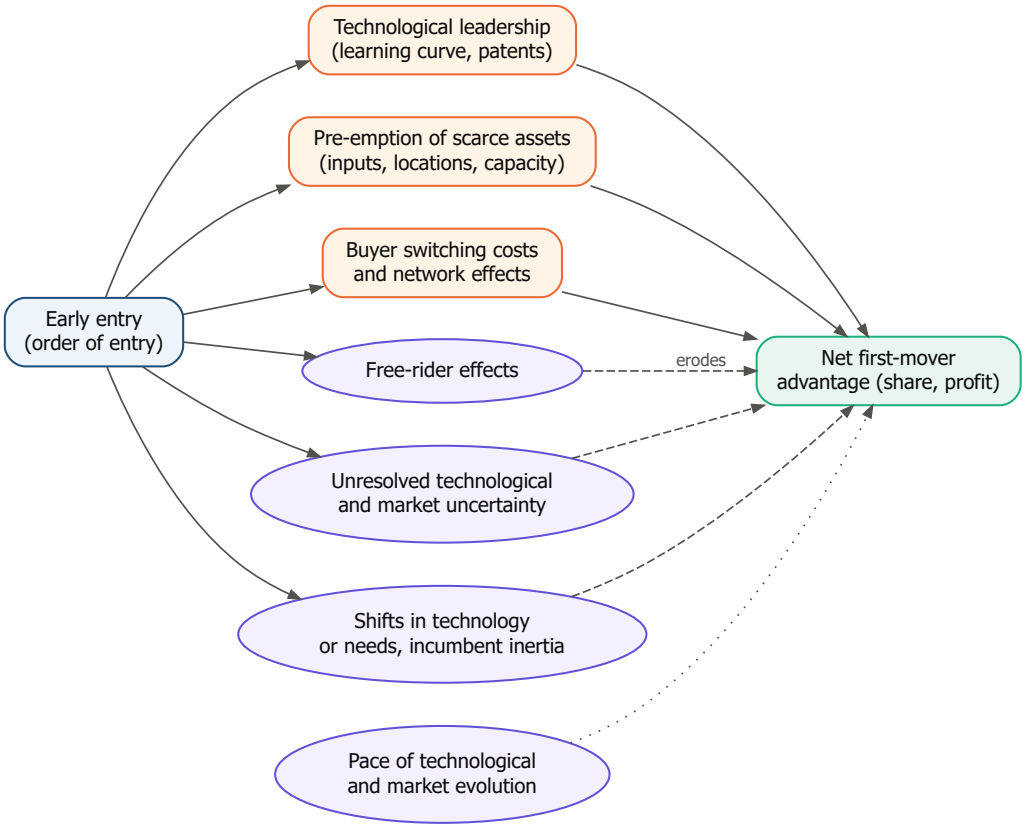


Figure 11.6. Mechanisms of first-mover advantage and disadvantage. Early entry can generate advantages through technological leadership, pre-emption of scarce assets, and buyer switching costs, but also exposes the pioneer to free riding, unresolved uncertainty, shifts in technology or needs, and its own inertia. The pace of technological and market evolution moderates the balance.

trants therefore confounds the effect of timing with the effect of the resources that led firms to choose that timing.

DEBATE AND EVIDENCE

Marketing logic or marketing legend? Survivor bias, self-reported pioneer status, and the endogeneity of entry timing all inflate apparent first-mover advantages (Golder and Tellis 1993, VanderWerf and Mahon 1997). Yet early entry does pay in some settings, particularly where network effects or switching costs are strong. The consensus is contingent: first-mover advantages exist but are neither universal nor automatic (Suarez and Lanzolla 2007).

11.5.3 Applications in business

BUSINESS CASE

Pioneers that held on and pioneers that did not. eBay, founded in 1995, was among the earliest online auction sites, and the network effects of a marketplace (buyers attract sellers and sellers attract buyers) allowed it to dominate U.S. online auctions for many years. Social

networking tells the opposite story. Friendster and MySpace attracted large audiences before Facebook, yet Facebook overtook both. Similarly, Google entered web search years after AltaVista and Yahoo. The mechanisms in Table 11.7 explain the difference. Where network effects were strong and technology stable, the pioneer's lead persisted. Where technology and user expectations shifted rapidly, better followers free-rode on the market education that pioneers had provided.

BUSINESS CASE

Platform entry decisions. Platform owners face a recurring timing choice: whether, and when, to enter the product spaces of their own complementors. F. Zhu and Liu (2018) studied Amazon's entry into product categories previously served by third-party sellers on its marketplace and found that Amazon was more likely to enter product spaces that had already proved popular and well rated. In the language of first-mover theory, the platform deliberately waits, allowing complementors to resolve market uncertainty, and then enters as a well-informed follower. For complementors, the implication is that being first on a platform carries a distinctive risk: success itself signals an attractive category to the platform owner.

11.5.4 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The market-share advantage of pioneers over early followers will be larger in categories with strong network effects than in categories without them.
- *H2.* After controlling for the selection of entry timing on firm resources, the effect of early entry on long-run survival will be smaller than in naive comparisons.

Designs. Historical identification of all entrants (including failures), selection models or instrumental variables for entry timing, matched comparisons of pioneers and followers with similar resources, and consumer experiments on pioneer prototypicality.

Measures. Order and date of entry, survival, long-run market share and profitability, switching-cost and network-effect proxies, pace of technical change (patent or performance-improvement rates).

COMMON PITFALL

Survivor bias. Studying only firms that still exist, or relying on firms to report whether they pioneered a market, systematically overstates first-mover advantages. Build the sample of entrants from historical sources before outcomes are known, include failures, and treat entry timing as a choice that must be modelled rather than as a random treatment.

11.6 Real Options Theory

THEORY CARD

Core claim. Many strategic investments create the right, but not the obligation, to take future actions. When the future is uncertain and investments are irreversible, such **real options** have value that static net present value ignores. Firms should invest small amounts to create options, keep them open while uncertainty resolves, and exercise or abandon them as information arrives.

Origins. Finance and strategic management. Myers (1977) framed growth opportunities as options on real assets. A. K. Dixit and Pindyck (1994) and Trigeorgis (1996) developed the theory of investment under uncertainty. Strategy applications by Kogut (1991), Bowman and Hurry (1993), McGrath (1997), and McGrath (1999). Boundaries examined by Adner and Levinthal (2004), reviewed by Trigeorgis and Reuer (2017).

Level of analysis. Investment project and the firm's portfolio of investments.

Key constructs. Uncertainty (volatility), irreversibility, flexibility, option to defer, stage, expand, abandon, or switch, exercise decision, growth options, option value.

Boundary conditions. Uncertainty must be resolvable over time through external events or the firm's own learning. The investment must be at least partly irreversible. Managers must be able and willing to exercise or abandon options on the basis of new information.

Related theories. Behavioral theory of the firm, Escalation of commitment (Section 14.2), Exploration and exploitation (Section 5.3), Portfolio theory (Section 16.1), Effectuation (Section 12.7).

11.6.1 Origins and intellectual history

Myers (1977) observed that much of a firm's value consists of **growth opportunities**: options to make future investments whose value depends on conditions not yet known. Treating them as call options on real assets gave rise to the term **real options**. Just as a call option gives the right to buy an asset at a fixed price, an R&D project, pilot plant, or joint venture gives a firm the right to invest more later if conditions prove favourable.

A. K. Dixit and Pindyck (1994) developed the economic theory. Their key result is that when an investment is irreversible and its value uncertain, the option to wait has value, so the firm should invest only when the expected value exceeds the investment cost by a margin that grows with uncertainty. The textbook rule "invest if NPV > 0" is wrong in such settings, because investing now *kills* the option to wait. Trigeorgis (1996) catalogued the main types of real options, which Table 11.8 summarises.

Table 11.8. Types of real options and business examples (organized following Trigeorgis, 1996).

Option type	What it allows	Business example
Option to defer	Wait for uncertainty to resolve before committing	Holding land or a licence without developing it
Option to stage	Invest in sequential steps, each conditional on the last	Drug development through clinical phases, venture capital rounds
Option to expand (growth option)	Scale up if the market proves attractive	Pilot plant, test-market launch, platform entry into a small market
Option to abandon	Exit and recover salvage value if outcomes are poor	Leasing rather than buying assets, kill criteria in R&D

Table 11.8. Types of real options and business examples (organized following Trigeorgis, 1996). *(continued)*

Option type	What it allows	Business example
Option to switch	Change inputs, outputs, or processes as prices change	Flexible manufacturing lines, dual-fuel power plants
Option to acquire	Buy out a partner when value becomes clear	Joint ventures that end in acquisition

11.6.2 Core logic and propositions

Real options reasoning makes three counterintuitive predictions. First, uncertainty can *increase* the value of an investment opportunity, because the option holder captures the upside and limits the downside. Second, small initial investments that look unprofitable on their own can be justified by the options they create. Third, the value of flexibility is greatest when uncertainty is high and when information arrives before the major commitment must be made. Figure 11.7 illustrates the second and third points.

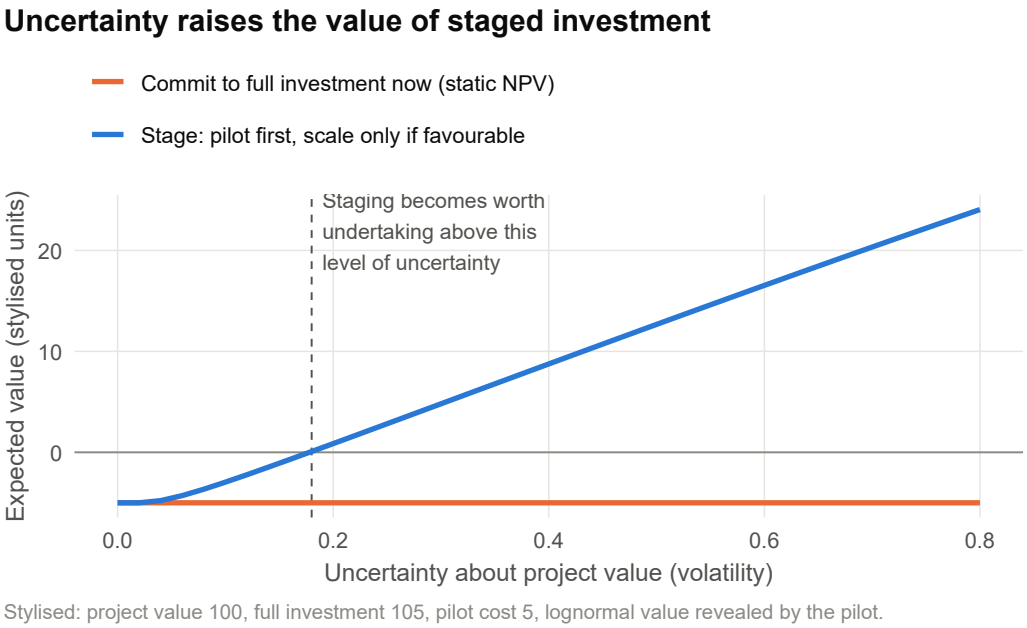


Figure 11.7. Stylised comparison of committing to a full investment now with staging it through a pilot. The full project has a static NPV of -5 at every level of uncertainty. Staging costs 5 up front but lets the firm scale only if the pilot reveals a favourable market. The value of staging rises with uncertainty, and above a threshold the staged project is worth undertaking even though the full project is not. Illustrative Monte Carlo values.

In propositional form:

- **P1 (Value of flexibility).** The value of an option to invest increases with the uncertainty of the underlying opportunity and with the time available before the decision must be made (A. K. Dixit and Pindyck 1994).
- **P2 (Staging).** Under high uncertainty, firms that invest in small, staged steps outperform firms that commit fully at the outset, provided that they abandon unpromising options (Mc-

Grath 1999).

- **P3 (Option chains).** Initial investments create **shadow options** that become recognised and valued only as firms learn, and options often come in chains in which exercising one creates the next (Bowman and Hurry 1993).
- **P4 (Acquisition timing).** Firms structure uncertain collaborations as joint ventures and exercise the option to acquire when the venture's value turns out to be higher than expected (Kogut 1991).

11.6.3 Evidence and extensions

Bowman and Hurry (1993) recast strategy as an **incremental-choice process** in which firms build portfolios of options through small resource investments and strike them when opportunities become clear. Kogut (1991) found that unexpected increases in the value of a joint venture's product market made acquisition by a partner more likely, consistent with the venture serving as an option to expand. McGrath (1999) argued that options reasoning reframes entrepreneurial failure as an expected outcome of a portfolio: the goal is to bound the cost of each failure and multiply the number of attempts, so that occasional large successes pay for many small losses.

Adner and Levinthal (2004) argued that real options logic is often misapplied. Financial options have fixed underlying assets and exercise prices, whereas real projects can be redefined along the way: when a technology disappoints, managers can always find a new application for it, which undermines the discipline of abandonment. A review by Trigeorgis and Reuer (2017) concludes that the theory is well supported where staging is clear (R&D, joint ventures, international expansion) but that measurement of option value and of managers' actual reasoning remains weak.

DEBATE AND EVIDENCE

Options or escalation? Real options reasoning assumes that managers will kill options when signals are negative. Behavioural research on escalation of commitment (Section 14.2) suggests that they often do not. Pilots become pet projects, sunk costs create pressure to continue, and the language of options can be used to justify investments that would otherwise be rejected. Critics also question whether real options valuation, built for traded financial assets, is meaningful when the underlying asset is not traded and its volatility cannot be estimated (Adner and Levinthal 2004). Proponents respond that the main value of the theory is *reasoning*, not precise valuation: framing decisions as options changes how firms structure investments, set milestones, and define abandonment criteria (McGrath 1999).

11.6.4 Applications in business

BUSINESS CASE

Staged R&D in pharmaceuticals. Drug development is a sequence of options. Discovery and preclinical work buy the right to start human trials, Phase I trials buy the right to conduct Phase II, and so on through Phase III and regulatory review. Each stage typically costs more than the last and resolves part of the technical and commercial uncertainty, and most candidates are abandoned along the way. A static valuation that ignores the right to stop understates the value of early-stage candidates. Staged venture capital rounds follow the same logic.

BUSINESS CASE

Illustrative scenario: entering a new market through a pilot. A consumer-goods firm is considering an irreversible factory investment in a new national market with highly uncertain demand and a slightly negative static NPV. Options reasoning suggests an import-and-distribution pilot with a local partner, explicit milestones (sales volume, retailer reorders), and a factory only if the milestones are met. The decisive question is whether the firm will actually walk away if they are missed.

11.6.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms will be more likely to use joint ventures rather than acquisitions to enter markets with high demand uncertainty.
- *H2.* The likelihood of acquiring a joint-venture partner will increase after positive, unexpected shocks to the venture's market value.
- *H3.* R&D projects with pre-specified kill criteria will be abandoned earlier after negative signals than projects without such criteria.

Measures. Uncertainty (volatility of industry returns, forecast dispersion), irreversibility (asset specificity, sunk-cost share), staging (number of investment rounds, milestone contracts), exercise and abandonment events.

Designs. Event histories of entry mode and exercise decisions (alliance databases, clinical trial registries, venture financing rounds), laboratory experiments on abandonment, and simulation models of option portfolios.

COMMON PITFALL

Calling everything an option. Any investment can be described as creating future opportunities, so the label "option" can rationalise almost anything. A real option exists only if uncertainty will resolve before the next decision, the next decision is genuinely discretionary, and the firm is prepared to abandon the position. If a pilot cannot produce decision-relevant information, or if the organization will never walk away, it is not an option but a down payment on a commitment already made (Adner and Levinthal 2004).

11.7 Behavioral Theory of the Firm

THEORY CARD

Core claim. Firms are coalitions of boundedly rational actors that pursue multiple goals expressed as **aspiration levels**. When performance falls below aspiration, firms engage in **problemistic search** and become more willing to change. When performance exceeds aspiration, **slack** accumulates and change is driven more by slack and less by problems. Aspirations adapt to past performance and to the performance of peers.

Origins. Organization theory and economics at the Carnegie Institute of Technology. Cyert and March (1963), building on Simon (1955). Performance feedback theory developed by

Greve (1998) and Greve (2003). Reviewed by Gavetti et al. (2012) and Shinkle (2012).
Level of analysis. Organization, with micro-foundations in individual and group decision making.
Key constructs. Organizational coalition, aspiration level (historical and social), performance feedback, problemistic search, slack search, organizational slack, quasi-resolution of conflict, uncertainty avoidance, standard operating procedures, adaptive learning.
Boundary conditions. Goals must be measurable and salient. The theory assumes that decision makers compare performance with reference points and that search is costly. Predictions weaken near the point of organizational survival, where attention shifts from aspirations to avoiding failure (March and Shapira 1992).
Related theories. Bounded rationality (Section 23.1), Prospect theory (Section 24.2), Organizational learning (Section 14.3), Exploration and exploitation (Section 5.3), Attention-based view, Garbage can model (Section 14.1).

11.7.1 Origins and intellectual history

Neoclassical economics treated the firm as a single actor maximising profit with complete information. Cyert and March (1963) replaced this abstraction with a model of how real firms make decisions about price, output, and resource allocation. Building on Simon (1955)'s concept of **bounded rationality (Section 23.1)**, they described the firm as a **coalition** of participants (managers, workers, shareholders, suppliers) with partly conflicting goals. Goals are not maximised but expressed as aspiration levels: targets that define satisfactory performance. Four relational concepts organized the theory:

- **Quasi-resolution of conflict.** Conflicting goals are attended to sequentially and in separate subunits rather than reconciled.
- **Uncertainty avoidance.** Firms prefer short-run feedback and negotiated environments to long-run forecasting.
- **Problemistic search.** Search is triggered by problems, begins near the symptom and current alternatives, and is biased by searchers' experience.
- **Organizational learning.** Goals and search rules adapt to experience.

Organizational slack, the excess of available resources over the payments needed to hold the coalition together, absorbs shocks, eases conflict, and funds experiments that would not survive scrutiny in lean times (Cyert and March 1963, Bourgeois 1981).

11.7.2 Core logic and propositions

Modern performance feedback theory formalises the model. Each period the firm compares its performance with an aspiration level that is a weighted average of its own past performance (the **historical aspiration**) and the performance of comparable firms (the **social aspiration**). Figure 11.8 shows the loop.

The main propositions are:

- **P1 (Problemistic search).** The further performance falls below aspiration, the more intensively firms search for solutions and the more willing they are to make risky changes (Cyert and March 1963, Greve 1998).

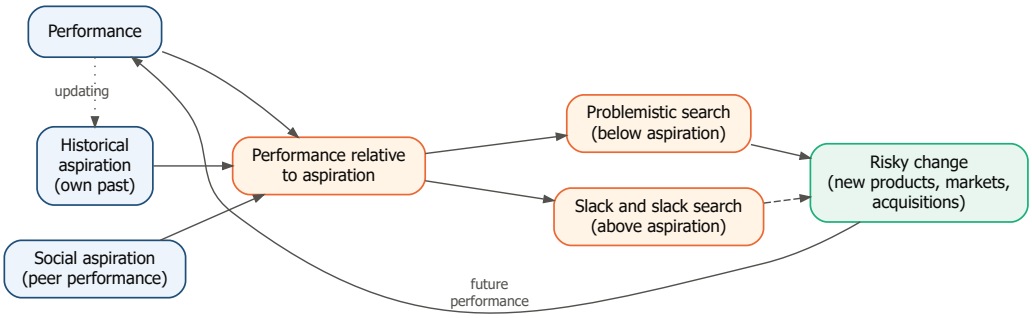


Figure 11.8. The behavioral theory of the firm as a performance feedback loop. Performance is compared with an aspiration level formed from the firm's own history and from peers. Shortfalls trigger problemistic search and greater willingness to take risks. Success builds slack, which funds slack search. Aspirations then adapt to experience.

- **P2 (Success and persistence).** As performance rises above aspiration, the likelihood of risky change declines, and past success breeds strategic persistence (Audia, Locke, and Smith 2000).
- **P3 (Slack).** Slack enables experimentation, so the relationship between slack and innovation is positive up to a point, beyond which slack breeds complacency and undisciplined projects (Nohria and Gulati 1996).
- **P4 (Adaptive aspirations).** Aspirations adjust toward recent own and peer performance, so persistent success raises the bar and persistent failure lowers it.
- **P5 (Survival focus).** Near the threshold of failure, attention can shift from the aspiration to survival, making firms more cautious than performance feedback alone would predict (March 1988, March and Shapira 1992).

Figure 11.9 shows the stylised shape that performance feedback research typically estimates.

11.7.3 Evidence and extensions

Performance feedback is one of the best-supported predictions in organization theory. Greve (1998) studied U.S. radio stations and found that stations were more likely to make risky changes of format when performance fell short of aspirations and less likely as performance rose above them, using both historical and social aspiration levels. Greve (2003) synthesised the program into a general theory of learning from performance feedback. Similar patterns appear for corporate risk taking (Bromiley 1991) and acquisitions (Iyer and Miller 2008). Audia, Locke, and Smith (2000) combined an archival study of firms facing deregulation with a laboratory study to show that past success increases strategic persistence after radical environmental change, even when persistence hurts performance. Nohria and Gulati (1996) found an inverted-U relationship between slack and innovation: too little slack starves experimentation, whereas too much reduces discipline.

The theory's reference-point logic parallels prospect theory (Section 24.2) (Kahneman and Tversky 1979), and its account of local search anticipates the trade-off between exploration and exploitation (Section 5.3) (March 1991) and the **myopia of learning** (Levinthal and March 1993). Reviews call for more attention to multiple, conflicting goals and to how firms choose the comparison group that defines their social aspiration (Gavetti et al. 2012, Shinkle 2012).

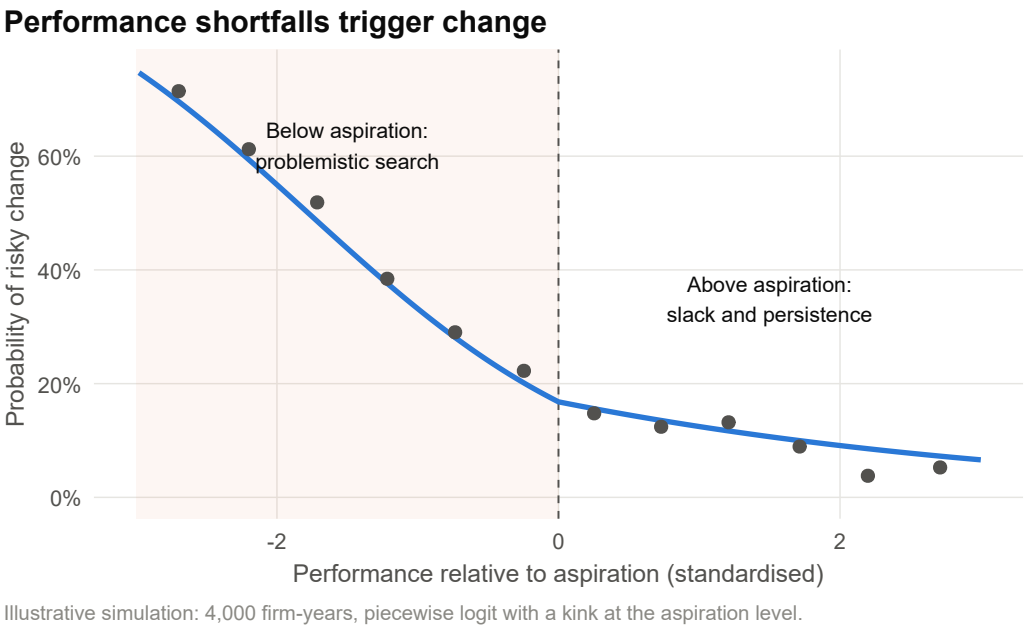


Figure 11.9. Stylised performance feedback relationship. The probability of a risky strategic change declines as performance rises relative to aspiration, with a change in slope (a kink) at the aspiration level. Points show binned averages from simulated firm-years, and the line shows the generating function. Illustrative data.

DEBATE AND EVIDENCE

Search or threat rigidity? The BTOF predicts that failure stimulates search and risk taking. A competing view holds that threat produces rigidity: under pressure, organizations centralise control and fall back on familiar routines. The two can be reconciled through the survival point (March and Shapira 1992): moderate shortfalls trigger search, whereas threats to survival trigger caution. A second debate concerns measurement. Aspiration levels are rarely observed and are usually computed from historical and peer performance with assumed weights, so results can be sensitive to how aspirations are constructed (Shinkle 2012).

11.7.4 Applications in business

BUSINESS CASE

Performance-feedback-driven change. The theory explains a familiar corporate pattern. Firms that miss their own targets, or fall behind peers, are more likely to reorganise, replace executives, launch acquisitions, or change strategy (Greve 1998, Iyer and Miller 2008). Firms that beat expectations tend to persist, which is efficient when the environment is stable and dangerous when it is not. Audia, Locke, and Smith (2000) showed that after deregulation transformed industries such as airlines and trucking, previously successful firms were slower to change their strategies. Practical implications follow. Managers can deliberately set aspirations by choosing comparison groups (a firm benchmarked against a struggling peer group will feel little pressure to change), and boards can counteract success-induced persistence by commissioning reviews when performance is strong rather than only when it is weak.

11.7.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The likelihood of a strategic change will increase as performance falls further below the social aspiration level.
- *H2.* Performance above aspiration will reduce the likelihood of risky change, and this effect will be weaker for firms with high absorbed slack.
- *H3.* The effect of performance shortfalls on risk taking will reverse as firms approach the threshold of financial distress.

Measures. Historical aspiration (exponentially weighted own past performance), social aspiration (peer mean or median), performance-aspiration gaps (spline coded above and below zero), slack (current ratio, selling and administrative expense ratio, unused borrowing capacity), and risky change (R&D intensity, acquisitions, entry into new markets, product changes).
Designs. Panel models with spline specifications at the aspiration level, event-history models of change, laboratory experiments that manipulate feedback, and agent-based simulations of aspiration adaptation.

COMMON PITFALL

Mechanical aspirations. Researchers often compute aspirations with arbitrary weights and comparison groups and then treat the resulting gap as an observed psychological reference point. Results can change with the weights. Report sensitivity to alternative specifications, justify the peer group, and where possible validate aspirations against stated targets (guidance, budgets, incentive thresholds).

11.8 Attention-Based View

THEORY CARD

Core claim. Firm behaviour is the result of how organizations channel and distribute the attention of their decision makers. What managers notice, and therefore what the firm does, depends on the issues and answers they focus on, the situation they find themselves in, and the rules, resources, and relationships that structure attention within the firm.
Origins. Strategic management and organization theory. Ocasio (1997), building on Simon (1955) and Cyert and March (1963), refined in Ocasio (2011).
Level of analysis. Organization and individual decision makers, linked through the firm's structures and communication channels.
Key constructs. Focus of attention, situated attention, structural distribution of attention, issues and answers, attention structures (rules of the game, players, structural positions, resources), procedural and communication channels, organizational moves.
Boundary conditions. Attention must be scarce relative to the issues competing for it. The view is most useful when outcomes depend on noticing and interpreting change rather than on executing a known plan.
Related theories. Behavioral theory of the firm, Upper echelons theory (Section 6.4), Sensemaking (Section 14.11), Dynamic capabilities (Section 5.1), Strategic groups.

11.8.1 Core logic

If rationality is bounded, attention is the scarce resource. Ocasio (1997) built a theory of the firm on three principles. **Focus of attention**: what decision makers do depends on the issues and answers they focus on. **Situated attention**: that focus depends on the context in front of them (meetings, reports, problems). **Structural distribution of attention**: the firm's rules, resources, players, and positions determine which issues reach which decision makers, through **procedural and communication channels** such as budget reviews, planning cycles, and executive committees. Organizational moves are the output of this attention-processing system (Figure 11.10).

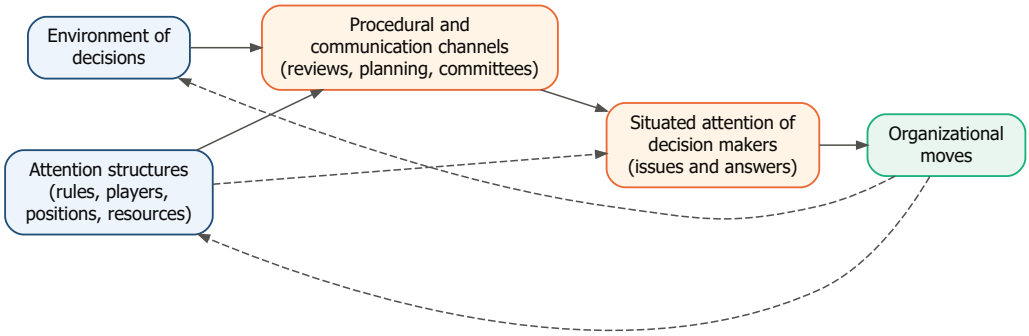


Figure 11.10. The attention-based view. The environment and the firm's attention structures (rules, players, positions, resources) feed issues and answers into procedural and communication channels. These channels situate the attention of decision makers, whose focus determines organizational moves, which in turn reshape the environment and the firm's structures.

The view yields three core propositions:

- **P1 (Selective focus).** Firms respond to the environmental changes that their decision makers attend to, not to all changes that matter (Ocasio 1997).
- **P2 (Structure channels attention).** The design of governance channels, committees, and reporting systems shapes which issues reach top management and therefore which adaptations occur (Joseph and Ocasio 2012).
- **P3 (Executive attention and adaptation).** When top managers attend to an emerging technology or market, firms adapt faster (Eggers and Kaplan 2009, Cho and Hambrick 2006).

11.8.2 Evidence and extensions

Cho and Hambrick (2006) studied U.S. airlines after the 1978 deregulation and found that the composition of top management teams influenced the extent to which executives attended to entrepreneurial issues, and that this attention mediated the relationship between team characteristics and strategic change. Eggers and Kaplan (2009) found that incumbent communications-equipment firms entered the emerging optical networking market faster when their CEOs paid more attention to the new technology. Joseph and Ocasio (2012) traced half a century of General Electric's history and showed that the architecture of cross-level governance channels shaped how the firm's attention, and with it adaptation, evolved. Tripsas and Gavetti (2000) showed that Polaroid developed strong digital imaging capabilities but that managerial beliefs anchored in its instant-film business model directed attention away from commercialising them. Ocasio (2011) refined the theory by distinguishing **attentional perspective** (the top-down schemas that make

some issues salient), **attentional engagement** (sustained, effortful allocation of attention), and **attentional selection** (the outcome of the process).

DEBATE AND EVIDENCE

Attention or cognition? Attention is hard to observe, and common proxies such as word counts in shareholder letters may reflect impression management. The view also overlaps with upper echelons theory, sensemaking, and managerial cognition. Its distinctive contribution is the *structural* claim: attention is distributed by the organization's channels and rules, not only by executives' personal characteristics (Ocasio 2011, Joseph and Ocasio 2012).

11.8.3 Applications in business

BUSINESS CASE

Redirecting a firm's attention. In May 1995, Bill Gates circulated an internal memo titled "The Internet Tidal Wave," which declared the internet the most important development for Microsoft since the IBM PC and directed managers across the company to make it a priority. Whatever one concludes about the subsequent strategy, the memo is a clear instance of a deliberate attempt to change the **focus of attention** of an entire organization. Intel offers a complementary structural example. Burgelman (1994) showed that Intel's exit from dynamic memory chips in the mid-1980s was prepared by middle managers who allocated scarce manufacturing capacity according to a margin-per-wafer rule, which steadily shifted resources toward microprocessors before top management formally decided to exit. The resource-allocation rule was an attention structure that channelled the firm toward its future.

11.8.4 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms whose top executives devote more attention to an emerging technology in public communications will enter the technology's market sooner.
- *H2.* The creation of a dedicated governance channel for a new issue (for example, a sustainability committee of the board) will increase the firm's subsequent investment in that issue.
- *H3.* Attention to competitors in executive communications will mediate the effect of market commonality on competitive responses.

Measures. Text-based attention measures from shareholder letters, earnings calls, and internal documents, agenda analysis of meetings, the structure of committees and reporting lines, time allocation of executives.

Designs. Longitudinal case studies of attention structures, panel models with text-based attention measures, and natural experiments in which reporting structures change.

11.9 Chapter Summary

KEY TAKEAWAYS

- The **structure-conduct-performance** paradigm explains profit differences by industry structure and entry barriers (Bain 1956). Variance-decomposition studies show that industry matters, but that firm-specific effects often matter more (Rumelt 1991, McGahan and Porter 1997).
- **Strategic groups** and **mobility barriers** explain persistent differences among firms within the same industry (Caves and Porter 1977), although the existence and performance consequences of groups remain contested (J. B. Barney and Hoskisson 1990).
- **Competitive dynamics** treats rivalry as action and response between specific rivals, shaped by market commonality, resource similarity, and the awareness-motivation-capability drivers. Multimarket contact can produce **mutual forbearance** (M.-J. Chen 1996, Gimeno 1999).
- **Game theory** identifies the structure of strategic interaction. One-shot dilemmas produce mutual defection, repeated interaction and credible commitment can sustain cooperation or deter entry (J. F. Nash 1950, A. Dixit 1980).
- **First-mover advantages** are real but contingent on isolating mechanisms, resources, and environmental dynamics, and early evidence was inflated by survivor bias (Lieberman and Montgomery 1988, Golder and Tellis 1993).
- **Real options theory** shows why uncertainty and irreversibility make staged investment valuable, provided that firms are willing to abandon options (A. K. Dixit and Pindyck 1994, McGrath 1999, Adner and Levinthal 2004).
- The **behavioral theory of the firm** explains change as a response to performance relative to aspirations, with slack and search as central mechanisms (Cyert and March 1963, Greve 2003).
- The **attention-based view** explains firm behaviour by how organizations structure and distribute the attention of decision makers (Ocasio 1997).

Table 11.9 presents the comparison of the chapter's theories.

Table 11.9. Comparison of the chapter's theories.

Theory	Unit of explanation	Key mechanism	Signature prediction	Typical business use
Structure-conduct-performance	Industry	Entry barriers and concentration enable coordination	Concentrated, protected industries earn higher profits	Industry attractiveness, antitrust analysis
Strategic groups	Group within industry	Mobility barriers	Persistent differences between groups in one industry	Competitor mapping, repositioning
Competitive dynamics	Rival dyad	Awareness, motivation, capability	Attack and response depend on commonality and similarity	Anticipating rival responses
Game theory	Interacting players	Equilibrium, commitment, repetition	Dilemmas, credible threats, tacit cooperation	Pricing, entry deterrence, co-opetition

Table 11.9. Comparison of the chapter's theories. *(continued)*

Theory	Unit of explanation	Key mechanism	Signature prediction	Typical business use
First-mover advantage	Entrant in new market	Isolating mechanisms versus free riding	Pioneers win only with appropriability	Entry timing decisions
Real options theory	Investment project	Value of flexibility under uncertainty	Staging beats commitment when uncertainty is high	R&D pipelines, pilots, joint ventures
Behavioral theory of the firm	Organization	Aspirations, search, slack	Shortfalls trigger change, success breeds persistence	Target setting, benchmarking, renewal
Attention-based view	Organization and executives	Structured allocation of attention	Firms adapt to what their structures make salient	Governance design, strategic agenda setting

11.10 Discussion Questions

1. Variance-decomposition studies reach different conclusions about how much industry matters. What features of the data and research design drive these differences, and what would a design that identifies the *causal* effect of industry structure on profitability look like?
2. Choose an industry you know well and draw a strategic group map. Which mobility barriers separate the groups, and which firms are best placed to cross them?
3. Using the awareness-motivation-capability framework, predict how incumbent carriers would respond to the entry of a low-cost airline into one of their hub markets. How would multimarket contact change your prediction?
4. The cola rivalry has involved intense advertising but relatively few sustained list-price wars. Use repeated-game logic to explain why, and identify the conditions under which a price war would become more likely.
5. Golder and Tellis argued that pioneering advantage was partly a "marketing legend." Design a study of entry timing in a recent technology market that avoids survivor bias and accounts for the endogeneity of entry timing.
6. A manager argues that every pilot project is a real option. Under what conditions is she right, and how would you design governance (milestones, kill criteria) so that the options logic and the behavioural tendency toward escalation do not conflict?

11.11 Further Reading

- Bain (1956): the classic statement of entry barriers and the foundation of the industrial-organization approach to strategy.
- Rumelt (1991) and McGahan and Porter (1997): the central exchange in the "how much does industry matter?" debate. Read them together.
- M.-J. Chen (1996): an integrative theory of competitor analysis that introduced market com-

monality, resource similarity, and competitor asymmetry.

- Brandenburger and Nalebuff (1996): an accessible translation of game theory into managerial strategy, built around the value net and the idea of changing the game.
- Lieberman and Montgomery (1988) and Golder and Tellis (1993): the theoretical case for first-mover advantages and the most influential methodological critique of the evidence.
- McGrath (1999) and Adner and Levinthal (2004): the case for real options reasoning in entrepreneurship and a careful statement of its boundaries.
- Cyert and March (1963) and Gavetti et al. (2012): the founding text of the behavioral theory of the firm and a modern assessment of its legacy and prospects.
- Ocasio (1997): the original statement of the attention-based view, with a framework that remains the starting point for research on organizational attention.

Chapter 12

Innovation and Entrepreneurship

In 1972 the British music and electronics company EMI unveiled the first commercial computed tomography (CT) scanner, the work of its engineer Godfrey Hounsfield, who later shared the 1979 Nobel Prize in Physiology or Medicine for the invention. Within a few years, established medical-equipment makers such as General Electric had overtaken EMI in the market it had created, and EMI soon left the business. Two decades later, Apple entered a mobile-phone industry it had not invented and captured a share of industry profits far larger than its share of units. Why do some pioneers capture the value of their innovations while others hand it to imitators? Why do well-managed incumbents so often lose to entrants with inferior products? And how do founders build ventures when the market they are aiming at does not yet exist?

This chapter assembles the theories that management scholars use to answer those questions. The first half concerns **innovation**: how new technologies replace old ones (**creative destruction**), why incumbents stumble (**disruptive** and **architectural innovation**), how industries settle on **dominant designs**, who profits from a new technology (**profiting from innovation**), and how firms draw on external knowledge (**open innovation**). The second half concerns **entrepreneurship**: how opportunities arise (**discovery and creation**), how expert founders reason under uncertainty (**effectuation**), how they build with whatever is at hand (**bricolage**), and why young organizations die at high rates (**liability of newness**). The theories complement those on Dynamic capabilities (Section 5.1) in Chapter 5 and on Diffusion of innovations (Section 10.1) in Chapter 10. Where those chapters ask how firms renew themselves and how adopters take up new products, this chapter asks where innovations come from, whom they destroy, and who captures their value (Table 12.1).

Table 12.1. Chapter at a glance: theories of innovation and entrepreneurship.

Theory	Core question	Level of analysis	Key source
Creative destruction	How does innovation drive economic change and growth?	Economy / industry	Schumpeter (1934, 1942), Aghion & Howitt (1992)
Disruptive innovation	Why do well-run incumbents lose to inferior entrants?	Firm / industry	Bower & Christensen (1995), Christensen (1997)
Profiting from innovation	Who captures the value an innovation creates?	Firm / industry	Teece (1986)
Architectural innovation	Why do small changes in product architecture defeat incumbents?	Firm / product	Henderson & Clark (1990)
Dominant design and technology cycles	How do industries converge on one design, and what follows?	Industry / technology	Abernathy & Utterback (1978), Anderson & Tushman (1990)

Table 12.1. Chapter at a glance: theories of innovation and entrepreneurship. *(continued)*

Theory	Core question	Level of analysis	Key source
Open innovation	When should firms use external ideas and paths to market?	Firm	Chesbrough (2003), Laursen & Salter (2006)
Effectuation	How do expert entrepreneurs decide under uncertainty?	Individual / venture	Sarasvathy (2001)
Opportunity discovery and creation	Do entrepreneurs find opportunities or make them?	Individual / market	Shane & Venkataraman (2000), Alvarez & Barney (2007)
Bricolage	How do resource-poor ventures build something from nothing?	Venture	Baker & Nelson (2005)
Liability of newness	Why do young organizations fail at higher rates?	Organization / population	Stinchcombe (1965), Freeman, Carroll & Hannan (1983)

12.1 Creative Destruction

THEORY CARD

Core claim. Capitalist economies develop through innovation, the carrying out of new combinations, which creates new products, firms, and industries while destroying the value of the old ones. Temporary monopoly profits from innovation are both the reward and the engine of the process.

Origins. Economics. Schumpeter (1934), first published in German in 1911, introduced the entrepreneur as the agent of "new combinations". Schumpeter (1942) coined the phrase *creative destruction* and shifted attention to the research laboratories of large firms. Aghion and Howitt (1992) formalized the idea in endogenous growth theory.

Level of analysis. Economy and industry, with the entrepreneur or innovating firm as the micro-level agent.

Key constructs. New combinations, entrepreneur, innovation rents, imitation, obsolescence, business stealing, quality ladder.

Boundary conditions. Strongest where innovations are rivalrous with existing products and where rents can be captured for a time. Weaker for innovations that complement rather than replace existing goods.

Related theories. Dynamic capabilities (Section 5.1), Disruptive innovation, Organizational ecology (Section 7.5), Competitive dynamics (Section 11.3).

12.1.1 Origins and intellectual history

Joseph Schumpeter broke with the equilibrium economics of his day by arguing that the essential feature of capitalism is not price competition among identical firms but change from within. In *The Theory of Economic Development* (Schumpeter 1934) he defined development as the carrying out of **new combinations**, and he listed five forms: a new good or a new quality of a good, a new method of production, the opening of a new market, the conquest of a new source of supply, and the reorganization of an industry, for example by creating or breaking up a monopoly position. The agent who carries out new combinations is the **entrepreneur**, a role defined by function rather than by ownership: a manager who introduces a new combination is acting entrepreneurially, while a capitalist who merely supplies funds is not.

Three decades later, *Capitalism, Socialism and Democracy* (Schumpeter 1942) gave the process its famous name. **Creative destruction** describes how new goods, methods, and organizational forms incessantly revolutionize the economic structure from within, destroying the old one and creating a new one. The book also changed the protagonist. The lone entrepreneur of 1911 gave way to the large corporation whose research laboratory routinizes innovation. Later scholars labelled these two pictures *Schumpeter Mark I* (entry by new firms drives innovation) and *Schumpeter Mark II* (established firms with large research budgets drive it). Whether small or large firms innovate more remains an active empirical question.

12.1.2 Core logic and propositions

The logic can be stated as a short chain. Innovation creates a temporary monopoly. The monopoly yields **innovation rents** (profits above the competitive level). The rents attract imitators and further innovators, who erode them and, in time, render the original innovation obsolete. Growth therefore comes in waves of innovation, imitation, and replacement rather than as a smooth accumulation of capital.

Aghion and Howitt (1992) built this idea into a formal model of economic growth. Research firms compete to produce the next improvement in an intermediate good. Each successful innovation makes the previous generation obsolete, so the innovator earns monopoly profits only until the next innovation arrives. Two implications follow. First, the expected value of research falls when future research is expected to be intense, because rents will be short-lived. Second, private and social returns to research diverge: innovators ignore the profits they destroy for incumbents (the **business-stealing effect**) and the benefits their ideas provide to future innovators (an intertemporal spillover). Equilibrium growth can therefore be either too fast or too slow relative to the social optimum.

- **P1 (Rents drive effort).** Innovative effort increases with the size and expected duration of the rents an innovation can earn.
- **P2 (Obsolescence).** Each successful innovation reduces the value of the assets, skills, and firms tied to the technology it replaces.
- **P3 (Turnover).** Industries with higher rates of innovation exhibit higher rates of entry, exit, and turnover in market leadership.
- **P4 (Competition and innovation).** The effect of product-market competition on innovation is non-monotonic: competition encourages firms to innovate to escape rivals but reduces the rents that reward innovation.

12.1.3 Mechanism

Figure 12.1 summarizes the cycle. The dashed feedback edge is the key to the dynamic: rents attract the next wave of innovators whose success destroys the current incumbents.

12.1.4 Evidence and extensions

The macroeconomic literature on Schumpeterian growth has grown into a large research programme that links innovation to firm entry, exit, and reallocation. An influential empirical extension concerns P4. Using panel data on British industries, Aghion et al. (2005) found an **inverted-U** relationship between product-market competition and patenting. When competition is weak, more of it spurs innovation because firms that innovate can escape their rivals. When

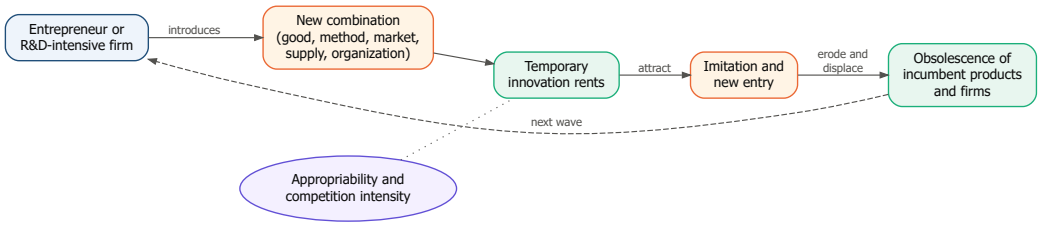


Figure 12.1. The cycle of creative destruction. A new combination yields temporary rents that attract imitators and new innovators. Their success erodes rents and makes the incumbent technology obsolete, restarting the cycle.

competition is already intense, more of it discourages innovation because laggards expect little reward from catching up. The result reconciles the Mark I view (competition spurs innovation) with the Mark II view (market power funds it), at least in the setting studied.

At the firm level, creative destruction underpins much of the strategy literature on technological change. Tushman and Anderson (1986) showed that discontinuous technologies alter the competitive landscape of entire industries, and the theories of disruptive, architectural, and competence-destroying change can be read as specifications of *when* the destructive half of the process falls on incumbents.

DEBATE AND EVIDENCE

Is destruction the right word? Critics note that many innovations complement rather than replace existing assets, and that incumbents often survive technological transitions by controlling complementary assets (Tripsas 1997, Teece 1986). In Tripsas's study of typesetting, incumbents survived three generations of competence-destroying technology when they owned specialized assets such as font libraries and sales networks. Creative destruction is therefore best treated as a tendency whose incidence on particular firms depends on the conditions specified by the more specific theories below. A second debate is normative: because the Aghion and Howitt model shows that equilibrium innovation can be either excessive or deficient, the theory does not by itself justify either stronger or weaker patent protection.

12.1.5 Applications in business

BUSINESS CASE

The smartphone as a destroyer of categories. The diffusion of smartphones after 2007 illustrates every element of Schumpeter's account. A single new good combined functions previously sold as separate products: standalone satellite-navigation devices, compact digital cameras, portable music players, and handheld game consoles. Each of these categories contracted as the smartphone absorbed its function, and firms built around them had to reinvent themselves or shrink. The same innovation created new industries (mobile apps, ride-hailing, mobile payments) and new rents for the firms that controlled the platform. For a strategist, the lesson is that the relevant competitive threat may come not from a rival in one's own category but from a new combination that makes the category unnecessary.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Industries with higher patenting intensity will exhibit higher annual turnover among the top four firms by market share.
- *H2.* The relationship between industry concentration and new-product introductions will be inverted-U shaped.

Designs. Firm and industry panels linking patents or product introductions to entry, exit, and market-share turnover. Natural experiments that change competitive pressure (trade liberalization, deregulation, patent-term changes).

Measures. Patent counts and forward citations, product-introduction databases, firm entry and exit from business registers, rank turnover among leaders, price-cost margins as a proxy for competition.

12.2 Disruptive Innovation

THEORY CARD

Core claim. Incumbents tend to win battles of **sustaining** innovation that improve products along dimensions their best customers value, but they tend to lose to **disruptive** innovations that start as inferior, cheaper, or more convenient offerings for overlooked customers and then improve until they satisfy the mainstream.

Origins. Strategy and technology management. Bower and Christensen (1995) introduced the idea in practitioner form. Christensen (1997) developed it from a study of the disk-drive industry. Christensen, Raynor, and McDonald (2015) restated the theory and its boundaries.

Level of analysis. Firm and industry, with resource-allocation processes inside the incumbent firm as the micro-mechanism.

Key constructs. Sustaining versus disruptive innovation, performance trajectories, performance oversupply, value network, low-end and new-market footholds, resource allocation.

Boundary conditions. Applies when technology improves faster than mainstream customers' needs, when the entrant begins in a segment incumbents find unattractive, and when incumbents' resource-allocation processes are tied to existing customers and margins. Does not describe every successful entrant.

Related theories. Exploration and exploitation (Section 5.3), Ambidexterity (Section 5.4), Resource dependence theory (Section 7.1), Architectural innovation.

12.2.1 Origins and intellectual history

Clayton Christensen began with a puzzle from the disk-drive industry. Leading manufacturers of 14-inch drives were technically excellent and closely attuned to their customers, yet they missed the shift to 8-inch drives, whose early products had lower capacity but suited the emerging minicomputer market. The same pattern repeated at the transitions to 5.25-inch drives (desktop computers) and 3.5-inch drives (portable computers). Christensen and Rosenbloom (1995) introduced the concept of the **value network**, the context of customers, suppliers, and cost structures within which a firm defines and solves problems, and argued that incumbents fail when a technology's initial value lies outside their value network. Christensen and Bower (1996) traced the mechanism to **customer power**: firms allocate resources to projects that their most profitable customers de-

mand, which starves projects aimed at small, low-margin markets. Christensen (1997) synthesized the argument in *The Innovator's Dilemma*.

12.2.2 Core logic and propositions

The theory rests on two trajectories. The first is the trajectory of **performance demanded** by customers in a market tier, which rises slowly. The second is the trajectory of **performance supplied** by a technology, which rises faster as firms pursue sustaining improvements. Because supply outpaces demand, incumbents eventually **overshoot** what mainstream customers can use. A disruptive entrant begins below mainstream needs but serves a foothold market, either the least demanding customers (a **low-end foothold**) or people who previously did not consume the product at all (a **new-market foothold**). As the entrant's technology improves, it becomes good enough for the mainstream while retaining its advantages in cost, simplicity, or convenience (Christensen, Raynor, and McDonald 2015).

- **P1 (Sustaining advantage).** Incumbents usually prevail in sustaining innovations, whether incremental or radical, that serve their existing customers.
- **P2 (Overshooting).** When performance supplied exceeds performance demanded, the basis of competition shifts toward convenience, customization, and price.
- **P3 (Asymmetric motivation).** Incumbents rationally decline to pursue foothold markets because those markets are small and offer lower margins than the core business.
- **P4 (Process, not event).** Disruption unfolds over years. Its outcome depends on whether the entrant's trajectory reaches mainstream needs before incumbents respond.

Figure 12.2 reproduces the classic diagram with stylised values.

Table 12.2 distinguishes the two footholds from sustaining innovation.

Table 12.2. Sustaining innovation and the two forms of disruption (after Christensen, Raynor, & McDonald, 2015).

Dimension	Sustaining innovation	Low-end disruption	New-market disruption
Target customers	Most demanding, most profitable customers	Over-served customers at the bottom of the market	Non-consumers who lacked money, skill, or access
Initial performance	Better on established attributes	Good enough on established attributes	Worse on established attributes, better on new ones
Business model	Existing margins and cost structure	Lower cost, lower price	New value network
Incumbent response	Fight and usually win	Retreat up-market to protect margins	Ignore, because the market looks small
Examples discussed in the literature	Successive gains in drive capacity	Minimills in steel	Personal computers relative to minicomputers

12.2.3 Evidence and extensions

The disk-drive evidence in Christensen (1997) remains the core of the theory, and Christensen and Rosenbloom (1995) and Christensen and Bower (1996) provide the more formal academic statements. R. Henderson (1993) offered a complementary analysis of photolithographic equipment, arguing that incumbents' failures reflected both rational **underinvestment**, because new technologies cannibalize existing rents, and organizational **incompetence**, because embedded routines hinder new learning. That distinction matters: a theory of incentives implies that incumbents *choose*

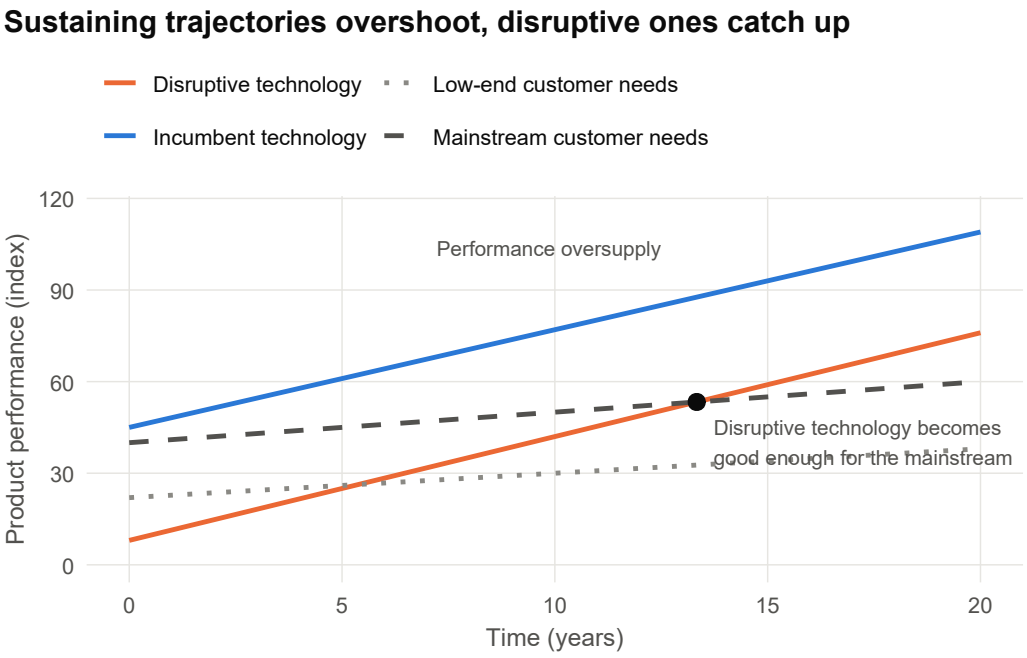


Figure 12.2. Stylised performance trajectories in disruptive innovation. Incumbents' sustaining improvements overshoot what mainstream customers can use, while a disruptive technology starts below mainstream needs and improves until it is good enough. Illustrative values only.

not to respond, whereas a theory of capabilities implies that they *cannot*. The managerial prescription in Christensen (1997), an autonomous unit with a cost structure suited to the new market, addresses both, and it anticipates the literature on structural ambidexterity (Section 5.4).

DEBATE AND EVIDENCE

How useful is the theory of disruptive innovation? In a widely read essay, Lepore (2014) criticized the theory's case selection, argued that several firms described as disrupted had in fact prospered, and questioned whether the theory could predict rather than merely redescribe. A. A. King and Baatartogtokh (2015) asked experts to assess cases cited as disruptions in Christensen's books against the theory's core conditions, such as incumbents overshooting customer needs and having the capability to respond but failing to do so, and found that only a small minority of cases met all of them. Christensen, Raynor, and McDonald (2015) replied in part by narrowing the concept: disruption refers to a process that starts from a low-end or new-market foothold, and many successful entrants (they cite Uber in its home market) are not disruptive in this sense. The exchange illustrates a general lesson about theory from Chapter 2: a concept stretched to cover every case loses its explanatory power. Scholars who use the theory should measure its conditions rather than infer them from outcomes.

12.2.4 Applications in business

BUSINESS CASE

Netflix and Blockbuster. Christensen, Raynor, and McDonald (2015) use Netflix as an example of disruption done right. Its DVD-by-mail service, launched in the late 1990s, appealed to a small set of customers, among them early adopters of DVD players and online shoppers, but did not suit the typical Blockbuster customer who wanted a new release on impulse. Blockbuster had little reason to respond. Netflix's move to streaming then improved the offering until it served the mainstream, and Blockbuster filed for bankruptcy in 2010. The case is analysed from the perspective of capability renewal in Chapter 5 (see Dynamic capabilities (Section 5.1)). Read through disruption theory, the instructive feature is that Netflix's initial *inferiority* on immediacy was what kept the incumbent from fighting back.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Incumbents will be slower to enter a new product segment when the segment's initial customers overlap little with their existing customer base.
- *H2.* The hazard of incumbent exit after the entry of a low-cost technology will be higher when the incumbent's products exceed the performance demanded by the median customer.

Designs. Industry histories that code product performance and customer requirements over time. Event-history models of incumbent entry into new segments. Pre-registered *ex ante* classification of innovations as sustaining or disruptive, followed by prospective tracking.
Measures. Product performance on key attributes, customer-requirement surveys, overlap between new-segment and core customers, internal resource allocation to new versus core projects.

COMMON PITFALL

Calling every successful entrant disruptive. Disruption is a specific process with stated conditions: an inferior foothold product, a faster-improving trajectory, and an incumbent with rational reasons not to respond. Labelling a case "disruptive" because the incumbent lost is circular. Classify innovations before outcomes are known and report the cases that did not follow the predicted path.

12.3 Profiting from Innovation

THEORY CARD

Core claim. Innovators often fail to profit from their innovations. Whether the innovator, imitators, or the owners of complementary assets capture the value depends on the **appropriability regime**, the stage of the **dominant design**, and access to **complementary assets**.
Origins. Strategy and industrial organization. Teece (1986), revisited in Teece (2006).
Level of analysis. Firm (innovator, imitator, asset owner) within an industry.

Key constructs. Appropriability regime (tight or weak), legal protection, tacit versus codified knowledge, pre-paradigmatic and paradigmatic phases, complementary assets (generic, specialized, cospecialized), contracting versus integration.
Boundary conditions. Most relevant for product and process innovations that require downstream manufacturing, distribution, or service to reach customers. Less relevant for innovations sold as pure intellectual property in efficient markets for technology.
Related theories. Resource-based theory (Section 4.1), Transaction cost economics (Section 6.3), First-mover advantage (Section 11.5), Network effects (Section 17.6).

12.3.1 Origins and intellectual history

Teece (1986) began from a set of puzzling outcomes. EMI invented the CT scanner and lost the market. RC Cola introduced diet cola before Coca-Cola and Pepsi, yet the large brands dominated the category. Other pioneers, such as Pilkington with float glass, did capture the value of their inventions. Teece argued that these outcomes are not accidents of management but reflect three structural features of the innovation's environment. The article became one of the most cited in strategic management and laid foundations for later work on ecosystems, platforms, and dynamic capabilities (Section 5.1) (Teece 2006).

12.3.2 Core logic and propositions

The first building block is the **appropriability regime**: the extent to which the innovator can prevent imitation. It depends on legal instruments (patents, copyright, trade secrets) and on the nature of the knowledge. Tacit knowledge that is hard to articulate is easier to protect than codified knowledge that can be read from a product or a patent. Regimes are **tight** when imitation is difficult and **weak** when it is easy. Teece stressed that tight regimes are the exception: in most industries patents are easy to invent around.

The second is the **dominant design paradigm**. Early in an industry's life (the pre-paradigmatic phase), many designs compete. Once a dominant design emerges, competition shifts to price and scale. An innovator with a weak regime may find that its early design is not the one that wins (see Dominant design).

The third is **complementary assets**: the capabilities and assets needed to commercialize an innovation, such as manufacturing, distribution, service, brand, and complementary technologies. **Generic** assets are available on competitive terms. **Specialized** assets have a one-way dependence with the innovation. **Cospecialized** assets have a mutual dependence. When the regime is weak and the needed complementary assets are specialized and scarce, the owners of those assets are in a strong bargaining position and capture much of the value.

- **P1 (Tight regimes favour innovators).** Under a tight appropriability regime, the innovator is likely to capture a large share of the value it creates, whether by producing or by licensing.
- **P2 (Weak regimes and timing).** Under a weak regime, the innovator's position deteriorates once a dominant design emerges and competition shifts to complementary assets.
- **P3 (Asset ownership).** When complementary assets are specialized and scarce, owners of those assets (incumbents or fast followers) tend to capture value at the innovator's expense.
- **P4 (Integration choice).** An innovator should integrate into specialized complementary assets when the regime is weak, the assets are critical, and the innovator can afford to build or buy them in time, and should contract for them otherwise.

12.3.3 Mechanism

Figure 12.3 expresses the argument as a decision logic.

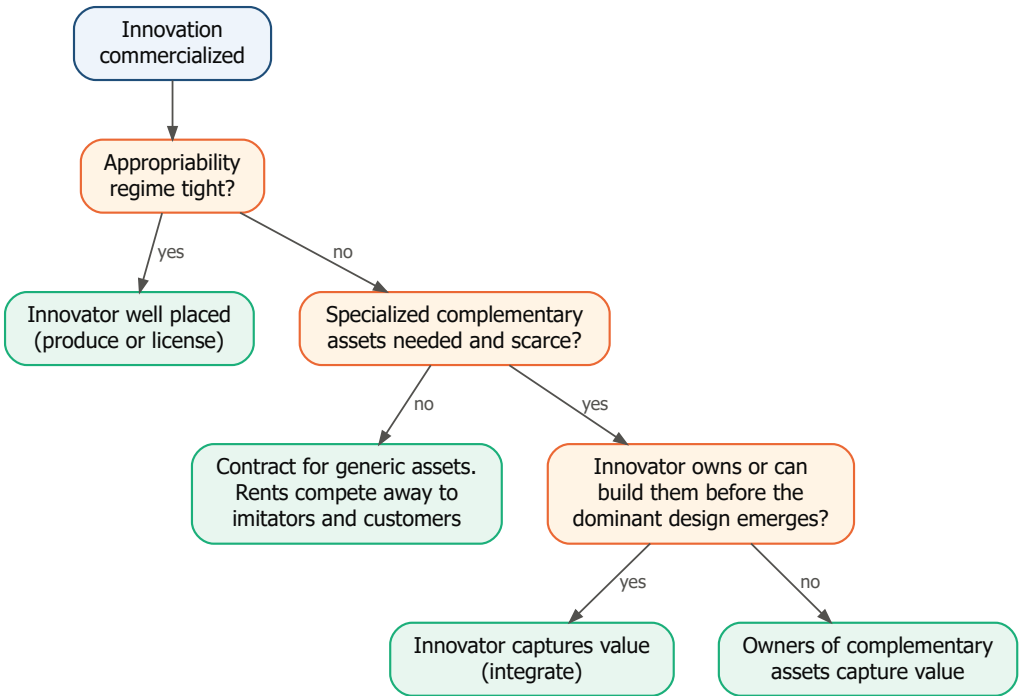


Figure 12.3. The logic of profiting from innovation. Tight appropriability favours the innovator. Under weak appropriability, value flows to whoever controls specialized complementary assets once a dominant design has emerged. Adapted from the argument in Teece (1986).

Table 12.3 summarizes who tends to win under each configuration, with examples Teece discussed.

Table 12.3. Who profits from innovation? Configurations in Teece's framework.

Configuration	Likely winner	Reason	Example in Teece (1986)
Tight regime, any asset position	Innovator	Imitation is blocked, so the innovator can produce or license on favourable terms	Pilkington float glass
Weak regime, generic complementary assets	Customers and many imitators	Anyone can copy and contract for the assets, so rents compete away	Many consumer products after a design stabilizes
Weak regime, specialized assets held by innovator	Innovator	Imitators cannot reach customers without assets the innovator controls	Integrated firms that own manufacturing and distribution
Weak regime, specialized assets held by others	Imitators or asset owners	Owners of manufacturing, brand, distribution, or service extract the value	EMI and the CT scanner, RC Cola and diet cola
Pre-paradigmatic phase, weak regime	Uncertain	The innovator's design may not become dominant, and positioning for the eventual design matters most	Early entrants in many new product categories

12.3.4 Evidence and extensions

Tripsas (1997) provided some of the most careful evidence. Across a century of the typesetter industry, incumbents that held specialized complementary assets survived waves of competence-destroying technology that would otherwise have removed them. Gans and Stern (2003) extended the framework to entrepreneurial strategy: a start-up should cooperate with incumbents (for instance by licensing or being acquired) when its intellectual property is strong and incumbents control the needed complementary assets, and compete in the product market when neither holds. In his reflections, Teece (2006) noted that the framework anticipated later work on business ecosystems and platforms, where control of a few critical complements, such as an operating system or an app store, can determine who captures value from many other firms' innovations.

DEBATE AND EVIDENCE

How weak are weak regimes? Teece's emphasis on complementary assets assumes that patents are usually weak. In pharmaceuticals and specialty chemicals, patents are strong and innovators often capture value through licensing. In software and digital services, by contrast, network effects and data can act as complementary assets that are themselves created by the innovation, blurring the line between the innovation and its complements. Critics also note that the framework is easier to apply in hindsight than in prospect, because which assets will become specialized depends on which design wins.

12.3.5 Applications in business

BUSINESS CASE

EMI and the CT scanner. EMI's scanner was a genuine breakthrough, but selling and supporting sophisticated equipment to hospitals required a sales force, service network, and relationships with radiologists, especially in the United States, that established medical-electronics firms possessed and EMI did not. The underlying ideas could be engineered around. Firms with those complementary assets, including General Electric, entered quickly, and EMI exited the scanner business within a decade of launch (Teece 1986). The lesson is not that EMI should not have innovated but that its commercialization strategy needed either a partner that owned the complementary assets or an early, deliberate investment in building them.

BUSINESS CASE

The iPhone and complementary assets. Apple's 2007 iPhone was neither the first smartphone nor the first touchscreen phone, and many of its features were quickly imitated. Yet Apple has captured a share of industry profits well above its share of unit sales. Teece's framework explains why. Apple held a portfolio of complementary assets that rivals could not easily replicate: a premium brand, its own retail stores (opened from 2001), the iTunes content and payment infrastructure, deep supply-chain relationships, and, from 2008, the App Store, which turned third-party developers into cospecialized complements. As a dominant design for touchscreen smartphones emerged, competition shifted to exactly these assets. Apple was a follower in the product category but an innovator in assembling complements around it.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The positive effect of being first to market on long-run market share will be weaker when patent protection in the industry is weak.
- *H2.* Technology start-ups will be more likely to license or be acquired than to enter the product market when incumbents control specialized manufacturing or distribution assets.

Designs. Industry histories with event-history models of survival, comparisons across industries that differ in patent effectiveness, and commercialization-mode choice models for start-ups.

Measures. Survey measures of appropriability (patent effectiveness, secrecy, lead time), ownership of downstream assets (sales force, service network, manufacturing plants), licensing and acquisition events.

COMMON PITFALL

Equating invention with value capture. Many business plans assume that a patent or a technical lead guarantees profits. Teece's framework implies that the right question is "Who controls the assets needed to bring this to customers, and what happens after the design stabilizes?" Map the complementary assets before choosing whether to build, partner, or license.

12.4 Architectural Innovation

THEORY CARD

Core claim. Innovations that reconfigure how existing components are linked together, while leaving the components largely unchanged, can destroy the value of incumbents' **architectural knowledge** and are especially hard for incumbents to recognize and exploit.

Origins. Technology management. R. M. Henderson and Clark (1990), based on the photolithographic alignment equipment industry.

Level of analysis. Product and firm, with the organization's communication channels and information filters as the micro-mechanism.

Key constructs. Component knowledge, architectural knowledge, core design concepts, linkages, incremental, modular, architectural, and radical innovation, communication channels, information filters, problem-solving strategies.

Boundary conditions. Applies to complex products made of interacting components, developed by organizations whose structure mirrors the product's architecture.

Related theories. Disruptive innovation, Absorptive capacity (Section 5.2), Knowledge-based view (Section 4.3), Dominant design.

12.4.1 Core logic

Earlier research had classified innovations as **incremental** or **radical**. R. M. Henderson and Clark (1990) observed that this dichotomy could not explain why incumbents sometimes failed in the face of seemingly modest innovations. They argued that designing a product requires two kinds of knowledge: **component knowledge** about each core concept and how it is implemented, and

architectural knowledge about how components are integrated into a coherent whole. Crossing these dimensions yields four types of innovation, shown in Figure 12.4.

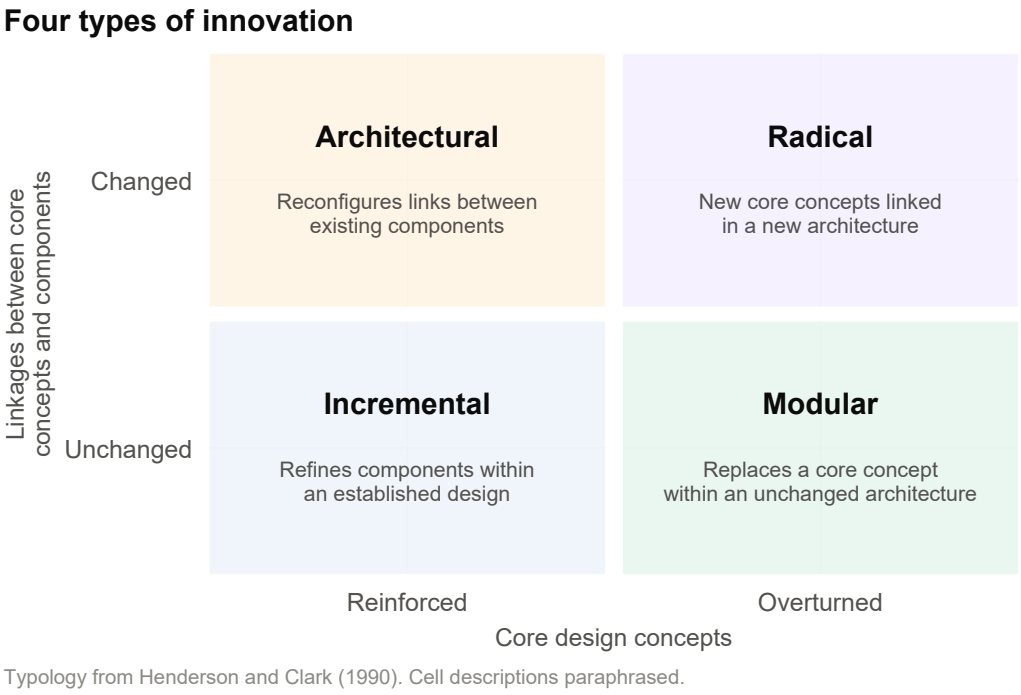


Figure 12.4. Henderson and Clark's (1990) typology of innovation. Architectural innovation leaves core concepts intact but changes how components are linked, which is why incumbents often misread it as incremental.

Their own example is the room fan. Improving blade design or motor power within a standard ceiling fan is incremental. Replacing the motor with a fundamentally different one is modular. Designing a portable fan, which uses largely the same components but arranges them in a different configuration with different interactions among blade size, motor power, and housing, is architectural.

The difficulty for incumbents is that architectural knowledge becomes embedded in the organization. Teams are structured around components, **communication channels** link the people whose components interact, **information filters** highlight the variables that matter in the existing design, and **problem-solving strategies** become routine. Because these structures were built for the old architecture, they screen out precisely the information that signals a new one. An incumbent may therefore treat an architectural innovation as incremental, apply its familiar routines, and fail.

- **P1.** Incumbents are more likely to fail in architectural and radical innovations than in incremental and modular ones.
- **P2.** The more an organization's communication structure mirrors the existing product architecture, the harder it is to recognize architectural change.
- **P3.** Architectural innovation is more damaging than its technical magnitude suggests because it is often misclassified as incremental (Table 12.4).

Table 12.4. Knowledge effects of the four innovation types (after Henderson & Clark, 1990).

Type	Component knowledge	Architectural knowledge	Typical incumbent outcome	Illustrative example
Incremental	Enhanced	Enhanced	Incumbent advantage	Successive improvements in a standard product
Modular	Destroyed	Enhanced	Incumbents can adapt by acquiring new component skills	Replacing an analogue component with a digital one
Architectural	Largely enhanced	Destroyed	Often fail, because the change is misread	Room fan to portable fan, contact to proximity aligners
Radical	Destroyed	Destroyed	Usually fail, but the threat is recognized	Shift to an entirely new technology base

12.4.2 Evidence and extensions

In the photolithographic alignment equipment industry, a series of innovations that appeared modest in technical terms (for example, the move from contact to proximity aligners) was accompanied by changes in industry leadership, with established firms repeatedly losing to entrants (R. M. Henderson and Clark 1990, R. Henderson 1993). The concept has since been used to explain organizational design: firms that separate component development into modules may innovate quickly within the architecture but struggle to change it. It links naturally to absorptive capacity (Section 5.2) (W. M. Cohen and Levinthal 1990), because recognizing architectural change requires knowledge outside the current filters.

DEBATE AND EVIDENCE

Architecture or cognition? Henderson and Clark locate the problem in organizational structure and routines. Later work emphasized managerial cognition. In their study of Polaroid's move into digital imaging, Tripsas and Gavetti (2000) found that the firm built strong technical capabilities but that managers' beliefs about the business model held it back. The two explanations are complementary but imply different remedies: restructuring communication channels versus changing managers' mental models.

BUSINESS CASE

Kodak and the architecture of photography. Digital imaging was radical for Kodak's chemistry, but it was also architectural for the photography business. The value chain shifted from film, processing, and prints to sensors, software, storage, and screens, and later to sharing through smartphones and social media. Kodak built an early digital camera in 1975 and invested in digital products, yet its organization, profit model, and information filters remained centred on film and photofinishing. H. C. Lucas and Goh (2009) attribute the failure less to ignorance of the technology than to an organization and culture unable to reconfigure around it. The company filed for Chapter 11 protection in 2012. The capability side of the story is covered in Chapter 5 under Dynamic capabilities (Section 5.1). The architectural lens adds that the hardest part to change was not knowledge of components but knowledge of how the business fitted together.

12.5 Dominant Design and Technology Cycles

THEORY CARD

Core claim. New industries begin with a period of experimentation in which many designs compete. A **dominant design** eventually emerges, after which the focus of innovation shifts from product to process, the number of firms falls, and competition turns to cost and scale until a new technological discontinuity restarts the cycle.
Origins. Technology management. Utterback and Abernathy (1975) and Abernathy and Utterback (1978) on the dynamics of product and process innovation. Tushman and Anderson (1986) and P. Anderson and Tushman (1990) on technological discontinuities and technology cycles. Synthesized in Utterback (1994).
Level of analysis. Industry and technology, with consequences for firm entry, exit, and survival.
Key constructs. Fluid, transitional, and specific phases, product versus process innovation, dominant design, era of ferment, era of incremental change, competence-enhancing versus competence-destroying discontinuities.
Boundary conditions. Fits assembled products with many interacting components better than non-assembled goods or services. The concept of a single dominant design is harder to apply in software and platform markets with continual updates.
Related theories. Product life cycle (Section 28.6), Network effects (Section 17.6), Organizational ecology (Section 7.5), Profiting from innovation.

12.5.1 Origins and intellectual history

Utterback and Abernathy (1975) proposed that the rate and type of innovation in a productive unit change as the unit matures. Early on, when performance requirements are unclear, firms experiment with product features, and production processes are flexible and inefficient. As the market learns what it wants, a design stabilizes, and innovation shifts to the production process, which becomes more specialized, capital-intensive, and efficient. Abernathy and Utterback (1978) named the three stages the **fluid**, **transitional**, and **specific** phases and illustrated them with the automobile industry, in which Ford's Model T became a dominant design around which mass production was organized.

12.5.2 Core logic and propositions

A **dominant design** is the configuration that wins allegiance in the marketplace and that competitors and innovators must adhere to if they hope to command significant market share. It is rarely the most advanced technology. It usually combines features pioneered by several firms, and it can be shaped by network effects, complementary assets, regulation, and alliances as much as by technical merit (P. Anderson and Tushman 1990). Examples often given include the QWERTY typewriter keyboard, the Model T, the VHS video format, and the IBM-compatible personal computer.

P. Anderson and Tushman (1990) embedded the concept in a **technology cycle**. A technological discontinuity opens an **era of ferment** marked by competition among variants of the new technology and between the new and old technologies. The selection of a dominant design ends the era of ferment and opens an **era of incremental change**, which lasts until the next discontinuity. Tushman and Anderson (1986) added that discontinuities differ in their effect on incumbents. **Competence-**

enhancing discontinuities build on existing know-how and tend to be introduced by incumbents. **Competence-destroying** discontinuities render existing know-how obsolete and tend to be introduced by new entrants.

- **P1 (Phase shift).** The rate of product innovation peaks before the emergence of a dominant design, and the rate of process innovation peaks after it.
- **P2 (Shakeout).** The number of firms in an industry rises during the era of ferment and falls sharply after a dominant design emerges.
- **P3 (Competence effects).** Competence-destroying discontinuities increase the share of new entrants among industry leaders, and competence-enhancing discontinuities consolidate incumbents' positions.
- **P4 (Selection is social).** Which design becomes dominant cannot be predicted from technical merit alone.

Figure 12.5 shows the stylised pattern behind P1.

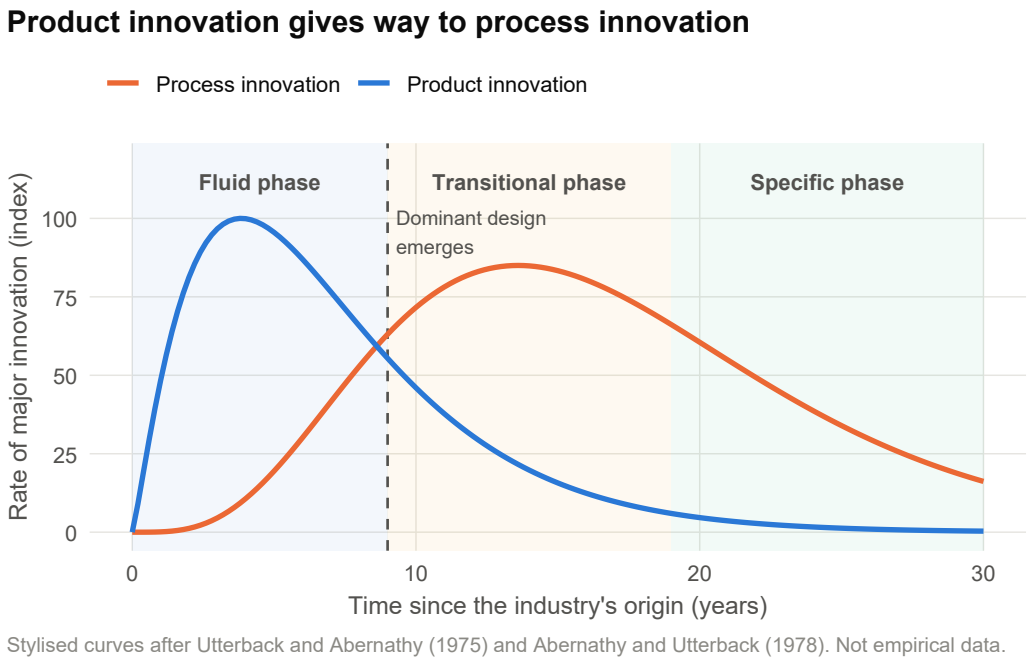


Figure 12.5. Stylised Abernathy-Utterback dynamics. Product innovation dominates the fluid phase. After a dominant design emerges, the rate of process innovation peaks and both rates eventually decline. Illustrative curves only.

12.5.3 Evidence and extensions

The empirical record broadly supports the cycle. P. Anderson and Tushman (1990) found the predicted sequence of discontinuity, ferment, and dominant design in the cement, glass, and minicomputer industries, and reported that the dominant design was seldom the technological frontier at the time of the discontinuity. Suárez and Utterback (1995) showed that the emergence of a dominant design shapes firm survival: across several industries, firms that entered before a dominant design emerged had better survival prospects than those that entered after. Klepper (1996) offered an economic explanation for shakeouts in which larger firms gain more from process innovation

because they spread its cost over more output, so that early entrants that grow large come to dominate.

DEBATE AND EVIDENCE

Is the dominant design a cause or a label? Critics point out that a dominant design is often identified only after the shakeout it is supposed to explain, which risks circularity. Others argue that in modular, software-based products, architectures evolve continuously and the cycle compresses or blurs. The competence-destroying versus competence-enhancing distinction has also been challenged: whether a technology destroys a firm's competence depends on the particular firm, as the typesetter evidence on complementary assets shows (Tripsas 1997). The most defensible use of the framework is as a set of conditional expectations about where an industry is in its cycle, not as a deterministic timetable.

Table 12.5 presents the competence-enhancing versus competence-destroying discontinuities (after Tushman & Anderson, 1986).

Table 12.5. Competence-enhancing versus competence-destroying discontinuities (after Tushman & Anderson, 1986).

Feature	Competence-enhancing discontinuity	Competence-destroying discontinuity
Effect on know-how	Builds on existing skills and knowledge	Makes existing skills and knowledge obsolete
Typical initiator	Incumbent firms	New entrants
Effect on industry structure	Reinforces incumbents' positions	Opens the door to new leaders
Managerial challenge	Keep investing in the current trajectory	Recognize the threat and build or buy new competences
Link to other theories	Profiting from innovation: incumbents' assets retain value	Creative destruction, Liability of newness for entrants

BUSINESS CASE

Format battles and the timing of commitment. In consumer electronics, the contest between VHS and Betamax in home video is a frequently cited case in which the eventual dominant design was not universally judged the technically superior one. Its victory owed much to licensing, recording time, and the availability of pre-recorded content: network effects and complementary assets rather than picture quality. For managers, the cycle has two practical implications. During the era of ferment, invest in flexibility and alliances that improve the odds of being on the winning design. Once a design is dominant, shift resources to process efficiency, cost, and the complementary assets that will determine profits.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* After a dominant design emerges, the ratio of process to product patents in an industry will increase.
- *H2.* Firms entering an industry after the emergence of a dominant design will face a higher hazard of exit than firms that entered before it.

Designs. Long industry histories coded from trade directories, product catalogues, and patent

data. Event-history models of entry and exit. Content analysis of product specifications to date dominant designs independently of outcomes.
Measures. Product-feature convergence across models, process versus product patent classes, number of producers over time, entry cohorts.

12.6 Open Innovation

THEORY CARD

Core claim. Firms can and should use purposive inflows and outflows of knowledge, bringing in external ideas and taking internal ideas to market through external paths, rather than relying solely on internal research and development.
Origins. Technology management. Chesbrough (2003), building on earlier work on absorptive capacity and user innovation. Laursen and Salter (2006) on the returns to external search. Dahlander and Gann (2010) on the forms of openness.
Level of analysis. Firm, and its relations with universities, suppliers, customers, start-ups, and communities.
Key constructs. Closed versus open innovation, inbound and outbound openness, search breadth and depth, pecuniary and non-pecuniary exchange, intellectual-property management, absorptive capacity.
Boundary conditions. Valuable where useful knowledge is widely distributed and markets for technology function. Less valuable where knowledge is highly tacit, integration costs are high, or appropriability concerns dominate.
Related theories. Absorptive capacity (Section 5.2), Structural holes (Section 9.4), Social capital (Section 9.5), Profiting from innovation.

12.6.1 Core logic

Chesbrough (2003) contrasted two paradigms. Under **closed innovation**, a firm generates its own ideas, develops them in its own laboratories, and brings them to market through its own channels. Under **open innovation**, the boundary of the firm is porous: valuable ideas can come from outside, and internal ideas that do not fit the current business can reach the market through licensing, spin-offs, or partnerships. Chesbrough argued that several erosion factors had undermined the closed model, notably the growing mobility of skilled workers, the availability of venture capital for spin-offs, and the rising capability of external suppliers and universities. His cautionary example was Xerox's Palo Alto Research Center, which generated influential technologies whose value was largely captured by other firms (Table 12.6).

Table 12.6. Closed versus open innovation (principles paraphrased from Chesbrough, 2003).

Principle	Closed innovation	Open innovation
Source of talent	The best people in the field work for us	Not all the best people work for us, so we must work with those outside
Profiting from R&D	We must discover, develop, and ship it ourselves	External R&D can create value, and internal R&D captures part of it
Speed to market	Being first to discover wins	Building a better business model beats being first to discover

Table 12.6. Closed versus open innovation (principles paraphrased from Chesbrough, 2003). *(continued)*

Principle	Closed innovation	Open innovation
Intellectual property	Control it so competitors cannot benefit	Profit from others' use of our IP and buy others' IP when it advances our model
Unused ideas	Shelved	Licensed, spun off, or contributed to partners

Dahlander and Gann (2010) clarified that openness comes in four forms by crossing direction with exchange type. **Inbound** openness includes *sourcing* external knowledge without payment (for example from users or scientific publications) and *acquiring* it through licences or purchases. **Outbound** openness includes *revealing* internal knowledge without immediate payment and *selling* it through licences or spin-offs. Each form has distinct costs and benefits.

12.6.2 Evidence and extensions

Laursen and Salter (2006) studied how the **breadth** (number of external sources used) and **depth** (intensity of use) of external search relate to innovation performance among UK manufacturing firms. Both were positively associated with the share of sales from new products, but with diminishing and eventually negative returns: an **inverted-U**. The explanation is attention. Each additional external relationship demands managerial time to identify, absorb, and integrate knowledge, and beyond some point the costs exceed the benefits. Figure 12.6 shows the shape with simulated data.

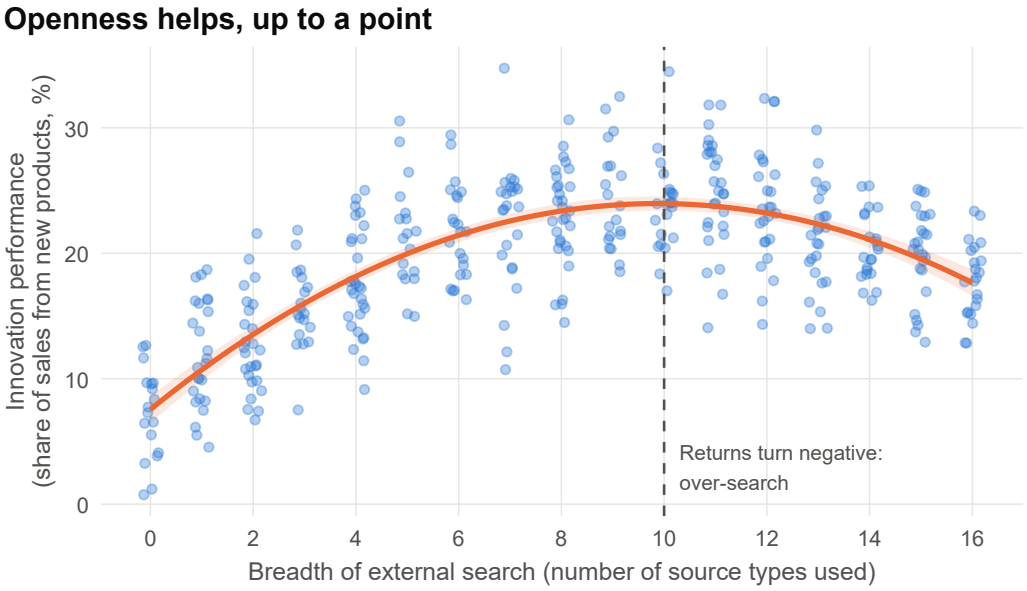


Figure 12.6. Illustrative simulation of the inverted-U relationship between breadth of external search and innovation performance reported by Laursen and Salter (2006). Points are simulated firms, the line is a quadratic fit.

Open innovation intersects with absorptive capacity (Section 5.2): firms need internal research to recognize and assimilate outside knowledge (W. M. Cohen and Levinthal 1990). It also connects to network theories, since external search draws on brokers and bridging ties (see Structural holes

(Section 9.4)).

DEBATE AND EVIDENCE

Is open innovation new? Critics argue that firms have long sourced ideas externally through acquisitions, licensing, and collaboration, and that open innovation relabels familiar practices. Defenders respond that the contribution lies in treating openness as a deliberate strategy with its own business model, and in measuring its forms and limits. A second concern is appropriability: revealing knowledge can help rivals as much as partners, which ties the theory back to Profiting from innovation.

12.6.3 Applications in business

BUSINESS CASE

Procter & Gamble's Connect and Develop. In the early 2000s, P&G's leadership concluded that internal research alone could not deliver the growth the company needed. It set a goal of sourcing half of its innovations from outside the company and created a programme, Connect and Develop, that used technology scouts, supplier networks, and online intermediaries to find external solutions to defined problems. Huston and Sakkab (2006), two P&G executives, describe how the firm specified needs in short technology briefs and searched its networks for existing solutions. One example they report is Pringles crisps printed with pictures and words, which used an ink-jet technique developed by a small bakery in Italy rather than one invented in-house. The programme illustrates inbound, pecuniary openness and the importance of internal capability to evaluate what is found outside.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The relationship between the number of external knowledge sources and new-product sales will be inverted-U shaped.
- *H2.* The optimal breadth of search will be higher for firms with greater internal R&D intensity (absorptive capacity).

Designs. Innovation surveys (such as the European Community Innovation Survey) with breadth and depth indices, panel studies of licensing and alliance portfolios, field studies of crowdsourcing contests.

Measures. Count of external source types used, intensity of use, in- and out-licensing revenues, share of new products with external origins.

COMMON PITFALL

More openness is always better. Openness has costs in attention, coordination, and appropriability. Firms that multiply partnerships without the capacity to absorb them can see innovation performance fall. Treat openness as a portfolio with an optimum, not a slogan.

12.7 Effectuation

THEORY CARD

Core claim. Under uncertainty, expert entrepreneurs often reason **effectually**: they start from the means they control and let goals emerge through commitments with stakeholders, rather than reasoning **causally** from a predetermined goal to the means needed to reach it.
Origins. Entrepreneurship and cognitive science. Sarasvathy (2001), drawing on think-aloud protocols with expert entrepreneurs and on Herbert Simon's work on decision making. Elaborated in Sarasvathy (2008).
Level of analysis. Individual entrepreneur and the venture's stakeholder network.
Key constructs. Causation, effectuation, means (who I am, what I know, whom I know), affordable loss, partnerships and pre-commitments, leveraging contingencies, non-predictive control.
Boundary conditions. Most applicable when the future is unpredictable, goals are ambiguous, and the environment is shaped by the actions of the participants (Knightian uncertainty). Causal reasoning is more effective when the market exists and outcomes are predictable.
Related theories. Bounded rationality (Section 23.1), Real options theory (Section 11.6), Opportunity discovery and creation, Bricolage.

12.7.1 Origins and core logic

Sarasvathy (2001) illustrated the difference with a chef. A chef reasoning causally is given a menu and buys the ingredients needed to cook it. A chef reasoning effectually looks in the cupboards and asks what meals can be made with what is there. Both are rational, but they suit different conditions. Causation is a logic of **prediction**: to the extent we can predict the future, we can control it. Effectuation is a logic of **control**: to the extent we can control the future, we do not need to predict it.

Sarasvathy's later work organized effectuation into a set of principles that are now standard in entrepreneurship teaching (Sarasvathy 2008) (Table 12.7).

Table 12.7. Principles of effectuation contrasted with causal reasoning (labels from Sarasvathy, 2008).

Principle	Effectual logic	Causal contrast	Business illustration
Bird in hand (means)	Start with who you are, what you know, and whom you know	Start with a goal and acquire the means to reach it	A founder builds a product for the industry she has worked in for a decade
Affordable loss	Commit what you can afford to lose	Pursue the option with the highest expected return	Testing a concept with a weekend pop-up rather than a leased store
Crazy quilt (partnerships)	Bring in self-selected stakeholders who make commitments and shape goals	Analyse competitors and position against them	An early customer co-funds development in exchange for exclusivity
Lemonade (contingencies)	Treat surprises as resources to exploit	Avoid deviations from the plan	A failed feature reveals a different, more valuable use
Pilot in the plane (control)	Shape the future through action rather than forecast it	Forecast the market and plan for it	Creating an industry standard with partners instead of predicting one

12.7.2 Mechanism

Effectuation describes a process rather than a single decision. Means suggest possible courses of action. Interactions with people yield commitments, which expand the venture's means and constrain its goals. The cycle repeats until the emerging artifact, a firm, product, or market, is stable enough to manage causally (Figure 12.7).

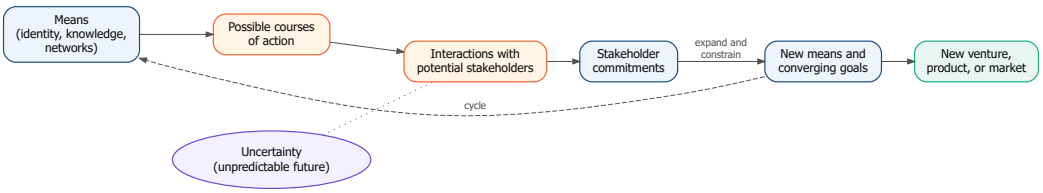


Figure 12.7. The effectual cycle. Means suggest actions, interactions produce stakeholder commitments, and commitments expand means while converging goals, until a new market or venture takes shape. Adapted from the dynamic model in Sarasvathy (2008).

12.7.3 Evidence and extensions

In a meta-analysis, S. Read, Song, and Smit (2009) found positive associations between new-venture performance and several effectual elements, especially the use of existing means and partnerships, though the evidence was uneven across principles and many studies relied on cross-sectional, self-reported measures. Effectuation has become one of the most influential frameworks in entrepreneurship education, and its emphasis on cheap experiments and stakeholder commitments overlaps with the practitioner movement known as the **lean startup**.

DEBATE AND EVIDENCE

Is effectuation a theory? Arend, Sarooghi, and Burkemper (2015) evaluated effectuation against criteria for theory and concluded that it falls short: its constructs are loosely defined, its predictions about performance are unclear, and it is hard to falsify. Proponents responded that effectuation is a theory of expertise and process, not a performance theory, and that its value lies in describing how entrepreneurs act under true uncertainty. The exchange is a useful case study in the criteria discussed in Chapter 2: precision, falsifiability, and scope. A further issue is causal direction. Experienced entrepreneurs may use effectual logic because they have succeeded rather than succeed because they use it.

12.7.4 Applications in business

BUSINESS CASE

Lean startups as structured experimentation. The lean startup method popularized by Ries (2011) and Blank (2013) advises founders to treat business ideas as hypotheses, build a **minimum viable product**, measure customer responses, and **pivot** or persevere on the basis of what they learn. The method shares effectuation's distrust of long-range forecasts and its emphasis on affordable experiments, but it is more causal in spirit: it starts from a hypothesized customer problem and tests it. Evidence is emerging. In a randomized field experiment with Italian start-ups, Camuffo et al. (2020) found that founders trained to formulate and test

hypotheses scientifically were more likely to pivot or abandon weak ideas and performed better than a control group. For managers in established firms, the same logic underlies internal venture programmes that fund small experiments with stage gates tied to evidence.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The positive relationship between effectual decision making and venture survival will be stronger in nascent markets than in established markets.
- *H2.* Ventures that secure pre-commitments from stakeholders before product launch will experience fewer and smaller losses on failure.

Designs. Think-aloud protocols and decision experiments comparing expert and novice entrepreneurs, longitudinal panels of nascent ventures, randomized training interventions.
Measures. Validated scales for causation and effectuation, counts of pre-commitments and partnerships, maximum funds at risk relative to founder wealth, number of pivots.

COMMON PITFALL

Treating effectuation as "no planning". Effectuation is not improvisation without discipline. It specifies a logic: bounded downside, stakeholder commitments, and the deliberate use of surprises. Studies that equate it with the absence of a business plan measure something else.

12.8 Opportunity Discovery and Creation

THEORY CARD

Core claim. Entrepreneurship is the study of how opportunities to create future goods and services are discovered, evaluated, and exploited. Theories disagree on whether opportunities exist objectively, waiting to be **discovered** by alert individuals, or are **created** through entrepreneurs' own actions.
Origins. Austrian economics and entrepreneurship research. Kirzner (1973) on entrepreneurial alertness. Shane and Venkataraman (2000) on the individual-opportunity nexus. Alvarez and Barney (2007) on discovery versus creation.
Level of analysis. Individual entrepreneur and market.
Key constructs. Entrepreneurial opportunity, alertness, prior knowledge, information asymmetry (Section 8.1), risk versus uncertainty, enactment.
Boundary conditions. Discovery logic fits markets disrupted by exogenous shocks with knowable outcomes. Creation logic fits settings where no market yet exists and outcomes cannot be known in advance.
Related theories. Effectuation, Information asymmetry (Section 8.1), Creative destruction, Sensemaking (Section 14.11).

12.8.1 Core logic

Kirzner (1973) offered a contrast to Schumpeter. Where Schumpeter's entrepreneur disrupts equilibrium with new combinations, Kirzner's entrepreneur moves markets *toward* equilibrium

by noticing price discrepancies and unexploited possibilities that others overlook. The capacity to notice is **alertness**, and it requires no special resources, only perception.

Shane and Venkataraman (2000) argued that entrepreneurship research had focused too much on who entrepreneurs are and too little on opportunities. They defined the field as the study of the **nexus** between enterprising individuals and valuable opportunities, and they posed three questions: why, when, and how opportunities come into existence, why, when, and how some people rather than others discover and exploit them, and why, when, and how different modes of action are used. In a companion study, Shane (2000) traced eight sets of entrepreneurs who founded firms to exploit a single MIT three-dimensional printing invention. Each saw a different application, in markets connected to their prior work, which suggests that **prior knowledge** shapes which opportunities people discover.

Alvarez and Barney (2007) sharpened the debate by setting out two internally consistent theories (Table 12.8).

Table 12.8. Discovery versus creation theories of entrepreneurial action (after Alvarez & Barney, 2007).

Assumption	Discovery theory	Creation theory
Nature of opportunities	Exist independently of entrepreneurs, formed by exogenous shocks to markets or technology	Do not exist until created through entrepreneurs' actions
Nature of entrepreneurs	Differ from others in alertness or prior knowledge	May not differ ex ante, differences may emerge through action
Decision context	Risky: outcomes and their probabilities can be estimated	Uncertain: outcomes cannot be known in advance
Implied actions	Search, evaluate, plan, and then exploit	Act, observe responses, and iterate
Associated approaches	Kirznerian alertness, information asymmetry	Effectuation, bricolage, enactment
Philosophical stance	Critical realism	Evolutionary realism and social construction

12.8.2 Evidence and extensions

The discovery view has generated a large literature on alertness, prior knowledge, and the cognitive and network sources of opportunity recognition. The creation view connects to effectuation, bricolage, and sensemaking. Recent work has questioned the opportunity construct itself. Ramoglou and Tsang (2016) propose that opportunities are best understood as **propensities**: latent conditions for profitable demand that exist whether or not anyone acts, but that are only actualized through entrepreneurial action. This realist position tries to keep the objectivity of discovery while accommodating the role of action stressed by creation theory.

DEBATE AND EVIDENCE

Can opportunities be observed before they are exploited? Critics argue that opportunities are identified only after successful exploitation, so "discovered opportunity" is often a retrospective label for success. This makes the construct hard to measure independently of outcomes. A practical response is to study *beliefs* about opportunities (perceived opportunities) and their formation, rather than opportunities as such, and to sample ventures that failed as well as those that succeeded.

BUSINESS CASE

Illustrative scenario: two founders, one technology. A university develops a low-cost sensor that detects moisture in materials. One founder, who previously worked in commercial roofing, sees a leak-detection product for flat roofs. A second founder, from agriculture, sees a soil-monitoring service for small farms. The discovery view explains the divergence through prior knowledge, as in Shane (2000). The creation view would add that each market takes shape only as the founders persuade early customers, adapt the product, and build distribution, so the eventual opportunity is partly a product of their actions. For a technology-transfer office, the implication is to license the same technology to several founders in different fields rather than search for the single best application.

12.9 Bricolage

THEORY CARD

Core claim. Resource-constrained entrepreneurs can create value by **making do** with resources at hand, recombining them for new purposes, and refusing to accept conventional limits on what those resources can do.
Origins. Anthropology and organization theory. Lévi-Strauss (1966) introduced the bricoleur. T. Baker and Nelson (2005) developed entrepreneurial bricolage from a field study of resource-constrained firms. Garud and Karnøe (2003) applied the idea to technology development.
Level of analysis. Venture and its local resource environment.
Key constructs. Resources at hand, making do, recombination, refusal to enact limitations, parallel versus selective bricolage, domains (physical inputs, labour, skills, customers, institutional rules).
Boundary conditions. Most relevant in resource-scarce environments. Persistent, wide-ranging bricolage may constrain growth.
Related theories. Effectuation, Resource-based theory (Section 4.1), Combinatorial creativity (Section 31.3).

12.9.1 Core logic

The anthropologist Claude Lévi-Strauss contrasted the **bricoleur**, who builds with a heterogeneous stock of materials at hand, with the engineer, who acquires materials designed for a project (Lévi-Strauss 1966). T. Baker and Nelson (2005) brought the concept into entrepreneurship. Their field study of small firms facing severe resource constraints found that some created something from nothing by treating undervalued or discarded resources (scrap materials, untrained but willing workers, ignored customers, ambiguous regulations) as raw material for new products and services. Resource environments, they argued, are partly **constructed** by the firms that act in them (Table 12.9).

Table 12.9. Domains of entrepreneurial bricolage (organized following Baker & Nelson, 2005).

Domain	What bricolage looks like	Example
Physical inputs	Using discarded, idle, or unconventional materials	Repurposing equipment or waste streams from another industry

Table 12.9. Domains of entrepreneurial bricolage (organized following Baker & Nelson, 2005). *(continued)*

Domain	What bricolage looks like	Example
Labour	Involving customers, friends, and suppliers as workers	Customers who help build or test the product
Skills	Applying amateur or self-taught skills to professional tasks	A founder writes software without formal training
Customers and markets	Serving customers that others ignore	Offering a basic service to buyers who were not served at all
Institutional environment	Testing ambiguous rules and standards rather than accepting them	Operating in a gap between regulatory categories

T. Baker and Nelson (2005) also distinguished two patterns. **Parallel bricolage**, applied habitually across many domains at once, tended to trap firms in a small, idiosyncratic niche, because customers, employees, and products all depended on improvisation. **Selective bricolage**, confined to a few domains and over a limited period, was associated with growth.

12.9.2 Evidence and extensions

Garud and Karnøe (2003) compared the emergence of wind-turbine industries in Denmark and the United States. The Danish path, a form of bricolage in which modest designs were improved incrementally through interaction among users, producers, and regulators, proved more successful than the American pursuit of technological breakthroughs. Senyard et al. (2014) found survey evidence that bricolage was positively associated with the innovativeness of resource-constrained new firms. The concept has been applied to social enterprises, frugal innovation in emerging markets, and crisis response.

BUSINESS CASE

Illustrative scenario: frugal clinic equipment. A start-up providing diagnostic services to rural clinics cannot afford new imaging equipment. Its founders refurbish units retired by urban hospitals, train local technicians whom larger providers do not recruit, and offer pay-per-scan pricing that clinics previously could not access. This is bricolage across physical inputs, labour, and customers. The theory's warning is that if the firm keeps improvising in every domain, it may never achieve the reliability required to win hospital contracts. A deliberate shift from parallel to selective bricolage, standardizing equipment while keeping its low-cost model, is the growth path.

12.10 Liability of Newness

THEORY CARD

Core claim. New organizations have higher failure rates than older ones because they must learn new roles, build trust among strangers, and establish ties with customers and suppliers, while lacking the routines and legitimacy (Section @ref(legitimacy (Section 7.3))) that come with age.

Origins. Sociology. Stinchcombe (1965) in the *Handbook of Organizations*. Formalized and tested in organizational ecology by J. Freeman, Carroll, and Hannan (1983), with alternatives

in Brüderl and Schüssler (1990) and Levinthal (1991).
Level of analysis. Organization within a population.
Key constructs. Age dependence, hazard of failure, role learning, trust, stable ties, reliability, accountability, initial endowments, liability of adolescence, liability of smallness.
Boundary conditions. Age effects can be confounded with size and initial resources. Very old organizations may face a liability of obsolescence.
Related theories. Organizational ecology (Section 7.5), Legitimacy (Section 7.3), Institutional theory (Section 7.2), Signaling theory (Section 8.2).

12.10.1 Core logic

Stinchcombe (1965) gave several reasons why new organizations are vulnerable. New roles must be learned, often by people who have not performed them before. Members of a new organization must develop trust and working relationships with strangers. Routines for coordination must be invented rather than inherited. And a new organization lacks stable ties with the customers, suppliers, and other parties on which it depends. Hannan and Freeman (1984) added that selection pressures favour organizations that perform **reliably** and can **account** rationally for their actions, properties that develop with age through the institutionalization of routines.

J. Freeman, Carroll, and Hannan (1983) tested the prediction that death rates decline with age using data on labour unions, newspaper publishers, and semiconductor manufacturers, and found the predicted negative age dependence. The liability of newness can also operate at the level of an entire new industry, whose firms lack both cognitive legitimacy (people do not understand what they do) and socio-political legitimacy (Aldrich and Fiol 1994).

- **P1 (Newness).** The hazard of organizational failure declines with age.
- **P2 (Endowments).** Initial resources such as capital, founders' experience, and commitments delay the period of highest risk.
- **P3 (Legitimacy).** New organizations in new forms or industries face higher failure rates than new organizations in established forms.

12.10.2 Mechanism

Brüderl and Schüssler (1990) proposed a **liability of adolescence**. Using data on German business registrations, they found that the hazard of failure first rose and then fell with age. Their explanation is that founders commit initial resources that buffer the venture for a time, and that stakeholders suspend judgement until performance information accumulates. Failure peaks when the buffer is exhausted. Levinthal (1991) showed that a simple model in which organizational capital follows a random walk, with failure when it falls below a threshold, generates both patterns depending on initial endowments. Figure 12.8 compares the shapes.

Table 12.10 presents the age-related liabilities and their explanations.

Table 12.10. Age-related liabilities and their explanations.

Liability	Predicted hazard pattern	Mechanism	Key source
Newness	Highest at founding, declining with age	Role learning, trust, routines, and external ties take time to build	Stinchcombe (1965), Freeman, Carroll & Hannan (1983)

Table 12.10. Age-related liabilities and their explanations. *(continued)*

Liability	Predicted hazard pattern	Mechanism	Key source
Adolescence	Low at founding, rising, then declining	Initial endowments and suspended judgement buffer young firms	Brüderl & Schüssler (1990)
Smallness	Higher for small organizations at any age	Limited capital, staff, and legitimacy	Freeman, Carroll & Hannan (1983)
Newness of the form	Higher for all firms in a new industry	Lack of cognitive and socio-political legitimacy	Aldrich & Fiol (1994)
Random walk	Depends on initial capital relative to the failure threshold	Stochastic accumulation of organizational capital	Levinthal (1991)

12.10.3 Evidence and extensions

Negative age dependence is among the most replicated findings in organizational ecology, but its interpretation has been contested. Because age and size are correlated, early studies that did not control for size may have attributed a **liability of smallness** to age. The adolescence pattern suggests that the riskiest period is not the founding moment but the point at which initial resources run out. Research on imprinting, also traced to Stinchcombe (1965), adds that conditions at founding leave durable marks on structure and strategy.

DEBATE AND EVIDENCE

Newness, smallness, or selection? Liability of newness may partly reflect heterogeneity rather than learning. If a population of new firms contains both strong and weak ventures, the weak fail early, and the average hazard among survivors falls even if no individual firm becomes less risky with age. Event-history models with controls for unobserved heterogeneity, and studies of within-firm changes in capabilities, are needed to separate these explanations.

BUSINESS CASE

Managing the newness gap. Practitioners respond to the liability of newness by borrowing age and legitimacy. Start-ups recruit experienced executives and board members, seek endorsements from reputable investors and customers, join accelerators, and adopt recognizable organizational templates. Each move substitutes for the track record they lack and links to signaling theory (Section 8.2). Corporate venturing units face a mirror problem: they can borrow the parent's legitimacy and resources but may inherit routines unsuited to the new business, which is why ambidexterity (Section 5.4) research recommends structural separation with senior integration.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The hazard of venture failure will first increase and then decrease with age, with the peak occurring later for ventures with larger initial funding.
- *H2.* Founding-team industry experience will reduce the hazard of failure more strongly in the first three years than thereafter.

Designs. Event-history (survival) analysis of founding cohorts from business registers, with

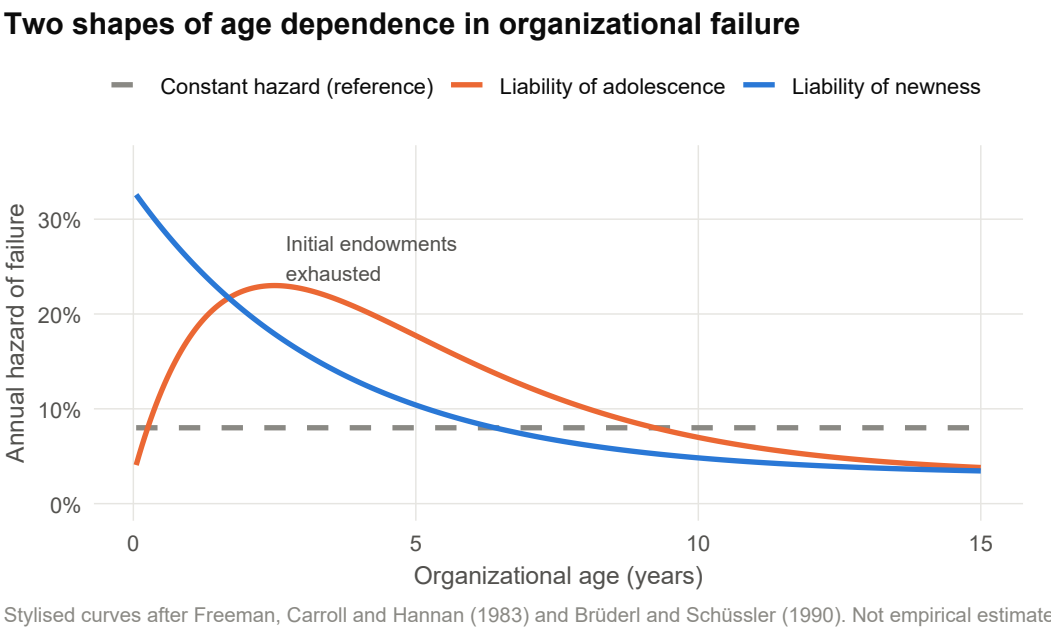


Figure 12.8. Stylised hazard rates of organizational failure by age. Liability of newness predicts a monotonic decline. Liability of adolescence predicts an initial rise while initial endowments buffer the venture, followed by a decline. Illustrative functional forms only.

time-varying covariates for size and funding. Piecewise-exponential or flexible hazard models to detect non-monotonic age dependence.
Measures. Founding and exit dates, initial capital, employment, founder experience, accelerator participation, investor reputation.

COMMON PITFALL

Confusing age with size or endowment. A negative coefficient on age in a survival model does not by itself demonstrate a liability of newness. Without controls for size, initial resources, and unobserved firm quality, the effect may reflect smallness, under-capitalization, or the early exit of weak ventures.

12.11 Chapter Summary

KEY TAKEAWAYS

- **Creative destruction** frames innovation as the engine of economic change: new combinations earn temporary rents and make old products and firms obsolete (Schumpeter 1942, Aghion and Howitt 1992).
- **Disruptive innovation** explains why well-managed incumbents lose to inferior entrants that start in footholds the incumbents rationally ignore, but its conditions must be measured, not inferred from outcomes (Christensen 1997, Christensen, Raynor, and McDonald 2015, A. A. King and Baatartogtokh 2015).

- **Profiting from innovation** shows that value capture depends on appropriability, the dominant design, and control of complementary assets (Teece 1986).
- **Architectural innovation** reveals that reconfiguring existing components can defeat incumbents because their organizations embed the old architecture (R. M. Henderson and Clark 1990).
- **Dominant designs and technology cycles** describe how industries move from ferment to standardization, shifting innovation from product to process (Abernathy and Utterback 1978, P. Anderson and Tushman 1990).
- **Open innovation** treats external knowledge as a strategic resource with diminishing returns (Chesbrough 2003, Laursen and Salter 2006).
- **Effectuation**, **opportunity creation**, and **bricolage** describe how entrepreneurs act under uncertainty and scarcity, starting from means at hand (Sarasvathy 2001, Alvarez and Barney 2007, T. Baker and Nelson 2005).
- The **liability of newness** reminds founders and investors that young organizations face elevated failure risk until routines, ties, and legitimacy develop (Stinchcombe 1965, J. Freeman, Carroll, and Hannan 1983).

Table 12.11 presents the comparison of the chapter's theories.

Table 12.11. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Typical business use
Creative destruction	Innovation rents attract imitators and replace incumbents	Waves of entry, exit, and leadership turnover	Scanning for category-killing innovations
Disruptive innovation	Incumbents' resource allocation favours core customers	Entrants from footholds overtake incumbents	Deciding whether and how to respond to low-end entrants
Profiting from innovation	Appropriability and complementary assets	Asset owners beat innovators under weak regimes	Build, partner, or license decisions
Architectural innovation	Organizational filters embed the old architecture	Incumbents fail at modest-looking reconfigurations	Diagnosing blind spots in product development
Dominant design	Market selection of a standard configuration	Shift from product to process innovation, shakeout	Timing entry and investment across the cycle
Open innovation	External knowledge flows and absorptive capacity	Inverted-U returns to search breadth	Designing partnership and licensing portfolios
Effectuation	Means-driven action and stakeholder commitments	Experts control rather than predict	Early-stage venture and internal-venture design
Discovery and creation	Alertness and prior knowledge versus enactment	Different actions under risk and uncertainty	Choosing planning versus iterative approaches
Bricolage	Recombining resources at hand	Selective bricolage supports growth	Frugal innovation and resource-constrained ventures
Liability of newness	Role learning, trust, and legitimacy take time	Failure hazard falls with age (or peaks in adolescence)	Venture funding, staging, and support

12.12 Discussion Questions

1. Choose a recent technology that commentators describe as “disruptive”. Using the conditions in Christensen, Raynor, and McDonald (2015), assess whether it fits the theory. What evidence, collected now, would let you test the prediction prospectively?
2. Apply Teece’s framework to a start-up developing a new medical device. Which complementary assets will it need, who controls them, and should it seek to compete, partner, or be acquired?
3. Henderson and Clark argue that architectural knowledge is embedded in communication channels. Design an organizational intervention that would help a product-development group detect architectural change earlier.
4. Using the inverted-U finding of Laursen and Salter (2006), how would you decide how many external innovation partners a mid-sized firm should maintain? What moderators would shift the optimum?
5. Design a study that distinguishes the discovery and creation views of opportunities. What sample would avoid selecting on success?
6. A venture fund observes that its portfolio companies fail most often in their third year. Which theory of age dependence does this support, and what changes in funding or support would the theory recommend?

12.13 Further Reading

- Schumpeter (1942): the source of the phrase creative destruction and still the best statement of why innovation, not price competition, drives capitalism.
- Teece (1986): the foundational analysis of who captures value from innovation. Read with Teece (2006) for the author’s later reflections.
- Christensen (1997), then A. A. King and Baatartogtokh (2015): the classic statement of disruption and a rigorous test of how well its cases fit, read together for balance.
- R. M. Henderson and Clark (1990): a concise, highly cited article that shows how organizations embed product architecture in their structure.
- P. Anderson and Tushman (1990): the clearest account of technology cycles and dominant designs.
- Laursen and Salter (2006): an influential empirical study of the benefits and limits of open search.
- Sarasvathy (2001) and Alvarez and Barney (2007): two articles that reframed how entrepreneurship scholars think about uncertainty and opportunity.
- T. Baker and Nelson (2005): a rich field study that shows how founders create resources from what others overlook.

Chapter 13

International Business

In 1997 Walmart, already the largest retailer in the world, entered Germany by acquiring the Wertkauf hypermarket chain and, a year later, the Interspar stores. It brought the formula that had worked at home: everyday low prices, a tightly managed supply chain, extended opening hours where the law allowed, and a distinctive service culture. In 2006 it announced the sale of its 85 German stores to Metro and left the country. Over roughly the same decades IKEA, a furniture retailer from a small town in southern Sweden, moved from Norway to Denmark, Switzerland, Germany, North America, China, and India, and became one of the most recognisable brands in the world. Why did a firm with enormous scale, capital, and operational know-how fail in a large, wealthy, and seemingly similar market, while a firm from a country of a few million people succeeded in dozens of very different ones?

The puzzle is the core of **international business** (IB) theory. Firms that cross borders face costs that local rivals do not, and they must possess something that offsets those costs. They must decide *where* to go, *how* to enter (exporting, licensing, joint ventures, acquisitions, or greenfield subsidiaries), *how fast* to commit resources, and *how much* to adapt. This chapter presents six theories that answer these questions. The **Uppsala model** explains internationalization as a gradual process of learning and commitment. The **OLI (eclectic) paradigm** explains why firms produce abroad rather than export or license. **Internalization theory** explains why firms replace cross-border markets with internal hierarchies. The **liability of foreignness** names the extra costs that foreign firms bear. **Institutional voids** describe the missing market intermediaries in emerging economies and the strategies they invite. **Born globals** challenge the gradualist story by internationalizing from inception. Together these theories form the backbone of research on multinational enterprises (MNEs), entry mode, and global strategy, and they draw heavily on Transaction cost economics (Section 6.3), Institutional theory (Section 7.2), and research on Cultural distance and [Hofstede](#hofstede) (Section @ref([hofstede](#hofstede) (Section 18.2))) (Section 18.2) (Table 13.1).

Table 13.1. Chapter at a glance: theories of international business.

Theory	Core question	Level of analysis	Key source
Uppsala model	How and in what order do firms internationalize?	Firm, firm-market relationship	Johanson & Vahlne (1977, 2009)
OLI paradigm	Why do firms produce abroad rather than export or license?	Firm, country, industry	Dunning (1980, 1988)
Internalization theory	Why do firms replace cross-border markets with internal hierarchies?	Transaction, firm	Buckley & Casson (1976), Rugman (1981)

Table 13.1. Chapter at a glance: theories of international business. *(continued)*

Theory	Core question	Level of analysis	Key source
Liability of foreignness	What extra costs do foreign firms bear, and how can they overcome them?	Subsidiary, firm	Hymer (1976), Zaheer (1995)
Institutional voids	How do missing market institutions shape strategy in emerging economies?	Country, business group, firm	Khanna & Palepu (1997, 2010)
Born globals	Why do some new ventures internationalize from inception?	New venture	Oviatt & McDougall (1994), Knight & Cavusgil (2004)

13.1 Uppsala Model

THEORY CARD

Core claim. Firms internationalize gradually. Because knowledge of foreign markets is mostly experiential, firms commit resources in small steps, entering psychically close markets first and moving from low-commitment to high-commitment modes as they learn.
Origins. Business studies at Uppsala University. Johanson and Wiedersheim-Paul (1975) documented the pattern in four Swedish firms. Johanson and Vahlne (1977) formalized it as a dynamic model of knowledge and commitment, and Johanson and Vahlne (2009) revised it around business networks.
Level of analysis. The firm and its relationships with particular foreign markets (and, after 2009, with network partners in those markets).
Key constructs. Psychic distance, experiential market knowledge, market commitment, commitment decisions, current activities, establishment chain, network position, liability of outsidership.
Boundary conditions. Strongest when market knowledge is tacit and country-specific, when the firm has limited resources, and when uncertainty about foreign markets is high. Weaker for firms with large resources, highly codified offerings, or managers with prior international experience.
Related theories. Organizational learning (Section 14.3), Absorptive capacity (Section 5.2), Cultural distance and Hofstede (Section 18.2), Born globals, [Real options](#real-options-theory) (Section 11.6) theory (Section 11.6).

13.1.1 Origins and intellectual history

In the early 1970s a group of researchers at Uppsala University studied how Swedish manufacturers had become international. Johanson and Wiedersheim-Paul (1975) traced the histories of four firms, Sandvik, Atlas Copco, Facit, and Volvo, and found two regularities. First, operations in a given country developed along a sequence they called the **establishment chain**: no regular exports, then exports through independent representatives (agents), then a sales subsidiary, and finally local production. Second, firms entered new markets in order of increasing **psychic distance**, defined as the factors that prevent or disturb the flow of information between the firm and the market, such as differences in language, education, business practices, culture, and industrial development. Swedish firms typically went to Norway, Denmark, and Finland before Germany or the United Kingdom, and

only later to more distant markets.

Johanson and Vahlne (1977) turned these observations into a theory. Their model has two **state aspects**, *market knowledge* and *market commitment*, and two **change aspects**, *commitment decisions* and *current activities*. Knowledge and commitment affect the decisions a firm makes and the activities it performs. Those activities, in turn, generate new experiential knowledge and deepen commitment. The model is therefore a feedback loop rather than a one-shot decision, and it borrows from the behavioural theory of the firm the assumptions of bounded rationality (Section 23.1), problemistic search, and incremental decision making.

Three decades later, Johanson and Vahlne (2009) revised the model in response to research on business networks and on rapidly internationalizing firms. In the revised model, markets are networks of relationships, and the main obstacle to internationalization is not foreignness as such but **outsidership**: lacking a position in the relevant network. The state variables become *knowledge and opportunities* and *network position*. The change variables become *relationship commitment decisions* and *learning, creating, and trust-building*.

13.1.2 Core logic and propositions

The engine of the model is the claim that the most important knowledge for operating abroad is **experiential**. Objective knowledge (market statistics, laws, tariffs) can be bought or taught. Knowledge of how customers decide, how local distributors behave, and which officials matter can mostly be acquired only by doing business in the market. Because experiential knowledge reduces perceived risk, commitment follows knowledge. The core propositions can be stated as follows:

- **P1 (Incremental commitment).** Additional commitment to a market will be made in small steps, unless the firm has large resources, market conditions are stable and transparent, or it has experience from similar markets (Johanson and Vahlne 1977).
- **P2 (Establishment chain).** Within a market, operations will tend to move from low-commitment modes (agents) to high-commitment modes (sales and production subsidiaries).
- **P3 (Psychic distance sequence).** Across markets, entry will tend to proceed from psychically close to psychically distant countries.
- **P4 (Network position).** After Johanson and Vahlne (2009), internationalization will follow the firm's relationships: firms enter markets where partners already have positions, and commitment grows as trust and mutual knowledge build.

Table 13.2 describes the establishment chain. Each stage involves more resources and creates more knowledge than the one before.

Table 13.2. The establishment chain (after Johanson and Wiedersheim-Paul, 1975).

Stage	Operating mode	Resource commitment	Knowledge generated	Typical risk
0	No regular export (sporadic orders)	Negligible	Almost none, information arrives by chance	Lost opportunities
1	Export through independent agent or distributor	Low: training, samples, margins	Second-hand market information filtered by the agent	Agent opportunism, weak control of positioning
2	Sales subsidiary	Moderate: staff, office, inventory	Direct customer and competitor knowledge	Fixed costs if the market disappoints

Table 13.2. The establishment chain (after Johanson and Wiedersheim-Paul, 1975). *(continued)*

Stage	Operating mode	Resource commitment	Knowledge generated	Typical risk
3	Production subsidiary	High: plant, local suppliers, workforce	Deep knowledge of labour, suppliers, regulation	Sunk investment, exposure to host-country policy

13.1.3 Mechanism

Figure 13.1 redraws the 1977 model. The loop explains why internationalization tends to be path-dependent: a firm's current position determines what it can learn next, and what it learns determines what it is willing to commit.

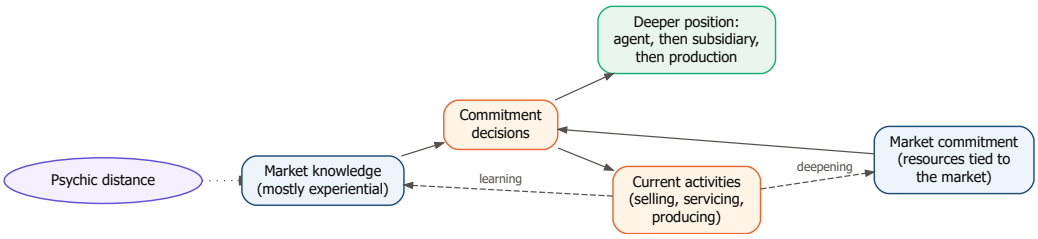


Figure 13.1. The Uppsala internationalization mechanism. Market knowledge and market commitment (state aspects) shape commitment decisions and current activities (change aspects), which generate new experiential knowledge and deepen commitment. Psychic distance slows the loop. Adapted from Johanson and Vahlne (1977).

Figure 13.2 shows the pattern the model predicts for a single firm entering three markets that differ in psychic distance. Entry into the close market comes earlier, and each subsequent step comes sooner, because knowledge accumulates faster where information flows more freely.

13.1.4 Evidence and extensions

The model fit the early history of Nordic and other European manufacturers well, and the establishment chain remains a useful description of how many small and medium-sized firms enter markets. Research on entry mode has confirmed that distance matters for how firms enter: Kogut and Singh (1988) found that firms from countries culturally distant from the United States were more likely to enter through joint ventures or greenfield investments than through acquisitions, which require integrating a foreign organization. Ghemawat (2001) broadened psychic distance into four dimensions of **CAGE distance**: cultural, administrative, geographic, and economic. He argued that firms systematically overestimate the attractiveness of large but distant markets because they ignore these frictions.

The 2009 revision moved the model toward network theory. If markets are webs of relationships, a firm can "jump" to a distant country when a key customer or partner takes it there, and a firm can fail in a close market if it cannot become an insider. This revision reconciles the model with evidence on rapid internationalization and with the view that relationships, not countries, are the unit of learning (Johanson and Vahlne 2009).

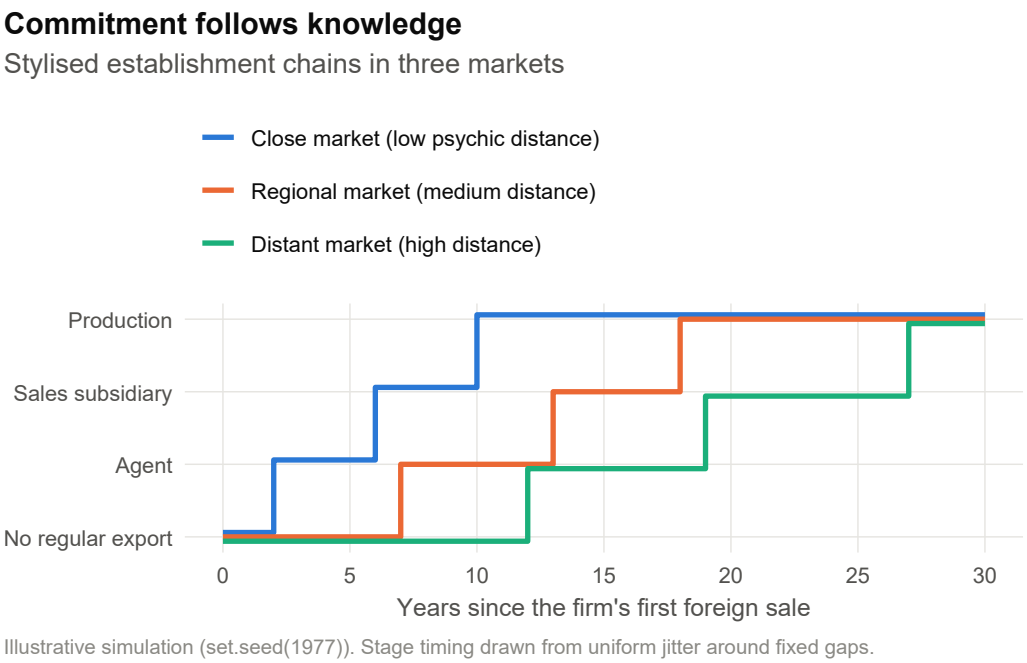


Figure 13.2. Illustrative stylised trajectories of commitment to three foreign markets that differ in psychic distance. Entry to more distant markets is later and each move along the establishment chain takes longer. Simulated data, not empirical.

DEBATE AND EVIDENCE

Is the Uppsala model a theory of learning or a description of caution? Forsgren (2002) argued that the model rests on a narrow conception of learning. Firms can also acquire knowledge by hiring experienced managers, imitating other firms, acquiring local companies, or searching deliberately, and these routes allow faster and less incremental internationalization than the model predicts. A second critique concerns **psychic distance** itself. Shenkar (2001) showed that distance measures built from national culture scores carry hidden assumptions of symmetry (the distance from A to B equals that from B to A), stability, linearity, and causality that rarely hold. A third critique is empirical: services, digital, and knowledge-intensive firms often skip stages or enter distant markets first, a pattern taken up in the section on born globals. Defenders reply that the model was never meant to predict a fixed sequence but to explain why commitment follows knowledge, and that the 2009 network version accommodates many of these exceptions.

13.1.5 Applications in business

BUSINESS CASE

IKEA's step-by-step internationalization. IKEA's expansion illustrates both the establishment chain and the psychic-distance sequence. After opening its first store in Sweden in 1958, the company opened its first foreign store in Norway in 1963 and in Denmark in 1969, both psychically close neighbours. Switzerland (1973) and Germany (1974) followed, then other

Western European markets, North America in the 1970s and 1980s, China in 1998, Russia in 2000, and India in 2018. Its experience in Japan shows the cost of moving too far too fast: IKEA entered Japan in the 1970s through a local partnership, withdrew in the 1980s, and returned in 2006 with wholly owned stores after studying Japanese homes and shoppers. In Uppsala terms, the firm's later commitments rested on knowledge that its first attempt had generated. IKEA also illustrates the model's limits: its concept is highly standardized (flat-pack products, self-service, a common catalogue), which reduced the amount of country-specific knowledge required and let it expand faster than a firm whose offering had to be rebuilt for each market.

13.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The greater the psychic distance between the home and host country, the more likely the firm's first entry mode will be low-commitment (exporting through agents rather than a subsidiary).
- *H2.* The firm's prior experience in a host country will increase the likelihood that its next investment there is wholly owned rather than shared.
- *H3.* The effect of psychic distance on entry timing will be weaker when a key domestic customer is already present in the host market (network position).

Designs. Event-history models of entry timing and mode transitions using firm-country panels. Sequence analysis of entry paths. Comparative case studies of entry and exit.
Measures. Psychic distance (perceptual surveys of managers) versus cultural distance (country scores from Hofstede (2001), combined with the index of Kogut and Singh (1988)). Experiential knowledge (years of operation in the country, number of prior entries in the region). Commitment (mode, assets, employees).
Data sources. Firm annual reports and subsidiary registries, ORBIS and similar ownership databases, export-transaction customs data, government surveys of outward investment.

COMMON PITFALL

Confusing psychic distance with cultural distance. Psychic distance is a *perception* held by managers about information flows, and it can differ across firms from the same country. Substituting a country-level cultural distance index assumes that all managers perceive the same distance and that culture is the only source of it. Measure perceptions where possible, include administrative, geographic, and economic distance (Ghemawat 2001), and do not assume distance is symmetric (Shenkar 2001).

13.2 OLI Paradigm

THEORY CARD

Core claim. A firm will engage in international production (foreign direct investment) when three conditions hold together: it has **ownership** advantages over local firms, there are **location** advantages to using those assets in a foreign country, and there are **internalization**

advantages to exploiting them within the firm rather than through the market.
Origins. International economics. Dunning (1980) proposed the "eclectic theory" as a synthesis of industrial-organization, location, and internalization ideas. Dunning (1988) restated it as a paradigm. It builds on Hymer (1976), Vernon (1966), and Buckley and Casson (1976).
Level of analysis. The firm (ownership and internalization) in its country and industry context (location).
Key constructs. Ownership advantages (asset and transaction advantages), location advantages, internalization advantages, mode of foreign involvement (export, licensing, foreign direct investment).
Boundary conditions. Explains the *existence* and *pattern* of foreign production at the level of industries and countries better than the timing of individual decisions. Assumes firms seek to exploit existing advantages abroad rather than acquire new ones.
Related theories. Internalization theory, Transaction cost economics (Section 6.3), Resource-based theory (Section 4.1), Liability of foreignness.

13.2.1 Origins and intellectual history

Before the 1960s, foreign direct investment (FDI) was usually treated as a form of capital movement: capital flowed to countries where returns were higher. Hymer's 1960 doctoral dissertation, published posthumously as Hymer (1976), overturned this view. Hymer observed that firms investing abroad do so to *control* operations, not merely to earn interest, and that a foreign firm faces disadvantages relative to local rivals. It must therefore possess some offsetting advantage, such as technology, brands, or scale, rooted in imperfect markets. Vernon (1966) added a dynamic account: products are developed in rich home markets, exported as they mature, and eventually produced abroad when costs dominate.

John Dunning combined these ideas. In Dunning (1980) he argued that no single theory explained international production, and that an "eclectic" approach was needed. The resulting **OLI** framework (also called the eclectic paradigm) asks three questions, summarised in Table 13.3. Dunning (1988) refined the ownership component by distinguishing **asset advantages** (proprietary resources such as technology or brands) from **transaction advantages** (benefits of common governance of a network of assets across countries).

Table 13.3. The three conditions of the OLI paradigm (after Dunning, 1980, 1988).

Advantage	Question it answers	Typical sources	If absent	Closest theory
Ownership (O)	Why can this firm compete abroad against locals?	Technology, brands, management know-how, scale, multinational coordination	No basis for foreign activity	Resource-based theory
Location (L)	Why produce in this country rather than at home?	Market size, factor costs, trade barriers, natural resources, incentives, clusters	Export from the home base	Trade and location theory
Internalization (I)	Why exploit the advantage inside the firm rather than sell it?	Tacit knowledge, risk of dissemination, quality control, weak contract enforcement	License, franchise, or contract	Transaction cost economics

13.2.2 Core logic and propositions

The paradigm is best read as a decision structure. Each condition is necessary for the next to matter, as Figure 13.3 shows.

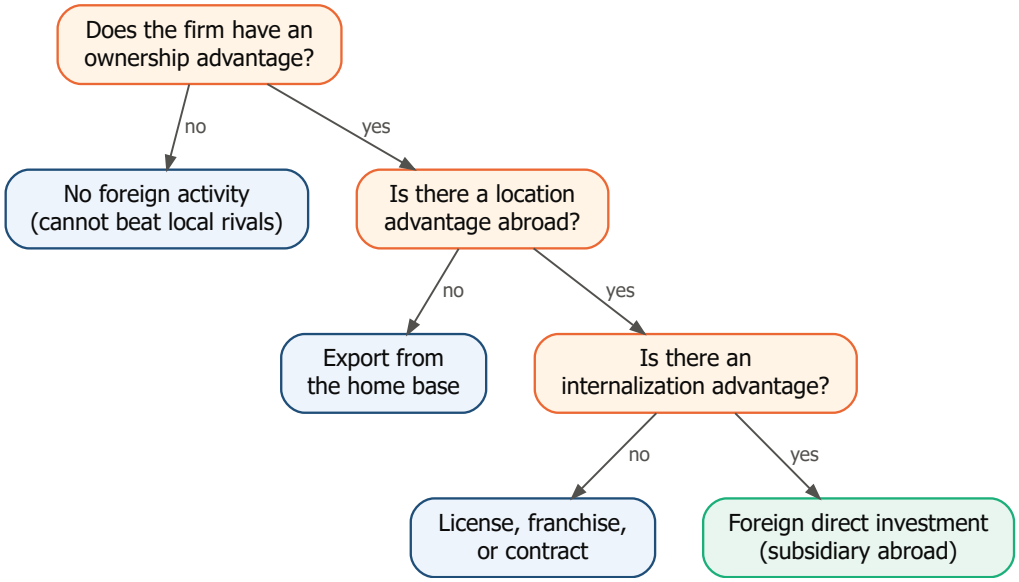


Figure 13.3. The OLI paradigm as a decision tree. Only when ownership, location, and internalization advantages are all present does foreign direct investment become the preferred mode. A simplified rendering of Dunning (1980, 1988).

- **P1.** The greater a firm's ownership advantages, the greater its propensity to engage in foreign activity of any kind (Dunning 1980).
- **P2.** Given ownership advantages, foreign production rather than export is more likely when host-country location advantages (market size, trade barriers, lower factor costs) are larger.
- **P3.** Given O and L advantages, FDI rather than licensing is more likely when markets for the firm's intangible assets are imperfect: when knowledge is tacit, when licensees could appropriate it, or when quality is hard to specify by contract.
- **P4.** The configuration of OLI advantages varies systematically by country, industry, and firm, so the pattern of international production differs across these contexts (Dunning 1988).

13.2.3 Evidence and extensions

The paradigm has organized a large empirical literature on the determinants of FDI and on entry-mode choice. Studies typically find that R&D intensity and advertising intensity (proxies for ownership advantages) predict outward FDI, that market size and trade barriers predict location, and that technological intensity predicts a preference for wholly owned subsidiaries over licensing. Later work added **strategic asset-seeking** investment, in which firms invest abroad to *acquire* advantages rather than exploit them, a motive that became prominent with acquisitions by emerging-market multinationals.

DEBATE AND EVIDENCE

Paradigm or laundry list? Critics argue that OLI is a classification scheme rather than a theory: almost any investment can be explained after the fact by naming some O, L, and I advantage, so the framework is hard to falsify. Internalization theorists add that ownership advantages are themselves the product of internalization and therefore not an independent condition (Buckley and Casson 2009). A related critique comes from Rugman and Verbeke (2004), who examined the largest MNEs and argued that most derive the bulk of their sales from their home region. If so, the "global" firm implied by location-free ownership advantages is rare, and ownership advantages are often regionally bounded. Dunning's response was that the paradigm is deliberately a general envelope within which more specific, testable theories can be nested.

13.2.4 Applications in business

BUSINESS CASE

Toyota's move from exporting to American production. Toyota built a strong **ownership** advantage in its production system, a set of tightly coupled practices for quality, inventory, and continuous improvement. Through the 1970s it served the United States mainly by exporting. In the early 1980s **location** advantages shifted as trade frictions grew and Japan agreed to limit car exports to the United States. Toyota's first American production was a joint venture with General Motors, New United Motor Manufacturing, Inc. (NUMMI), in Fremont, California, which began production in 1984. Its own wholly owned plant in Georgetown, Kentucky, began production in 1988. The **internalization** logic is visible in the sequence: the production system was largely tacit, embedded in routines and supplier relationships, and could not simply be licensed. The joint venture let Toyota learn about American workers and suppliers, and full ownership let it reproduce its system with control.

13.2.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms with higher R&D intensity will be more likely to serve foreign markets through wholly owned subsidiaries than through licensing.
- *H2.* Tariff increases in a host country will increase the probability that exporters to that country switch to local production, and more so for firms with strong ownership advantages.
- *H3.* Emerging-market firms will be more likely than advanced-economy firms to invest in countries with strong technology clusters (asset-seeking FDI).

Designs. Multinomial models of mode choice (export, license, joint venture, wholly owned). Gravity-style models of bilateral FDI. Natural experiments from trade policy shocks.

Measures. Ownership advantages (R&D and advertising intensity, patents). Location advantages (market size, wage levels, tariffs, cluster measures). Internalization advantages (tacitness, intellectual property protection, contract enforcement indices).

COMMON PITFALL

Labelling after the fact. Because almost any variable can be called an O, L, or I advantage, the paradigm invites post hoc explanation. State in advance which advantage each variable captures, specify the mode it should predict, and test the interaction implied by the paradigm: internalization advantages should matter only when O and L advantages are present.

13.3 Internalization Theory

THEORY CARD

Core claim. Multinational enterprises exist because markets for intermediate products, especially knowledge, are imperfect across borders. Firms internalize these markets, replacing arm's-length contracts with internal hierarchies, up to the point where the marginal benefits of internalization equal its marginal costs.
Origins. International economics. Buckley and Casson (1976), building on Coase (1937) and on Hymer's analysis of market imperfections (Hymer 1976). Generalized by Rugman (1981) and reviewed in Buckley and Casson (2009).
Level of analysis. The transaction (the cross-border transfer of an intermediate product) and the firm.
Key constructs. Intermediate products (knowledge, components, brands), market imperfection, internalization benefits and costs, dissemination risk, entry-mode control.
Boundary conditions. Applies when the asset being exploited is hard to price, protect, or transfer by contract. When contracts are enforceable and knowledge is codified, licensing and other market modes become efficient.
Related theories. Transaction cost economics (Section 6.3), OLI paradigm, Knowledge-based view (Section 4.3), Profiting from innovation (Section 12.3).

13.3.1 Core logic

Buckley and Casson (1976) argued that modern firms carry out a wide range of activities linked by flows of intermediate products: knowledge, expertise, components, and semi-finished goods. When markets for these products are perfect, firms can buy and sell them at arm's length. When markets are imperfect, firms gain by bringing the transaction inside the organization. Knowledge is the leading case. A buyer cannot value knowledge without seeing it, but once it has seen it, it need not pay for it. Knowledge is often tacit, so it cannot be fully specified in a contract. A licensee can learn the technology and become a competitor. Each of these problems raises the cost of the market relative to the hierarchy. When the knowledge is used in more than one country, internalization creates a multinational enterprise.

Rugman (1981) argued that internalization is a general theory of the MNE: the multinational is an internal market that substitutes for missing or imperfect external markets. The logic parallels Transaction cost economics (Section 6.3) (Williamson 1985), and E. Anderson and Gatignon (1986) made the connection explicit for entry mode. They classified foreign entry modes by the degree of **control** they confer and proposed that firms choose higher-control modes when transaction-specific assets are valuable, when the environment is uncertain, and when free-riding by partners is likely, trading control against the resource commitment that control requires. Table 13.4 summarizes the trade-offs.

Table 13.4. Foreign entry modes and the control-commitment trade-off (after Anderson and Gatignon, 1986).

Entry mode	Control	Resource commitment	Dissemination risk	Flexibility	Favoured when
Export (via agent)	Low	Low	Low	High	Small or uncertain market, little local adaptation
Licensing or franchising	Low to moderate	Low	High	Moderate	Codified knowledge, strong IP protection, enforceable contracts
Joint venture	Shared	Moderate	Moderate	Moderate	Partner holds local knowledge or required permits
Acquisition	High	High	Low	Low	Need for speed, existing local assets, low cultural distance
Greenfield subsidiary	Full	Highest	Lowest	Lowest	Tacit know-how, need to replicate routines exactly

13.3.2 Mechanism

Figure 13.4 simulates the core prediction. As the tacitness of the knowledge being transferred rises, the probability that a firm chooses a wholly owned subsidiary over licensing rises, and the curve shifts upward when host-country protection of intellectual property is weak.

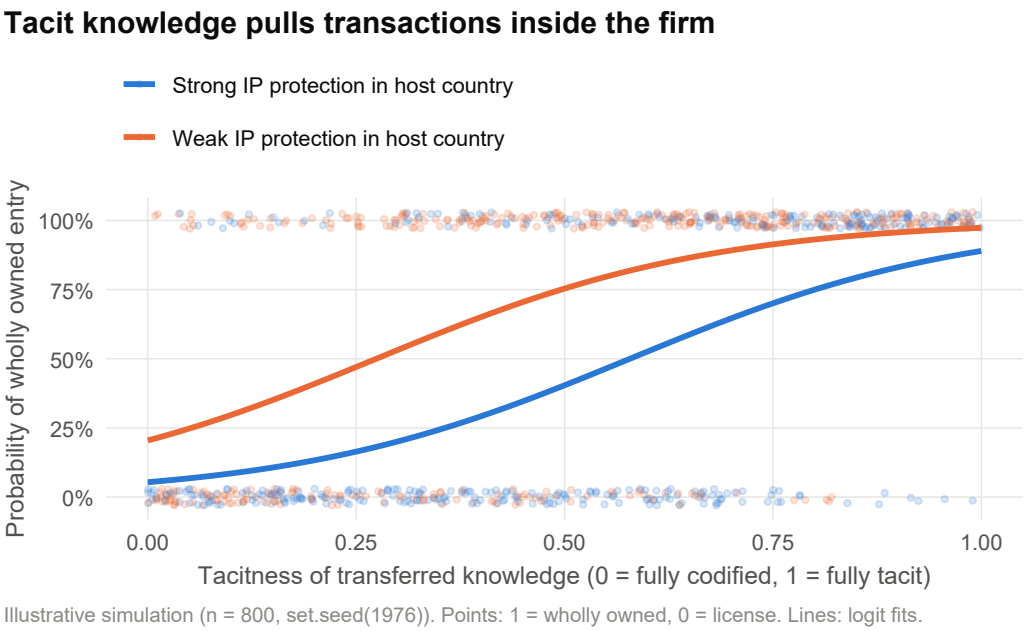


Figure 13.4. Illustrative simulation of internalization logic: the probability of choosing a wholly owned subsidiary rather than licensing rises with the tacitness of the transferred knowledge and is higher where intellectual property protection is weak. Simulated data, not empirical.

13.3.3 Evidence and extensions

Internalization theory underpins much research on entry mode, licensing, and the boundaries of the multinational. Its most important challenge came from the knowledge-based view. Kogut and Zander (1993) argued that MNEs exist not because markets for knowledge fail through opportunism, but because firms are social communities that are especially *efficient* at transferring knowledge internally. Studying transfers of technology by Swedish firms, they found that the less codifiable and the harder to teach the technology, the more likely it was transferred to a wholly owned subsidiary rather than to a licensee or joint venture. The prediction overlaps with internalization theory, but the mechanism differs: capability rather than market failure.

DEBATE AND EVIDENCE

Market failure or superior knowledge transfer? Transaction-cost versions of internalization theory explain the MNE by the hazards of the market (opportunism, appropriation). Knowledge-based versions explain it by the advantages of the organization (shared codes, routines, identity). The two explanations predict the same mode choice in many settings, which makes them hard to separate empirically. Distinguishing designs exploit cases in which contractual hazards are low but knowledge is tacit, or the reverse. Buckley and Casson (2009) argue that internalization theory can absorb both mechanisms, critics reply that this flexibility weakens its predictive content.

BUSINESS CASE

Illustrative scenario: a medical-device maker choosing a mode. A mid-sized European maker of surgical instruments wants to serve a fast-growing Asian market. A local firm offers to license its designs and manufacture them. Internalization theory directs management to three questions. Can the know-how be codified? (Much of the firm's advantage lies in precision finishing learned on the shop floor, so not fully.) Can contracts protect it? (Patent enforcement in the host country is slow.) Will quality failures damage the global brand? (Yes, surgeons buy on reputation.) All three point toward internalization, perhaps a wholly owned assembly and finishing plant supplied with critical components from home. If the know-how were codified and courts reliable, licensing would preserve capital and flexibility.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The tacitness of a firm's core technology will be positively associated with the probability of wholly owned entry.
- *H2.* Strengthening of host-country intellectual property protection will increase licensing relative to FDI, especially in industries with codifiable technologies.

Designs. Mode-choice models with measures of knowledge characteristics. Difference-in-differences around intellectual property reforms. Matched comparisons of licensed and internalized transfers of the same technology.

Measures. Codifiability and teachability scales (Kogut and Zander 1993), patent citations and claims, R&D intensity, indices of intellectual property protection and contract enforcement.

13.4 Liability of Foreignness

THEORY CARD

Core claim. Firms operating abroad incur costs that local firms do not, arising from unfamiliarity, distance, discrimination, and lack of legitimacy (Section @ref(legitimacy (Section 7.3))). These costs put foreign subsidiaries at a disadvantage that they must offset with firm-specific advantages or reduce over time.
Origins. Hymer's "costs of doing business abroad" (Hymer 1976). Named and tested by Zaheer (1995), with dynamics studied by Zaheer and Mosakowski (1997) and institutional foundations by Kostova and Zaheer (1999).
Level of analysis. The foreign subsidiary relative to comparable local firms.
Key constructs. Costs of doing business abroad, unfamiliarity hazards, discrimination hazards, relational hazards, institutional distance, legitimacy, liability of outsidership.
Boundary conditions. Larger when institutional and cultural distance are high, when products are visible or politically sensitive, and early in the subsidiary's life. Smaller for firms with strong, transferable advantages or local partners.
Related theories. Legitimacy (Section 7.3), Institutional theory (Section 7.2), Liability of newness (Section 12.10), Uppsala model, Cultural distance and Hofstede (Section 18.2).

13.4.1 Core logic

Hymer (1976) argued that a foreign firm is at a disadvantage relative to local firms because it lacks knowledge of local conditions, may be treated unfavourably by governments, consumers, and suppliers, and bears exchange-rate and communication costs. Zaheer (1995) named these costs the **liability of foreignness** (LOF) and identified four sources, summarised in Table 13.5.

Table 13.5. Four sources of the liability of foreignness (after Zaheer, 1995).

Source of liability	Example	Typical mitigation
Spatial distance	Travel, transport, time-zone coordination, communication	Regional hubs, digital coordination, local decision rights
Unfamiliarity with the local environment	Misreading customers, labour law, or supplier norms	Local hires, partners, acquisitions, experiential learning
Host-country environment	Lack of legitimacy, economic nationalism, discriminatory treatment by regulators or buyers	Local identity, community engagement, local ownership stakes
Home-country environment	Home-country restrictions on what may be sold where, home-country reputation	Compliance structures, distancing from home-country image

Zaheer's own evidence came from the currency-trading rooms of banks in New York and Tokyo. Foreign-owned trading rooms showed a disadvantage relative to local ones, and the study suggested that importing the parent's distinctive organizational capabilities did more to overcome the liability than imitating local practices (Zaheer 1995). Zaheer and Mosakowski (1997) then followed trading rooms worldwide over two decades and found that the survival disadvantage of foreign rooms was concentrated in their early years and diminished over time, consistent with learning. Mezias (2002) found that foreign subsidiaries in the United States faced more adverse labour-lawsuit judgments than domestic firms and identified strategies that reduced the gap, an unusually direct measure of discrimination and unfamiliarity hazards.

Kostova and Zaheer (1999) tied the concept to Institutional theory (Section 7.2). An MNE must gain legitimacy (Section 7.3) in many host environments at once, faces audiences that judge it with stereotypes and limited information, and may be held to different standards from local firms. Its legitimacy is complex because it depends on the institutional distance between home and host and on spillovers from the conduct of other foreign firms. Johanson and Vahlne (2009) extended the argument in a different direction: what hurts the entrant most is not being foreign but being an **outsider** to local business networks.

13.4.2 Mechanism

Figure 13.5 sketches the dynamic implied by Zaheer and Mosakowski (1997) and the Uppsala model. The foreign entrant starts with a cost disadvantage that shrinks as it learns and builds legitimacy. The shaded area is the cumulative liability the firm's advantages must cover.

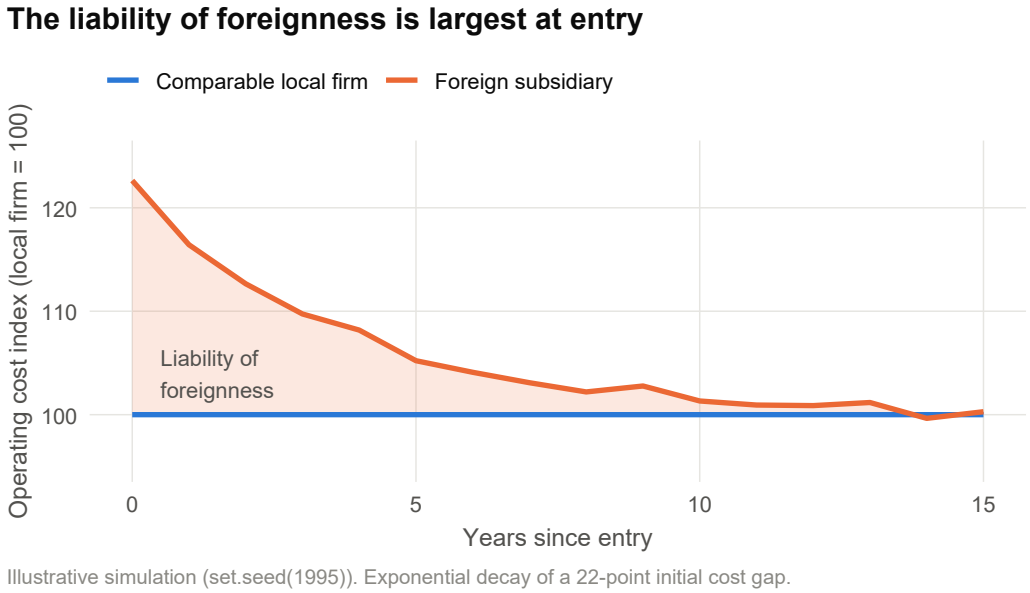


Figure 13.5. Stylised dynamics of the liability of foreignness. A foreign subsidiary's operating cost disadvantage relative to comparable local firms is largest at entry and declines as the subsidiary learns and gains legitimacy. Simulated values, not empirical.

13.4.3 Applications in business

BUSINESS CASE

Walmart in Germany. Walmart's German venture (1997-2006) is the most frequently cited case of liability of foreignness in retailing. The company entered by acquiring two chains with very different store formats, which complicated integration. Its model of "lean retailing" depended on practices common in the United States: control over large-format sites, flexible labour deployment, and strong bargaining power over suppliers. Christopherson (2007) argues that German institutions constrained these practices: land-use planning limited large out-of-town stores, labour-market institutions such as works councils and collective agreements limited flexible staffing, and the market was already served by efficient hard discoun-

ters such as Aldi and Lidl, which blunted the price advantage on which the model relied. Imported management practices also met resistance, and parts of the company's code of conduct were challenged in German labour courts. In 2006 Walmart sold its 85 German stores to Metro. In LOF terms, the firm suffered unfamiliarity and host-environment hazards, and its ownership advantages proved less transferable than expected because they were embedded in American institutions.

BUSINESS CASE

Uber in China. Uber launched in China in 2014 and competed with local ride-hailing firms that merged in 2015 to form Didi Kuaidi (later Didi Chuxing), whose predecessors were backed by Tencent and Alibaba. The local rival was embedded in the dominant payment and messaging ecosystems and in relationships with local governments. Uber's then chief executive publicly acknowledged heavy losses in the Chinese market, and in August 2016 Uber agreed to sell its China operations to Didi in exchange for a stake in the combined company. The case fits the **liability of outsidership** of Johanson and Vahlne (2009) as well as classic LOF: Uber's technology and brand were strong, but it lacked a position in the networks that shaped access to users, payments, and regulators.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Foreign subsidiaries will have lower survival rates than comparable local firms, and this gap will decline with subsidiary age.
- *H2.* The liability of foreignness will be larger when institutional distance between home and host country is greater.
- *H3.* Subsidiaries that hire local top managers will experience a smaller liability of foreignness in regulatory and legal outcomes.

Designs. Matched comparisons of foreign and domestic firms in the same industry and location. Survival models with time-varying effects. Natural experiments that change the salience of foreignness (for example, diplomatic disputes affecting consumer attitudes toward a home country).

Measures. Performance or survival gaps between foreign and domestic firms, institutional distance indices, regulatory outcomes, media sentiment toward foreign firms.

COMMON PITFALL

Attributing every foreign underperformance to foreignness. A foreign subsidiary may underperform because it is new, small, or poorly managed, not because it is foreign. Without a comparison to local firms of similar age and size, a "liability of foreignness" may simply be a liability of newness (Section 12.10). Separate the two by comparing foreign and local entrants of the same vintage.

13.5 Institutional Voids

THEORY CARD

Core claim. Emerging markets often lack the specialized intermediaries that make markets work in advanced economies: credit-rating agencies, reliable courts, independent auditors, executive search firms, retail distribution, and consumer protection. These **institutional voids** raise transaction costs, shape which organizational forms succeed (notably diversified business groups), and determine which strategies foreign entrants should use.
Origins. Strategy and international business. Khanna and Palepu (1997) introduced the idea to explain why diversified business groups prosper in emerging markets. Khanna and Palepu (2000) provided evidence from India, and Khanna and Palepu (2010) developed a framework for strategy in emerging markets.
Level of analysis. Country (the institutional context), business group, and firm.
Key constructs. Market intermediaries, product, labour, and capital market voids, business groups, internal markets, void-filling strategies.
Boundary conditions. Strongest where formal market institutions are underdeveloped. As institutions mature, the advantages of internal markets should decline.
Related theories. Institutional theory (Section 7.2), Transaction cost economics (Section 6.3), Liability of foreignness, Signaling theory (Section 8.2).

13.5.1 Core logic

In advanced economies, firms can raise capital from investors who rely on audited accounts and rating agencies, hire managers through executive search firms and business schools, and reach consumers through organized retail and credible advertising. Khanna and Palepu (1997) argued that these intermediaries are often missing or weak in emerging markets, and that the gap changes the logic of corporate strategy. Where external capital, labour, and product markets function poorly, a large diversified **business group** can replicate them internally: it can allocate capital across affiliates, train and rotate managers, and use its group name as a signal of quality to consumers and partners. Advice to "focus" on core businesses, appropriate where markets work, may therefore be wrong where they do not.

Khanna and Palepu (2000) tested the argument on Indian firms and found that the performance of group affiliates was related to the diversification of the group, with affiliates of the most diversified groups outperforming unaffiliated firms. Khanna and Palepu (2010) turned the concept into a framework for managers. They proposed mapping voids in product, labour, and capital markets and in the political and social system, and identified four strategic choices for firms entering emerging markets: replicate the home business model or adapt it, compete alone or collaborate with local partners, accept the market context or try to change it, and enter, wait, or exit. Firms can also treat voids as business opportunities by **filling** them, for instance by building distribution, certification, or payment infrastructure (Table 13.6).

Table 13.6. Institutional voids and responses (after Khanna and Palepu, 1997, 2010).

Market	Typical void	Business-group response	Foreign entrant's options
Capital	Weak disclosure, few credible rating agencies, shallow equity markets	Internal capital market across affiliates	Fund locally through parent, partner with local banks

Table 13.6. Institutional voids and responses (after Khanna and Palepu, 1997, 2010). *(continued)*

Market	Typical void	Business-group response	Foreign entrant's options
Labour	Few certified managers, limited executive search, weak training	Group training programmes, rotation of managers across affiliates	Build training academies, expatriates, local talent pipelines
Product	Little consumer protection, weak quality certification, fragmented retail	Group brand as a quality signal, own distribution	Invest in distribution, use brand and warranty as signals
Contracting	Slow courts, uncertain enforcement	Relational contracting within the group	Joint ventures, local partners, arbitration clauses
Payments and data	Low banking penetration, little credit history	In-house finance arms	Fill the void as a business (mobile money, credit scoring)

13.5.2 Mechanism

Figure 13.6 shows the institutional voids raise transaction costs in external markets.

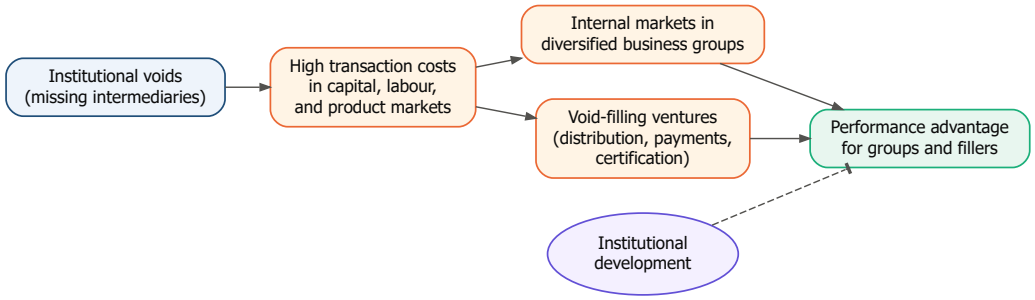


Figure 13.6. Institutional voids raise transaction costs in external markets. Business groups and void-filling firms respond by internalizing or supplying the missing intermediation. As institutions develop, the advantage of internal markets declines (dashed moderator). Based on the argument of Khanna and Palepu (1997, 2010).

13.5.3 Evidence and extensions

The concept has been widely used to explain the prevalence of business groups in Asia and Latin America, the strategies of emerging-market multinationals, and the design of market entry. It has also shaped thinking about market-creating innovation: Safaricom's M-Pesa, launched in Kenya in 2007, is often cited as a firm filling a financial-services void with mobile money. The framework links naturally to Institutional theory (Section 7.2) (voids are the absence of the regulative and normative pillars that make markets legible) and to Transaction cost economics (Section 6.3) (voids raise the costs of market governance).

DEBATE AND EVIDENCE

Efficiency or entrenchment? The void-filling account treats business groups as efficient responses to market failure. An alternative view holds that groups also reflect political connections, family control, and the expropriation of minority shareholders through transfers among affiliates. Both mechanisms may operate at once, and their balance likely varies with the stage of institutional development. A second debate concerns the term itself: critics note

that "voids" are rarely empty. Informal institutions such as kinship networks, reputation, and community enforcement often fill them, and a foreign firm that sees only absence may miss the rules that actually govern exchange.

13.5.4 Applications in business

BUSINESS CASE

Tata and Indian business groups. The Tata Group, whose origins go back to the trading business founded by Jamsetji Tata in 1868, is a classic diversified Indian business group, spanning steel, automobiles, information technology services, consumer goods, hotels, and power. Khanna and Palepu's account suggests why such a structure made sense in India's earlier institutional environment: the group name served as a signal of trustworthiness where consumer protection and quality certification were weak, group companies trained and rotated managers when external managerial labour markets were thin, and the group could channel capital to new ventures when equity markets were shallow (Khanna and Palepu 1997, 2000). As Indian markets developed, the group used its capabilities to expand abroad through acquisitions such as Tetley (2000), Corus (2007), and Jaguar Land Rover (2008). The case also illustrates the boundary condition: as voids close, the case for broad diversification weakens, and groups face pressure to show that each business can stand on its own.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The performance premium of business-group affiliation will be larger in countries and periods with weaker capital-market institutions.
- *H2.* Pro-market reforms will reduce the performance premium of group affiliation over time.
- *H3.* Foreign entrants will be more likely to use joint ventures in markets with larger contracting voids.

Designs. Cross-country panels of group-affiliated and independent firms. Difference-in-differences around institutional reforms (for example, securities regulation or bankruptcy law). Case studies of void-filling ventures.

Measures. Group affiliation (from national business-group directories), indices of institutional quality, transfer and related-party transaction data, diversification measures.

13.6 Born Globals

THEORY CARD

Core claim. Some firms internationalize from or near inception, deriving competitive advantage from resources and sales in multiple countries rather than expanding abroad gradually after building a domestic base.

Origins. International entrepreneurship. Oviatt and McDougall (1994) defined the **international new venture** and specified the elements needed for it to exist. Knight and Cavusgil (2004) linked born-global firms to innovation and organizational capabilities.

Level of analysis. The new venture and its founders.
Key constructs. International new venture, born global, speed and scope of internationalization, founder international experience, unique resources, international entrepreneurial orientation, digital and network reach.
Boundary conditions. Most common in knowledge-intensive, niche, or digital industries, where offerings are standardized, fixed costs are high, home markets are small, and global networks are accessible.
Related theories. Uppsala model, Effectuation (Section 12.7), Network effects (Section 17.6), Liability of newness (Section 12.10), Dynamic capabilities (Section 5.1).

13.6.1 Core logic

By the early 1990s researchers and consultants had noticed small firms that exported a large share of output within a few years of founding, often to distant markets. Oviatt and McDougall (1994) proposed a theory of the **international new venture** (INV), defined as a business organization that, from inception, seeks significant competitive advantage from the use of resources and the sale of outputs in multiple countries. They argued that four elements are necessary for an INV to exist: internalization of some transactions, alternative governance structures (such as networks and alliances, because a new firm cannot own many assets), foreign location advantages, and unique resources that are hard to imitate. The theory explicitly builds on internalization theory and the OLI paradigm, but replaces the assumption of a mature firm with that of a resource-poor start-up.

Knight and Cavusgil (2004) asked what makes such firms succeed. Drawing on case and survey data, they argued that born-global firms are characterized by an **innovative culture** and by knowledge and capabilities, including international entrepreneurial orientation and international marketing orientation, that support early international performance. The constructs tie the born-global phenomenon to the resource-based view and to research on entrepreneurial orientation. Operational definitions vary across studies but usually combine a short time between founding and first foreign sales with a minimum share of foreign sales.

Table 13.7 contrasts the two views, and Figure 13.7 shows their stylised paths.

Table 13.7. Gradual versus born-global internationalization.

Dimension	Uppsala-type firm	Born global
Starting point	Established domestic business	Global orientation at or near founding
Knowledge source	Experiential, accumulated in each market	Founders' prior experience, networks, codified offerings
Market sequence	Psychically close markets first	Lead markets and niches, often distant
Entry modes	Agents, then subsidiaries	Digital channels, partners, alliances, lean subsidiaries
Key constraint	Market knowledge and risk tolerance	Resources, legitimacy, and managerial bandwidth
Typical industries	Manufacturing, industrial goods	Software, digital media, biotechnology, niche high-tech
Theoretical anchor	Uppsala model	International entrepreneurship, Effectuation

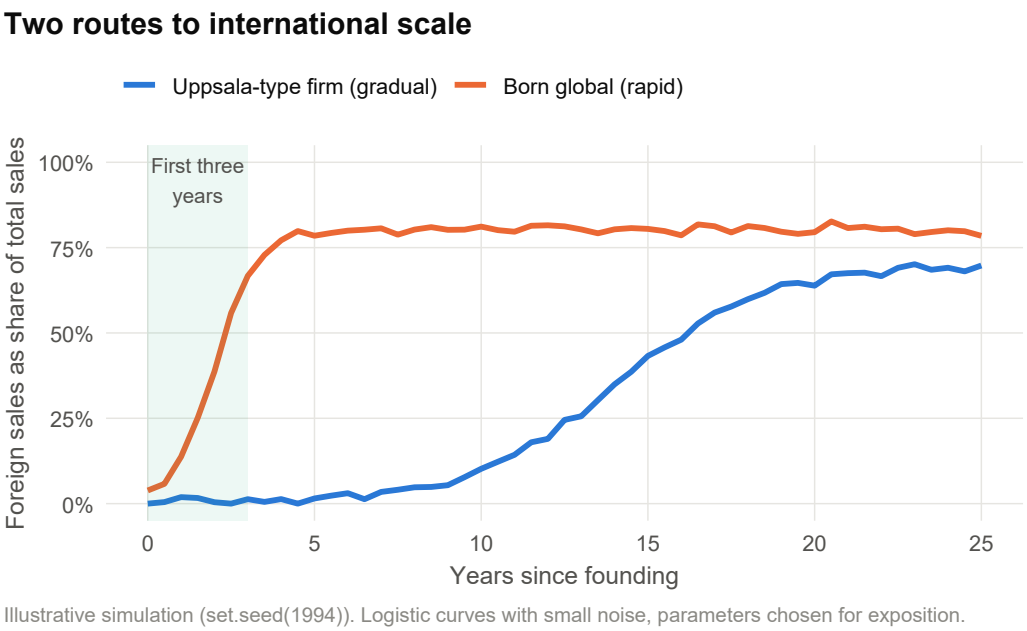


Figure 13.7. Illustrative stylised internationalization paths. An Uppsala-type firm builds a domestic base and raises its foreign sales share gradually. A born global reaches a high foreign share within a few years of founding. Simulated data, not empirical.

13.6.2 Evidence and extensions

Born globals have been documented in many countries, especially small open economies whose home markets are too small to support specialized products. The phenomenon has grown with digital distribution, which lowers the cost of reaching foreign customers and allows a firm to learn about many markets at once from usage data. Research has examined the role of founders' international experience, the importance of networks and alliances for resource-poor ventures, and the consequences of early internationalization for growth and survival.

DEBATE AND EVIDENCE

Do born globals refute the Uppsala model? Early studies presented born globals as a challenge to gradualism. The Uppsala authors replied that their model concerns learning and commitment, not age at entry: founders who bring experience and networks from previous jobs have already accumulated the knowledge the model emphasises, and the 2009 network version predicts rapid entry where relationships exist (Johanson and Vahlne 2009). Others note that many "born globals" are really "born regionals" that sell mostly within their home region, in line with evidence that even large MNEs are predominantly regional (Rugman and Verbeke 2004). A further concern is survivorship bias: the visible born globals are the successes, while early internationalization may also raise the risk of failure.

13.6.3 Applications in business

BUSINESS CASE

Skype and Spotify as born globals. Skype was founded in 2003 by Niklas Zennström and Janus Friis, with software developed by engineers in Estonia. As an internet telephony service distributed by download, it had no physical establishment chain to climb: users anywhere could adopt it, and its value grew with the number of users it could connect. It was acquired by eBay in 2005 and by Microsoft in 2011. Spotify, founded in Stockholm in 2006, launched its music streaming service in 2008 in several European markets at once and entered the United States in 2011. For Spotify the binding constraint was not market knowledge but licensing agreements with record labels and rights organizations in each country, a reminder that born globals still face country-specific institutions. Both cases fit the elements of Oviatt and McDougall (1994): unique, hard-to-imitate resources (software and network effects), alternative governance through partnerships, and location-independent delivery.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Founders' prior international work experience will be positively associated with the speed of a new venture's first foreign entry.
- *H2.* The positive effect of early internationalization on growth will be stronger for ventures with codified, digitally delivered offerings.
- *H3.* Early internationalization will increase the variance of venture outcomes (both rapid growth and failure).

Designs. Longitudinal cohorts of new ventures (for example, from national start-up registers). Survival and growth models with time to first foreign sale as the key variable. Founder-level data on experience and networks.

Measures. Time from founding to first foreign sale, share of foreign sales, number of countries served, founders' international experience, innovative culture and international entrepreneurial orientation scales (Knight and Cavusgil 2004).

COMMON PITFALL

Defining born globals by outcome alone. Studies that classify firms as born global because they succeeded internationally sample on the dependent variable. Define international new ventures by their *intent and activity at founding*, follow full cohorts including failures, and report how results change across alternative thresholds for speed and foreign sales share.

13.7 Chapter Summary

KEY TAKEAWAYS

- The **Uppsala model** explains internationalization as a loop in which experiential knowledge drives commitment, producing an establishment chain and a psychic-distance sequence. The 2009 revision locates the main barrier in network outsidership (Johanson and Vahlne 1977, 2009).

- The **OLI paradigm** states that FDI requires ownership, location, and internalization advantages at once. It organizes much research but is criticized as hard to falsify (Dunning 1980, 1988).
- **Internalization theory** explains the MNE as a response to imperfect markets for knowledge and other intermediate products. The knowledge-based view offers a competing mechanism with similar predictions (Buckley and Casson 1976, Kogut and Zander 1993).
- The **liability of foreignness** captures the extra costs of operating abroad, from unfamiliarity, discrimination, and lack of legitimacy. It is largest at entry and declines with learning (Zaheer 1995, Zaheer and Mosakowski 1997).
- **Institutional voids** in emerging markets favour business groups and void-filling strategies and call for strategies tailored to local market institutions (Khanna and Palepu 1997, Khanna and Palepu 2010).
- **Born globals** internationalize from inception using unique resources, networks, and digital reach, challenging but also extending gradualist accounts (Oviatt and McDougall 1994, Knight and Cavusgil 2004).

Table 13.8 presents the comparison of the chapter's theories.

Table 13.8. Comparison of the chapter's theories.

Theory	Unit of analysis	Key mechanism	Signature prediction	Typical business use
Uppsala model	Firm-market relationship	Experiential learning drives commitment	Gradual entry, close markets first	Sequencing market entry, pacing investment
OLI paradigm	Firm in country and industry context	Joint presence of O, L, I advantages	FDI only when all three hold	Choosing between export, licensing, and FDI
Internalization theory	Cross-border transaction	Market failure for intermediate products	Tacit knowledge leads to wholly owned modes	Entry-mode design, licensing decisions
Liability of foreignness	Foreign subsidiary versus locals	Unfamiliarity, discrimination, legitimacy deficits	Foreign firms underperform early, gap narrows	Localization, partner choice, legitimacy building
Institutional voids	Country and business group	Missing intermediaries raise transaction costs	Groups and void-fillers outperform where voids are large	Emerging-market strategy, partnering, new business models
Born globals	New venture	Unique resources, networks, digital reach	Early, rapid, and distant internationalization	Start-up global strategy, founder team design

13.8 Discussion Questions

1. The Uppsala model predicts that firms enter psychically close markets first. Using the CAGE framework, explain why Walmart's entry into Germany may have been more "distant" than it appeared. What would the model have advised?
2. Is the OLI paradigm a theory or a taxonomy? Propose a test of the paradigm that could, in principle, produce disconfirming evidence.
3. Internalization theory and the knowledge-based view both predict that tacit knowledge is

transferred to wholly owned subsidiaries. Design a study that could distinguish their mechanisms.

4. How would you separate a liability of foreignness from a liability of newness in data on foreign entrants? Specify the comparison group and the key variables.
5. Choose an emerging market you know. Map its institutional voids in capital, labour, and product markets, and identify one void that a new venture could fill profitably.
6. Do digital platforms make the Uppsala model obsolete? Use the Spotify case and the 2009 revision of the model to argue both sides.

13.9 Further Reading

- Johanson and Vahlne (1977): the original Uppsala model. Short and clear, and still the reference point for process theories of internationalization.
- Johanson and Vahlne (2009): the network revision that introduced the liability of outsidership.
- Dunning (1988): the fullest journal statement of the eclectic paradigm and its extensions.
- Buckley and Casson (2009): the authors' own review of three decades of internalization theory.
- Kogut and Zander (1993): the knowledge-based challenge to market-failure explanations of the MNE.
- Zaheer (1995): the article that named and measured the liability of foreignness.
- Khanna and Palepu (2010): a practical framework for strategy in emerging markets built on institutional voids.
- Oviatt and McDougall (1994): the foundational theory of international new ventures.

Chapter 14

Organizations, Teams, and Decision Processes

On the night of 27 January 1986, engineers at a NASA contractor argued against launching the space shuttle *Challenger* in record cold. The next morning the shuttle broke apart 73 seconds after lift-off. Seventeen years later, a piece of insulating foam struck the wing of *Columbia* during launch, engineers asked for better imagery of the damage, and the request went nowhere. The official investigation concluded that NASA's organizational culture and decision processes were as much a cause of the second accident as the foam itself (Columbia Accident Investigation Board 2003). Both tragedies involved talented people, abundant data, and formal review procedures. Neither can be explained by individual incompetence. They are failures of *organizing*: of how groups notice problems, interpret ambiguous signals, learn from the past, and decide whether to speak up or press on.

This chapter assembles the theories that explain how decisions are actually made inside organizations and teams, as opposed to how they should be made in a textbook. We begin with two accounts of decision processes that depart sharply from rational choice: the **garbage can model** and **escalation of commitment**. We then turn to how organizations accumulate and create knowledge (**organizational learning** and **knowledge creation**), to the shared self-understandings and values that steer behaviour (**organizational identity**, **organizational culture**, and **paradox theory**), to the interpersonal climate of teams (**psychological safety**), to how the composition and social context of the workforce shape attitudes (**attraction-selection-attrition** and **social information processing**), and finally to **sensemaking**, the process through which people construct the situations they then act upon. Together these theories form the core of the behavioural and interpretive tradition in organization theory, a tradition that starts from the observation that organizations are boundedly rational, socially constructed, and path dependent (Cyert and March 1963, Weick 1979) (Table 14.1).

Table 14.1. Chapter at a glance: theories of organizations, teams, and decision processes.

Theory	Core question	Level of analysis	Key source
Garbage can model	How are choices made when goals, technology, and participation are all unclear?	Organization	Cohen, March & Olsen (1972)
Escalation of commitment	Why do decision makers keep investing in failing courses of action?	Individual / group / organization	Staw (1976)

Table 14.1. Chapter at a glance: theories of organizations, teams, and decision processes. *(continued)*

Theory	Core question	Level of analysis	Key source
Organizational learning	How do organizations encode experience into routines, and when do they question them?	Organization	Argyris & Schön (1978), Levitt & March (1988)
Knowledge creation	How is new knowledge created through conversion of tacit and explicit knowledge?	Individual to organization	Nonaka (1994)
Organizational identity	Who are we as an organization, and how does that shape action?	Organization	Albert & Whetten (1985)
Organizational culture	What shared assumptions and values guide behaviour?	Organization / group	Schein (2010)
Paradox theory	How can organizations sustain contradictory demands over time?	Individual / organization	Smith & Lewis (2011)
Psychological safety	When do team members feel safe to speak up, admit errors, and learn?	Team	Edmondson (1999)
Attraction-selection-attrition	Why do organizations become homogeneous over time?	Individual / organization	Schneider (1987)
Social information processing	How do social cues shape job attitudes?	Individual / workgroup	Salancik & Pfeffer (1978)
Sensemaking	How do people construct meaning from ambiguous events and act on it?	Individual / group / organization	Weick (1995)

14.1 Garbage Can Model

THEORY CARD

Core claim. In "organized anarchies", decisions are not the output of a sequence that runs from problem to analysis to choice. They emerge when independent streams of problems, solutions, participants, and choice opportunities happen to meet. Timing and access, not intention, explain much of what gets decided.

Origins. Organization theory and public administration. M. D. Cohen, March, and Olsen (1972), extending the behavioural theory of Cyert and March (1963). Applied to public policy by Kingdon (1984).

Level of analysis. Organization (universities, government agencies, professional firms, new ventures), and policy systems.

Key constructs. Problematic preferences, unclear technology, fluid participation, problem stream, solution stream, participant stream, choice opportunities, decision by resolution, oversight, and flight, load, access structure.

Boundary conditions. Strongest when goals are inconsistent or ill-defined, cause-effect knowledge is weak, and attention is scarce and shifting. Weaker in tightly coupled, well-understood operational settings.

Related theories. Behavioral theory of the firm (Section 11.7), Attention-based view (Section

> 11.8), Bounded rationality (Section 23.1), Sensemaking.

14.1.1 Origins and core logic

M. D. Cohen, March, and Olsen (1972) began with an empirical observation about universities. Their goals are multiple and contested (**problematic preferences**), nobody fully understands how teaching or research produce results (**unclear technology**), and faculty, students, and administrators drift in and out of decisions depending on their other commitments (**fluid participation**). The authors called such settings **organized anarchies** and argued that standard models of choice, which assume that problems precede and motivate solutions, describe them poorly.

Instead, an organization can be viewed as a collection of choices looking for problems, issues and feelings looking for decision situations in which they might be aired, solutions looking for issues to which they might be the answer, and decision makers looking for work. Four relatively independent **streams** flow through the organization. A **choice opportunity**, such as a budget meeting, a hiring committee, or a strategic review, is the "garbage can" into which participants dump whatever problems and solutions are currently active. What comes out depends on the mix in the can at that moment.

The authors formalized the argument in a computer simulation and identified three ways in which choices are made:

- **Resolution.** A choice resolves the problems attached to it, the classic rational pattern.
- **Oversight.** A choice is made quickly, before problems attach to it, so it resolves nothing that was not already settled.
- **Flight.** Problems attach to a choice, but the choice is delayed until the problems leave for another, more attractive choice opportunity. The choice is then made without resolving them.

A central result of the simulation was that resolution is *not* the dominant style of decision making except when the load on the system is light. Under heavy load, oversight and flight dominate, important problems are more likely to remain unresolved, and decision makers spend much of their time on the same problems in different arenas (M. D. Cohen, March, and Olsen 1972).

14.1.2 Mechanism

Figure 14.1 shows the garbage can model: four loosely coupled streams meet in choice opportunities.

The garbage can model does not claim that organizations are irrational. It claims that **temporal coincidence** matters. A solution championed by a powerful participant can wait years for a problem to which it can be attached, and a problem that arrives when attention is elsewhere may be ignored even if it is severe. Kingdon (1984) adapted this idea to public policy as the **multiple streams framework**: a policy change becomes likely when a problem stream, a policy (solution) stream, and a political stream converge during a brief **policy window**, often with help from a **policy entrepreneur** who couples them (Table 14.2).

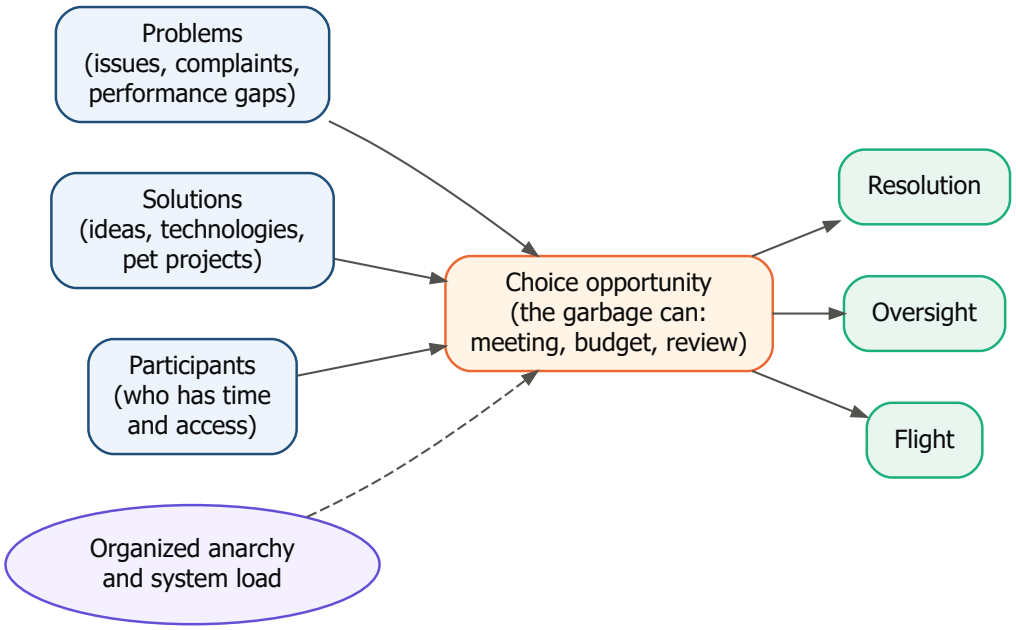


Figure 14.1. The garbage can model: four loosely coupled streams meet in choice opportunities. Whether a choice resolves problems depends on timing, load, and access rules rather than on a planned problem-solving sequence. Based on Cohen, March & Olsen (1972).

Table 14.2. Rational choice versus the garbage can account of organizational decisions.

Classic rational sequence	Garbage can account	Managerial implication
Problems are identified, then solutions are sought	Solutions often exist first and search for problems	Track which solutions are being promoted, not only which problems exist
Goals are clear and stable	Preferences are discovered through action	Expect goals to be rationalised after the fact
Participants are fixed	Participation depends on time, energy, and other arenas	Agenda and attendance shape outcomes as much as arguments
Choices resolve problems	Many choices are made by oversight or flight	Audit whether decisions actually closed the issues they were meant to address
Timing is incidental	Timing and coincidence are central	Prepare solutions in advance and act when windows open

14.1.3 Evidence and extensions

The original article was a thought experiment illustrated by simulation, and its empirical support comes mostly from qualitative studies of universities, public agencies, and policy making. Kingdon's interview-based study of United States federal policy agendas is the most influential application (Kingdon 1984). The model also influenced work on attention allocation and organizational routines, including the attention-based view (Section 11.8).

DEBATE AND EVIDENCE

Does the simulation match the theory? Bendor, Moe, and Shotts (2001) offered the sharpest critique. They argued that the computer model implements a far more structured organization than the verbal theory describes (for example, access rules and energy loads are fixed by design), that individual agency disappears from the model, and that several of the verbal claims do not follow from the simulation. Defenders reply that the garbage can is best read as a sensitizing metaphor about loose coupling and timing, not as a precise predictive model. A second limitation is falsifiability: almost any decision outcome can be narrated as the product of colliding streams, so researchers need to specify ex ante which stream configurations should produce which outcomes.

14.1.4 Applications in business

BUSINESS CASE

Illustrative scenario: the technology looking for a problem. A retailer's analytics team has built a sophisticated recommendation engine. No business unit asked for it. For two years it is presented at quarterly planning meetings without adoption. Then a sharp drop in online conversion coincides with a new chief digital officer's arrival and an open budget cycle. Within weeks the engine is funded as "the answer" to conversion, although nobody has diagnosed why conversion fell. The garbage can lens explains the timing and warns about the risk: the problem may not be the one the solution solves. A practical response is to separate the *window* (the moment of attention) from the *diagnosis* by requiring a short root-cause analysis before any solution already in the pipeline is attached to a newly salient problem.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The probability that a pre-existing internal proposal is adopted rises sharply in periods following an exogenous performance shock, independent of the proposal's fit with the shock.
- *H2.* Under high decision load (many concurrent choice opportunities), a larger share of decisions will be recorded without resolution of the issues raised in meeting minutes.

Designs and data. Meeting minutes and board agendas coded for problems, solutions, and attendance. Event-history models of proposal adoption. Agent-based simulation that explicitly varies access structures. Policy-agenda datasets for public-sector applications.

COMMON PITFALL

Using the garbage can as an all-purpose excuse. Calling a decision process a "garbage can" explains nothing unless the analyst shows that the streams were in fact independent and that timing, rather than deliberate choice, determined the outcome. The model is a description of particular conditions (organized anarchy and high load), not a general claim that organizations are chaotic.

14.2 Escalation of Commitment

THEORY CARD

Core claim. Decision makers who are responsible for a course of action tend to continue, and even increase, investment in it after receiving negative feedback, beyond the point that a forward-looking analysis would justify.
Origins. Organizational behaviour and social psychology. Staw (1976), with the theoretical framework of Staw (1981). The related **sunk cost effect** was documented by Arkes and Blumer (1985).
Level of analysis. Individual, group, and organization (projects, product lines, acquisitions, public programmes).
Key constructs. Sunk costs, personal responsibility, self-justification, negative feedback, project completion, framing of losses, public commitment, de-escalation.
Boundary conditions. Requires a prior decision, ambiguous or negative feedback, and an opportunity to reinvest. Effects are weaker when decision makers are not responsible for the initial choice, when the cost of continuation is salient, and when exit is institutionally easy.
Related theories. Cognitive dissonance (Section 21.1), Prospect theory (Section 23.3), Mental accounting (Section 23.5), Agency theory (Section 6.1), Real options (Section 11.6), Groupthink (Section 30.6).

14.2.1 Origins and intellectual history

The title of Staw's 1976 article, "Knee-deep in the big muddy", alludes to a Vietnam-era protest song, and the war itself, in which the United States government repeatedly increased its commitment in the face of discouraging results, was a motivating example. Staw (1976) tested the underlying idea in a role-playing experiment. Business students acting as a corporate financial officer allocated research and development funds to one of two divisions, then learned how the division had performed, then allocated a second round of funds. Participants who had personally chosen the initial division and then learned it had performed poorly allocated significantly more money to it in the second round than participants who had not made the original choice. Responsibility for a losing decision produced more, not less, investment in it.

Staw (1981) interpreted the finding through **self-justification**: people are motivated to demonstrate, to themselves and others, that an earlier decision was correct. Reinvesting keeps open the possibility that the decision will be vindicated. Admitting failure closes it. Four years later, Arkes and Blumer (1985) documented the closely related **sunk cost effect**, the tendency to continue an endeavour once an investment of money, effort, or time has been made. In one of their field studies, theatre patrons who had been randomly assigned to pay full price for season tickets attended more performances in the first part of the season than patrons who had received discounts, although the money was spent either way.

The biologist Richard Dawkins and a colleague had independently named the error the **Concorde fallacy**, after the Anglo-French supersonic airliner whose development continued despite mounting evidence that it would not be a commercial success (Dawkins and Carlisle 1976). Arkes and Ayton (1999) reviewed the animal and human evidence and argued that the effect is distinctively human, rooted in an overgeneralised rule against wasting resources.

14.2.2 Core logic and propositions

Escalation is an error only relative to a normative benchmark. In rational choice, past costs are **sunk**: they cannot be recovered whatever one does next, so they should not influence a forward-looking decision. The right question is always whether the *incremental* cost of continuing is justified by the *incremental* expected value. Escalation occurs when sunk costs, prior commitments, or identity concerns enter that comparison. Building on Staw (1981), Brockner (1992), and the meta-analysis of Sleesman et al. (2012), the theory can be summarised in propositions:

- **P1 (Responsibility).** Decision makers who are personally responsible for an initial decision escalate more after negative feedback than decision makers who inherit it.
- **P2 (Sunk costs).** The larger the prior investment, the greater the willingness to continue, holding forward-looking prospects constant.
- **P3 (Loss framing).** Because continuing offers a chance to avoid a certain loss, decision makers in the loss domain take the risky option of continuation, consistent with prospect theory (Section 23.3).
- **P4 (Completion).** Willingness to continue rises as a project approaches completion, even when the remaining cost exceeds the remaining value.
- **P5 (Structural and social binding).** Public commitment, political support, and institutionalised projects make exit costly for the organization, even after individual decision makers have lost confidence.

Figure 14.2 illustrates the economic logic with a stylised project. The value of completion is fixed. Each quarter, the estimate of the cost still required to finish rises. A forward-looking decision maker exits once the remaining cost exceeds the remaining value. An escalating decision maker, anchored on what has already been spent, continues long after that point.

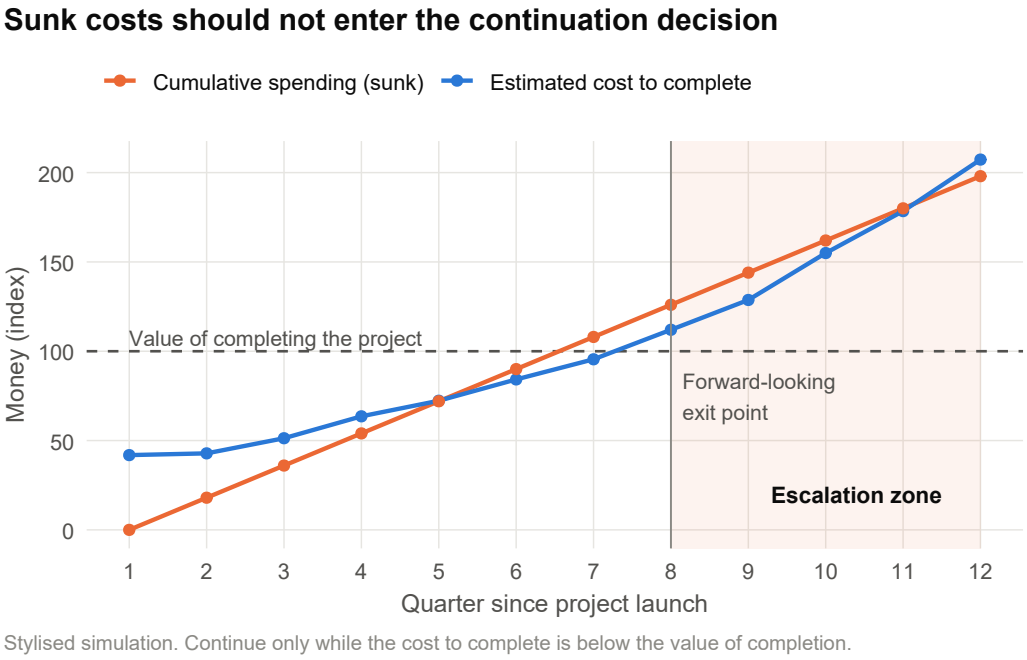


Figure 14.2. Stylised escalation of commitment. The rational exit point is where the estimated cost to complete first exceeds the value of completion. Funds spent after that point (shaded) represent escalation. Illustrative numbers only.

14.2.3 Determinants of escalation

J. Ross and Staw (1993) organized the many determinants of escalation into four classes, a framework that later reviews retained (Sleesman et al. 2012, 2018) (Table 14.3).

Table 14.3. Determinants of escalation of commitment, organized following Ross and Staw (1993).

Class of determinant	Examples	Theoretical root	De-escalation lever
Project	Ambiguous feedback, large payoff on completion, long time horizons, closeness to completion	Expected value, information economics	Pre-set milestones and kill criteria tied to forward-looking value
Psychological	Personal responsibility, self-justification, loss framing, sunk cost reasoning	Dissonance, prospect theory	Rotate decision makers and separate initial and continuation decisions
Social	Public commitment, face-saving, norms of consistency for leaders	Impression management	Reward accurate early reporting of bad news
Structural	Political coalitions, institutionalised projects, side bets, sunk investments in people and assets	Institutional theory, resource dependence	Independent review boards and staged funding

14.2.4 Evidence and extensions

The phenomenon is well supported. Sleesman et al. (2012) meta-analysed escalation research and confirmed that sunk costs, personal responsibility, and project-related factors such as proximity to completion predict continued commitment, while showing that effect sizes vary considerably across settings. Field studies trace escalation in nuclear power, information systems, banking, and professional sports. J. Ross and Staw (1993) reconstructed the history of the Shoreham nuclear power plant on Long Island, whose cost estimates rose from tens of millions to billions of dollars over two decades, and which was completed but never entered commercial operation. They showed how project, psychological, social, and structural forces dominated at different stages: early commitment was driven by projected economics, later commitment by political and structural binding. In information systems, Keil, Mann, and Rai (2000) surveyed audit professionals about runaway software projects and compared four explanations (self-justification, prospect theory (Chapter 24), agency theory, and approach-avoidance theory), finding that several of them contributed to explaining escalation.

DEBATE AND EVIDENCE

Error or rational persistence? Not every decision to continue is escalation. Persistence can be rational when feedback is noisy, when early losses are expected in innovation, or when reputational benefits of consistency are real. Critics note that many laboratory designs make it impossible to tell whether participants are biased or are updating sensibly on ambiguous information. Case labels can also oversimplify. The Concorde programme was governed by an Anglo-French intergovernmental agreement that made unilateral cancellation diplomatically costly, a structural rather than a purely psychological driver of continuation. The theoretical frontier, reviewed by Sleesman et al. (2018), is a multilevel account that separates individual biases from team and organizational forces and studies **de-escalation** as a process in its own right.

14.2.5 Applications in business

BUSINESS CASE

Large IT programmes. Enterprise software and public-sector IT projects are the modern home of escalation research because they combine long horizons, ambiguous progress measures, and highly visible sponsors. The United Kingdom's National Programme for IT in the National Health Service, launched in 2002 and dismantled in 2011 after sustained criticism from parliamentary committees and the National Audit Office, is a widely discussed public example. Research on such projects (Keil, Mann, and Rai 2000) suggests three practical safeguards: **stage-gated funding** that requires a fresh forward-looking business case at each gate, **separation of roles** so that the sponsor who initiated a project does not alone decide whether it continues, and **pre-registered kill criteria** written when the project is launched and the sponsor's ego is not yet invested.

BUSINESS CASE

Venture capital and real options. Venture investors explicitly structure funding in stages, so that each round is a new decision about the incremental investment. In principle, staging converts a single large commitment into a sequence of real options (Section 11.6). In practice, escalation can reappear through follow-on investments made to protect earlier positions. Many funds therefore have partners other than the original sponsor evaluate follow-on rounds.

14.2.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Executives who championed an acquisition will be less likely to divest the acquired unit after poor performance than successor executives, controlling for unit performance.
- *H2.* The positive effect of cumulative project spending on continuation will be weaker when funding is staged and continuation decisions are made by an independent committee.
- *H3.* Proximity to completion will increase continuation more strongly when progress is measured by spending than when it is measured by delivered functionality.

Designs. The classic two-stage allocation experiment (Staw 1976) remains the workhorse. Field designs exploit executive turnover as a shift in personal responsibility, compare initiators and successors, or use project-level data from IT portfolios, pharmaceutical pipelines, and sports drafts.

Measures. Additional resources allocated after negative feedback, time to project termination, responsibility (initiator versus successor), sunk cost magnitude, and proximity to completion.

COMMON PITFALL

Labelling every failed persistence as bias. Escalation is defined relative to a forward-looking benchmark. To claim escalation, researchers and managers need to show that the

information available at the time implied that continuation had negative incremental expected value. Hindsight makes this look easier than it was. Conversely, managers who quote "sunk costs are sunk" to kill every struggling innovation project risk the opposite error of premature abandonment.

14.3 Organizational Learning

THEORY CARD

Core claim. Organizations learn by encoding inferences from experience into routines that guide behaviour. Most learning refines existing routines (single-loop), while deeper learning questions the goals and assumptions behind them (double-loop).
Origins. Organization theory and the behavioural theory of the firm. Argyris and Schön (1978), B. Levitt and March (1988), G. P. Huber (1991), and the learning-curve tradition summarised by Argote and Epple (1990).
Level of analysis. Organization, with links to teams and individuals.
Key constructs. Routines, single-loop and double-loop learning, theories in use versus espoused theories, knowledge acquisition, information distribution, interpretation, organizational memory, learning curves, knowledge depreciation and transfer, competency traps, superstitious learning.
Boundary conditions. Learning requires feedback that can be observed and attributed. Learning is weaker when experience is rare, outcomes are noisy, or cause and effect are separated in time.
Related theories. Exploration and exploitation (Section 5.3), Absorptive capacity (Section 5.2), Dynamic capabilities (Section 5.1), Behavioral theory of the firm (Section 11.7), Psychological safety.

14.3.1 Origins and intellectual history

Three intellectual lines converge in organizational learning theory. The first is the **behavioural theory of the firm**, in which organizations adapt their goals, attention, and search rules in response to experience (Cyert and March 1963). The second is the action-science tradition of Argyris and Schön (1978), which focused on the gap between what managers say they believe and what their behaviour reveals. The third is the economics and engineering literature on **learning curves**, dating back to the observation that the labour required to build an airframe falls predictably as cumulative production rises (Wright 1936).

B. Levitt and March (1988) synthesised the first line in an influential review. Organizational learning, they argued, is **routine-based**, **history-dependent**, and **target-oriented**. Organizations encode inferences from history into routines (rules, procedures, technologies, conventions) that persist even as personnel change. Routines are adjusted when performance falls short of aspirations. G. P. Huber (1991) then mapped four constructs that organize the field: **knowledge acquisition**, **information distribution**, **information interpretation**, and **organizational memory**. Fiol and Lyles (1985) distinguished lower-level learning, which adjusts behaviour within existing rules, from higher-level learning, which changes the rules themselves.

14.3.2 Core logic: single-loop and double-loop learning

Argyris (1976) and Argyris and Schön (1978) distinguished two kinds of learning. In **single-loop learning**, an organization detects an error and corrects it by changing its action strategy while leaving its governing values intact, much as a thermostat adjusts heating to hold a set temperature. In **double-loop learning**, the organization questions the governing variables themselves: it asks whether the target temperature is right. Argyris and Schön also distinguished **espoused theories** (what people say guides their action) from **theories in use** (what their behaviour reveals). They argued that defensive routines, designed to avoid embarrassment or threat, systematically block double-loop learning (Figure 14.3).

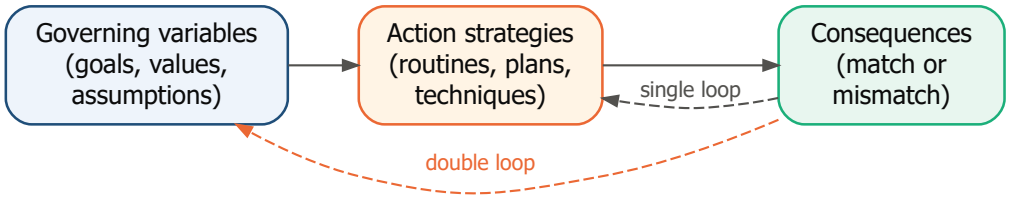


Figure 14.3. Single-loop and double-loop learning. Single-loop learning changes action strategies to close a gap between consequences and governing variables. Double-loop learning questions the governing variables themselves. Based on Argyris and Schön (1978).

Table 14.4 presents the single-loop versus double-loop learning.

Table 14.4. Single-loop versus double-loop learning.

Feature	Single-loop learning	Double-loop learning
Question asked	Are we doing things right?	Are we doing the right things?
What changes	Action strategies and routines	Goals, norms, and underlying assumptions
Typical trigger	Performance below target	Repeated failure, crisis, or new framing
Typical method	Process improvement, root-cause fixes	Reframing, strategic review, surfacing assumptions
Main obstacle	Weak feedback or poor measurement	Defensive routines that make assumptions undiscussable
Related idea	Exploitation	Exploration

14.3.3 Learning curves: experience, transfer, and forgetting

The quantitative heart of the field is the **learning curve**: unit cost or labour hours fall as a power function of cumulative output. Wright (1936) observed this regularity in aircraft manufacturing. Argote and Epple (1990) reviewed evidence from many industries and drew three conclusions that transformed the learning curve from an engineering rule of thumb into an organizational theory. First, **rates of learning vary dramatically across organizations**, even those producing the same product. Second, **knowledge depreciates**: it decays when production is interrupted or staff turn over. Third, **knowledge transfers** across units, but imperfectly. Darr, Argote, and Epple (1995) showed these patterns in pizza franchises: stores learned from cumulative experience, knowledge transferred across stores owned by the same franchisee but not across stores with different owners, and acquired knowledge depreciated quickly (Figure 14.4).

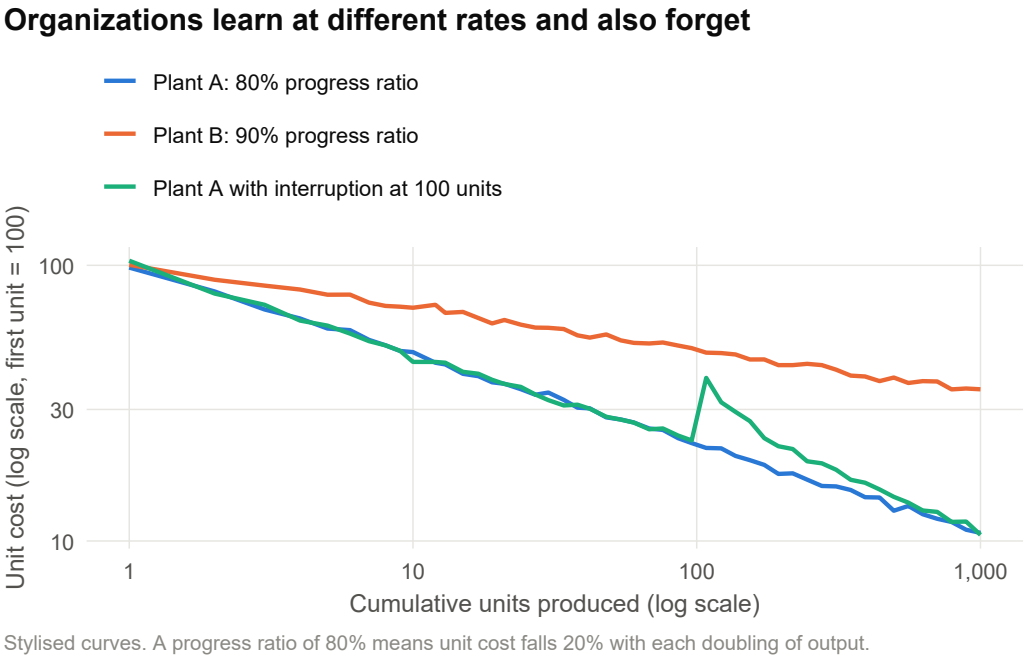


Figure 14.4. Stylised learning curves. Unit cost falls with cumulative output, but organizations learn at different rates, and an interruption causes knowledge to depreciate. Axes are logarithmic, so a constant learning rate appears as a straight line. Illustrative numbers only.

14.3.4 Evidence and extensions

Learning-curve heterogeneity is now well documented in manufacturing, services, and medicine. Pisano, Bohmer, and Edmondson (2001) found large differences across hospitals in how quickly surgical teams improved after adopting a new minimally invasive cardiac procedure, and A. C. Edmondson, Bohmer, and Pisano (2001) traced the differences to team learning practices, including whether team members felt able to speak up. Research on failure shows that organizations often learn more from failures than from successes: in the global orbital launch vehicle industry, Madsen and Desai (2010) found that knowledge from failures was retained longer than knowledge from successes. Haunschild and Sullivan (2002) found that airlines learned more from accidents and incidents with heterogeneous causes, which force broader search. Argote and Miron-Spektor (2011) offer an integrative framework in which experience interacts with organizational context to create knowledge that is embedded in members, tools, and routines.

DEBATE AND EVIDENCE

The dark side of learning. Learning can be myopic. Levinthal and March (1993) argued that organizations overweight the short run, the near neighbourhood, and success, so that learning itself drives out exploration. B. Levitt and March (1988) described **competency traps**, in which favourable performance with an inferior routine leads the organization to accumulate more experience with it, making a superior routine unattractive to try, and **superstitious learning**, in which the link between action and outcome is misread. Others question whether "organizations" learn at all, or whether the construct is simply the aggregation of individual learning. Most contemporary scholars locate organizational learning in changes to routines,

tools, and transactive memory that persist when individuals leave (Argote and Miron-Spektor 2011).

14.3.5 Applications in business

BUSINESS CASE

Toyota and learning as an operating system. The Toyota Production System is widely studied as an organization designed for continuous single-loop learning on the shop floor. Spear and Bowen (1999) argued that its distinctive feature is not any particular tool but a set of tacit rules that make every activity, connection, and flow a structured hypothesis that can be tested. Problems are surfaced immediately (for example, by stopping the line), and countermeasures are developed by the people closest to the work. Taiichi Ohno's practice of asking "why" repeatedly until a root cause is reached is a disciplined form of problem-oriented search (Ohno 1988). The system also illustrates the limits of transfer: many firms copied visible tools such as kanban cards without the learning routines behind them, and saw smaller gains.

BUSINESS CASE

Learning from failure in safety-critical industries. Commercial aviation's incident reporting systems, which protect reporters who disclose their own near misses, are an institutional design for converting rare and costly failures into shared knowledge. Research on airlines and launch vehicles suggests that such systems work best when they encourage analysis of diverse causes rather than single-cause attributions (Haunschild and Sullivan 2002, Madsen and Desai 2010). Cannon and Edmondson (2005) argue that most organizations fail to learn from failure because of social and technical barriers to identifying, analysing, and deliberately experimenting with failure, and propose a sequence of practices to overcome each barrier.

14.3.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Production units with higher rates of team learning behaviour will show steeper learning curves for a newly introduced product.
- *H2.* Knowledge depreciation (the loss of productivity after a production gap) will be smaller in units with codified routines than in units relying on individual expertise.
- *H3.* Organizations learn more from failures with heterogeneous causes than from failures with a single dominant cause.

Designs and data. Learning-curve regressions of unit cost or quality on cumulative experience with depreciation parameters (Argote and Epple 1990, Darr, Argote, and Epple 1995). Panel data on accidents, recalls, or surgical outcomes. Ethnographies of routines. Field experiments that vary after-action reviews.

Measures. Cumulative output, progress ratio, depreciation rate, transfer across units, frequency of after-action reviews, changes in codified procedures.

COMMON PITFALL

Equating learning with improvement. Organizations can learn the wrong lessons. A measured fall in errors may reflect less reporting rather than fewer errors, and success can teach complacency. Studies should separate changes in knowledge (routines, cognition) from changes in performance and test whether the experience-performance link runs through the proposed knowledge mechanism.

14.4 Knowledge Creation

THEORY CARD

Core claim. Organizations create new knowledge through a continuous dialogue between **tacit** and **explicit** knowledge, captured in four modes of conversion (socialization, externalization, combination, internalization) that spiral from individuals to groups to the whole organization.
Origins. Management and Japanese innovation studies. Nonaka (1994) and Nonaka and Takeuchi (1995), building on Polanyi's concept of tacit knowledge (Polanyi 1966).
Level of analysis. Individual, group, organization, and interorganizational networks.
Key constructs. Tacit and explicit knowledge, SECI modes, knowledge spiral, redundancy, requisite variety, middle-up-down management.
Boundary conditions. Most relevant for product development and knowledge-intensive work in which much know-how cannot be written down. Cultural and organizational context shapes the ease of socialization.
Related theories. Knowledge-based view (Section 4.3), Absorptive capacity (Section 5.2), Organizational learning.

14.4.1 Core logic

Polanyi (1966) observed that people know more than they can tell. Such **tacit knowledge** (the feel of a skilled baker's hands, the judgement of an experienced engineer) is personal, context specific, and hard to formalize. **Explicit knowledge** can be codified in documents, formulas, and databases. Organizational learning theory tended to treat organizations as processors of information. Nonaka (1994) argued instead that the core of innovation is the *creation* of new knowledge, and that it happens when tacit and explicit knowledge are converted into one another. The four modes form the **SECI** model shown in Figure 14.5.

Table 14.5 presents the four modes of knowledge conversion.

Table 14.5. The four modes of knowledge conversion.

Mode	Conversion	Typical practices	Business example
Socialization	Tacit to tacit	Apprenticeship, job rotation, site visits, working side by side	A new engineer shadows a senior field technician
Externalization	Tacit to explicit	Metaphors, analogies, concept papers, prototypes	A design team turns a customer frustration into a product concept
Combination	Explicit to explicit	Reports, specifications, knowledge bases, training manuals	Merging test results into a standard operating procedure

Table 14.5. The four modes of knowledge conversion. *(continued)*

Mode	Conversion	Typical practices	Business example
Internalization	Explicit to tacit	Simulation, practice, learning by doing	Staff practise a new procedure until it becomes routine

Nonaka and Takeuchi (1995) illustrated the model with Japanese product development cases. The best-known is the development of an automatic home bread-making machine at Matsushita in the 1980s. Early prototypes produced poor bread, so a software developer trained alongside the head baker at a leading Osaka hotel to absorb his kneading technique (socialization), and the team then translated what she had learned into a design specification for the machine (externalization). The authors also emphasised organizational conditions that enable the spiral, including intention, autonomy, creative chaos, redundancy of information, and requisite variety, and argued for **middle-up-down** management in which middle managers translate top management's vision into concepts that front-line teams can work with.

DEBATE AND EVIDENCE

Is tacit knowledge convertible at all? The most persistent critique concerns externalization. Following Polanyi, some scholars argue that tacit knowledge is by definition inexpressible, so it cannot be "converted" into explicit form, only supported or demonstrated. Gourlay (2006) questioned the empirical basis of the model and whether the four modes are distinct processes. Nonaka and Krogh (2009) responded by clarifying that tacit and explicit knowledge lie on a continuum and that conversion is a social process of articulation, not a literal transfer. A second critique is cultural: the model was developed from Japanese cases, and its assumptions about long tenure and dense socialization may not transfer to organizations with high turnover and contractual employment.

BUSINESS CASE

Illustrative scenario: capturing field know-how. An industrial equipment company finds that its best service technicians fix faults in half the time of others, but its manuals do not explain why. A knowledge-creation approach would not start with a database. It would pair junior and senior technicians on difficult jobs (socialization), hold structured debriefs in which experts explain their diagnostic hunches through analogies (externalization), consolidate the debriefs into decision trees embedded in the service app (combination), and use simulated faults in training (internalization). The spiral then repeats as the decision trees are tested in the field.

RESEARCH TOOLKIT

Designs and measures. Survey scales for the four SECI modes, case studies of product development, network analysis of knowledge flows, and studies of knowledge codification and its effects on transfer. Hypotheses typically link the intensity of particular modes to innovation outcomes (for example, that externalization practices predict new product concept quality more strongly than combination practices do).

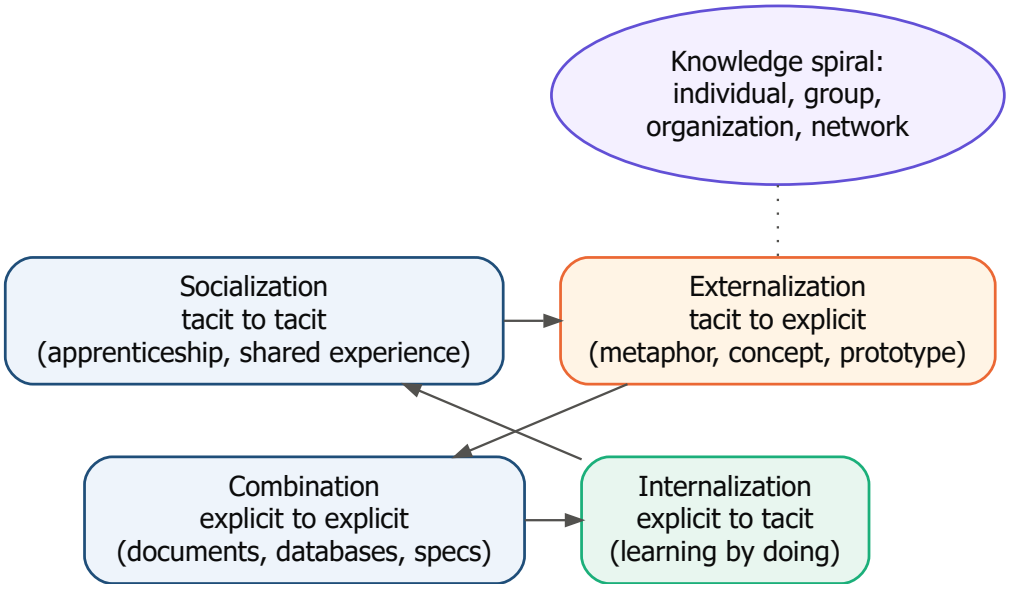


Figure 14.5. The SECI model of knowledge creation. Knowledge moves through four modes of conversion and spirals outward from individuals to groups, the organization, and partners. Based on Nonaka (1994) and Nonaka and Takeuchi (1995).

14.5 Organizational Identity

THEORY CARD

Core claim. Organizational members share beliefs about what is central, enduring, and distinctive about their organization. These beliefs filter how issues are interpreted, which actions feel legitimate, and how members respond to outside images that conflict with them.
Origins. Organization theory. Albert and Whetten (1985), with major empirical work by Dutton and Dukerich (1991) and the social identity foundation of Ashforth and Mael (1989).
Level of analysis. Organization (collective identity) and individual (organizational identification).
Key constructs. Central, enduring, and distinctive attributes, organizational identification, construed external image, reputation, identity threat, adaptive instability.
Boundary conditions. Identity matters most when issues are ambiguous and when members identify strongly. In loosely coupled or newly formed organizations, identity claims may be weakly shared.
Related theories. Social identity theory (Section 22.2), Organizational culture, Legitimacy (Section 7.3), Optimal distinctiveness (Section 22.3).

14.5.1 Core logic

Albert and Whetten (1985) proposed that an organization's identity consists of members' self-referential claims about what is **central** to the organization, what is **enduring** about it over time, and what is **distinctive** compared with similar organizations. Identity questions ("Who are we?") become salient at critical junctures such as founding, loss of a sustaining resource, mergers, rapid

growth, or retrenchment. Whetten (2006) later sharpened the concept, arguing that identity claims are most visible when organizations face "fork in the road" choices that would commit them to one self-definition over another.

At the individual level, Ashforth and Mael (1989) applied social identity theory (Section 22.2) to organizations. **Organizational identification** is the perception of oneness with or belongingness to an organization. When people identify strongly, the organization's successes and failures feel personal, which motivates effort and loyalty but also defensiveness when the organization is criticised (Table 14.6).

Table 14.6. Organizational identity and neighbouring concepts.

Concept	Core question	Whose view	Example measure
Organizational identity	Who do we believe we are?	Members about the organization	Members' ratings of central, enduring, and distinctive attributes
Organizational identification	How much is the organization part of who I am?	Individual member	Identification scale items on oneness with the organization
Construed external image	What do outsiders think of us, in our view?	Members about outsiders' views	Members' beliefs about how the public sees the firm
Reputation	What do outsiders actually think of us?	External stakeholders	Rankings, media tone, stakeholder surveys
Organizational culture	How do we do things here, and why?	Members, often tacit	Value profiles, observed practices, underlying assumptions

14.5.2 Evidence and extensions

Dutton and Dukerich (1991) provided the field's foundational case study. In the 1980s the Port Authority of New York and New Jersey faced a growing number of homeless people in its bus terminal, train stations, and airports. The agency's members saw it as a professional, technically excellent transportation organization, not a social service agency, and they initially treated homelessness as a policing and facilities problem. As media coverage and public criticism damaged the agency's **construed external image**, members felt the gap between how they saw themselves and how they believed outsiders saw them. That discomfort helped push the agency to reframe the issue and invest in more humane and collaborative responses. Identity and image, the authors argued, act as a mirror that shapes how organizations interpret and respond to issues.

Elsbach and Kramer (1996) found a similar dynamic when *Business Week* rankings of business schools threatened members' perceptions of their schools. Members responded by highlighting alternative comparison categories and dimensions on which their school excelled. Gioia, Schultz, and Corley (2000) challenged the assumption that identity is enduring, proposing **adaptive instability**: identity labels may stay constant while their meanings shift as members reinterpret them in light of changing images. Tripsas (2009) showed how a firm's identity as a "digital photography company" made it difficult to pursue opportunities that technology made possible but that did not fit that identity, complementing the cognitive account of Polaroid's difficulties in digital imaging in Tripsas and Gavetti (2000).

> **DEBATE AND EVIDENCE**
>
> **Enduring or fluid?** The original definition required identity to be enduring, but empirical work repeatedly shows identity claims changing in response to image, strategy, and leader-

ship. Social-actor perspectives treat identity as institutional claims that change slowly. Social-constructionist perspectives treat identity as shared understandings that are continually renegotiated. Whetten (2006) defends a narrower definition to keep the concept distinct from culture and image. A second debate concerns measurement: identity is easily conflated with culture or reputation, and many studies infer identity from self-presentation materials that may be impression management rather than belief.

BUSINESS CASE

Identity in strategic change. Firms that redefine their business often discover that identity, not capability, is the binding constraint. When a company that has seen itself as a hardware manufacturer moves to services or software subscriptions, engineers may resist work that feels peripheral to "who we are". Effective leaders in such transitions tend to reinterpret existing identity labels rather than discard them (for example, moving from "we build the best machines" to "we keep our customers' operations running"), which is the adaptive instability that Gioia, Schultz, and Corley (2000) describe.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms whose stated identity claims are inconsistent with a new strategic initiative will adopt it more slowly, controlling for capabilities.
- *H2.* Negative media coverage that conflicts with members' identity beliefs will trigger more issue-reframing activity than negative coverage that is consistent with those beliefs.

Designs and data. Longitudinal case studies with interviews and archival documents. Text analysis of letters to shareholders, mission statements, and internal communications. Survey measures of organizational identification. Event studies of identity threats such as rankings or scandals.

COMMON PITFALL

Treating the mission statement as the identity. Formal statements are identity *claims*, often written for external audiences. Identity theory concerns what members actually believe is central and distinctive. Valid studies triangulate official claims with members' own accounts and with choices made at identity-relevant forks in the road.

14.6 Organizational Culture

THEORY CARD

Core claim. Organizations develop shared, largely taken-for-granted assumptions, learned as the group solved problems of external adaptation and internal integration. These assumptions are expressed in espoused values and visible artifacts, and they shape behaviour more powerfully than formal structures.

Origins. Organizational psychology and anthropology. Schein (2010) (first edition 1985) and

Schein (1990). Quantitative measurement through the Organizational Culture Profile of C. A. O'Reilly, Chatman, and Caldwell (1991).
Level of analysis. Organization and subgroups (functions, units, professions).
Key constructs. Artifacts, espoused beliefs and values, basic underlying assumptions, culture strength (intensity and consensus), culture content, subcultures, person-culture fit.
Boundary conditions. Culture is most consequential where formal controls are incomplete and behaviour cannot be fully specified. Culture strength is valuable in stable environments and can become a liability when the environment shifts.
Related theories. Organizational identity, Attraction-selection-attrition, Hofstede's cultural dimensions (Section 18.2), Tight and loose cultures (Section 18.7), Institutional theory (Section 7.2).

14.6.1 Core logic: Schein's three levels

Schein (2010) defines culture as a pattern of shared basic assumptions that a group has learned as it solved its problems of adapting to its environment and integrating its members, which has worked well enough to be considered valid and is therefore taught to new members as the correct way to perceive, think, and feel. The definition has two important implications. Culture is **learned history**, not management decree, and it is **deep**: its most important layer is taken for granted and rarely discussed (Figure 14.6).

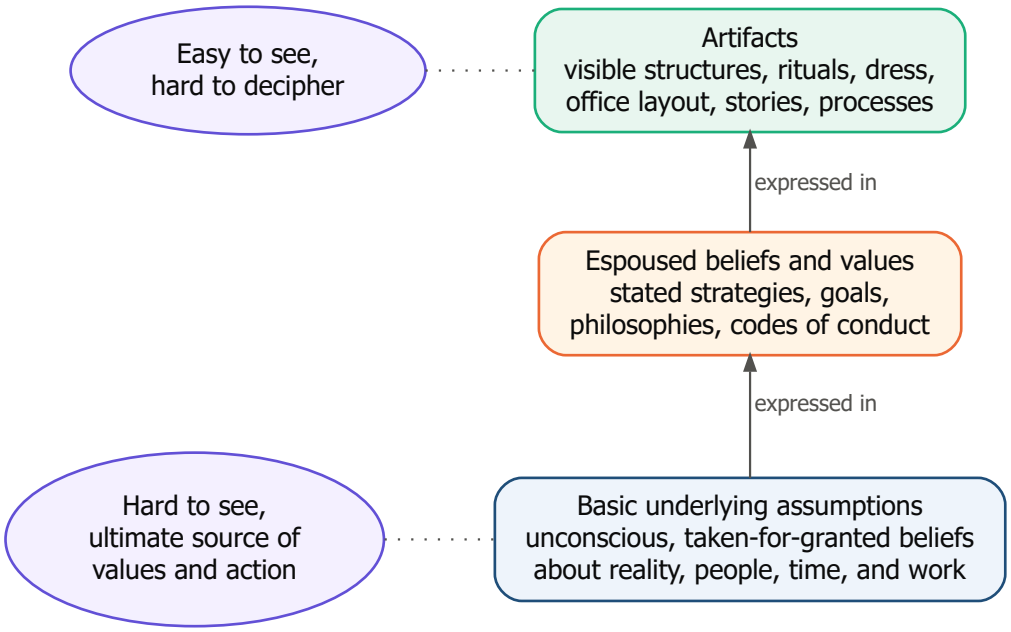


Figure 14.6. Schein's three levels of organizational culture. Artifacts are easy to observe but hard to interpret. Basic assumptions are hard to observe but explain the levels above them. Based on Schein (2010).

Because artifacts are ambiguous, an outsider who sees an open-plan office cannot know whether it reflects a belief in egalitarian collaboration or simply a cost reduction. Espoused values may also diverge from the assumptions actually in use, the same gap Argyris and Schön identified between espoused theories and theories in use (Argyris and Schön 1978). Schein therefore argues that cul-

ture can be diagnosed only by probing the assumptions that make artifacts and values coherent, usually through iterative interviews with insiders.

14.6.2 Measuring culture: the Organizational Culture Profile

Quantitative research needed a way to compare cultures and to assess fit between people and organizations. C. A. O'Reilly, Chatman, and Caldwell (1991) developed the **Organizational Culture Profile (OCP)**, in which respondents sort 54 value statements according to how characteristic they are of an organization, or how desirable they are to the individual. Comparing the two profiles yields a measure of person-organization fit. Factor analyses revealed dimensions such as innovation, stability, respect for people, outcome orientation, attention to detail, team orientation, and aggressiveness. In the original studies, person-culture fit predicted later job satisfaction, organizational commitment (Section 15.10), and turnover. Chatman and O'Reilly (2016) argue that the field should distinguish clearly between culture **content** (which norms) and culture **strength** (how widely shared and intensely held), and should connect culture explicitly to strategy.

14.6.3 Evidence and extensions

The most cited quantitative finding concerns culture strength. Sørensen (2002) showed that firms with strong cultures have more reliable (less variable) performance in relatively stable environments, but that the advantage diminishes in volatile environments, where strong cultures make it harder to detect and respond to change. This finding connects culture to the exploration and exploitation (Section 5.3) tension: strong culture supports exploitation and can impede exploration.

DEBATE AND EVIDENCE

Can culture be managed? Schein treats leaders as the primary creators and embedders of culture, especially founders. Critics from interpretive and critical traditions argue that culture is emergent, contested, and fragmented into subcultures, so that "managing culture" often means managing artifacts and rhetoric rather than assumptions. A related measurement debate concerns whether survey instruments capture culture or only climate (shared perceptions of policies and practices). The current consensus is that both qualitative diagnosis and quantitative profiles are needed, and that norms are best measured by what is rewarded and sanctioned in practice (Chatman and O'Reilly 2016).

14.6.4 Applications in business

BUSINESS CASE

Culture integration in mergers. Cultural incompatibility is a common explanation for disappointing mergers. The 1998 combination of Daimler-Benz and Chrysler, announced as a merger of equals, is a frequently cited public example: reports of friction between German and American management styles and decision-making norms accompanied the combined firm's difficulties, and Daimler sold a majority of Chrysler in 2007. Because many factors affect merger outcomes, case evidence alone cannot isolate culture. R. A. Weber and Camerer (2003) therefore created cultures in the laboratory. Groups developed their own shorthand codes for describing pictures of office scenes, and then two groups were merged. Performance fell after the merger, participants overestimated how well the merged group would perform, and they

tended to attribute the decline to members of the other group. The experiment shows that even trivially created cultures generate coordination costs and attribution conflict when combined. Practical implications include culture due diligence before a deal, explicit integration of decision rights and vocabulary, and patience about early performance dips.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Acquisitions in which the acquirer's and target's culture profiles differ more will show larger post-merger turnover among target executives.
- *H2.* Culture strength will reduce performance variability more in stable than in volatile industries (Sørensen 2002).

Data and measures. The OCP and its variants, employer review text (analysed with dictionary or embedding methods for culture content), ethnography, interviews probing assumptions, and archival indicators of rewarded behaviour.

COMMON PITFALL

Mistaking espoused values for culture. Values posted on walls and websites are espoused values. Culture lives in the assumptions that predict what people do when values conflict or nobody is watching. A study that measures only stated values cannot test claims about culture's effects on behaviour.

14.7 Paradox Theory

THEORY CARD

Core claim. Organizations face persistent contradictory yet interrelated demands (for example, exploration and exploitation, control and flexibility, global and local). Sustainable performance comes not from choosing one pole but from accepting the tension and continually attending to both over time, a dynamic equilibrium.
Origins. Organization theory. W. K. Smith and Lewis (2011), building on Lewis (2000) and reviewed by Schad et al. (2016).
Level of analysis. Individual (leaders' cognition), team, and organization.
Key constructs. Latent and salient tensions, paradox categories (learning, belonging, organizing, performing), acceptance, resolution through splitting and integration, virtuous and vicious cycles, paradoxical cognition.
Boundary conditions. Applies to tensions that are persistent and interdependent. Simple trade-offs that can be resolved once by a choice are dilemmas, not paradoxes.
Related theories. Ambidexterity (Section 5.4), Exploration and exploitation (Section 5.3), Contingency theory (Section 7.6).

Contingency theory asks which pole of a tension fits a given environment. Paradox theory asks how organizations can live with both. W. K. Smith and Lewis (2011) define **paradox** as contradictory yet interrelated elements that exist simultaneously and persist over time. Such tensions are **latent**, embedded in organizing, and become **salient** under conditions of **plurality** (multiple stakeholders and views), **change**, and **scarcity** of resources. Responses that suppress one pole (defensiveness,

splitting without reconnection) produce **vicious cycles**. Responses that combine **acceptance** of the tension with periodic, iterative **resolution** produce **virtuous cycles** and long-term sustainability (Table 14.7).

Table 14.7. Four categories of organizational paradox, following Smith and Lewis (2011).

Paradox category	Underlying tension	Organizational example
Learning	Building on the past versus destroying it to create the future	Investing in the core product while cannibalising it with a new platform
Belonging	Identity and affiliation: the individual versus the collective, one group versus another	Professional identities of physicians versus hospital-wide standardisation
Organizing	Structuring and leading: control versus flexibility, collaboration versus competition	Centralised compliance controls alongside autonomous product teams
Performing	Competing goals of different stakeholders	Social mission versus financial returns in a social enterprise

DEBATE AND EVIDENCE

Theory or vocabulary? Paradox theory has spread rapidly, and Schad et al. (2016) warn that the term is often used loosely for any tension or trade-off. Critics argue that "accept and attend to both" can become unfalsifiable advice. The research frontier focuses on precise constructs (for example, paradoxical cognition and paradox mindset at the individual level), on the conditions under which acceptance improves performance, and on multilevel dynamics of how tensions at one level are managed at another.

BUSINESS CASE

Illustrative scenario: the social enterprise. A microfinance organization must serve poor borrowers (social mission) while remaining financially self-sustaining (commercial logic). If leaders let the commercial pole dominate, mission drift follows. If the social pole dominates, the organization runs out of capital. A paradox approach would separate the poles structurally in some respects (distinct metrics and governance reviews for each), integrate them elsewhere (a board that deliberates on both at once), and revisit the balance regularly rather than fixing it once.

RESEARCH TOOLKIT

Designs and measures. Longitudinal process studies of how leaders oscillate between poles, survey measures of paradox mindset, and comparative cases of hybrid organizations. A testable hypothesis: leaders with higher paradoxical cognition will sustain investment in both exploratory and exploitative projects for longer under resource scarcity.

14.8 Psychological Safety

THEORY CARD

Core claim. Team members learn, speak up, and report errors when they share a belief that the team is safe for interpersonal risk taking. Psychological safety enables learning behaviour, which in turn improves team performance in uncertain, interdependent work.
Origins. Organizational behaviour. A. Edmondson (1999), drawing on Kahn (1990) and earlier work on organizational change. Reviewed by A. C. Edmondson and Lei (2014) and meta-analysed by Frazier et al. (2017).
Level of analysis. Team (shared climate), with individual-level variants.
Key constructs. Psychological safety, team learning behaviour (asking questions, seeking feedback, discussing errors, experimenting), voice, leader inclusiveness, status differences, accountability.
Boundary conditions. Matters most when work is uncertain, interdependent, and learning-intensive. In routine, independent work, its effects on performance are weaker.
Related theories. Organizational learning, Leader-member exchange (Section 15.5), Spiral of silence (Section 22.8), Groupthink (Section 30.6), Transformational leadership (Section 15.6).

14.8.1 Origins and intellectual history

The research programme started with a puzzle. In a study of hospital nursing units, A. C. Edmondson (1996) expected that better-led, better-functioning units would make fewer medication errors. Instead, they appeared to make *more*. The explanation was that these units were more willing to report errors, while in units with punitive managers errors went unreported. Detected error rates partly measured the climate for speaking up, not only the underlying frequency of mistakes.

A. Edmondson (1999) developed the construct of **team psychological safety**, a shared belief that the team is safe for interpersonal risk taking, and distinguished it from trust (which concerns one party's expectations of another) and from cohesiveness (which can actually suppress dissent). Kahn (1990) had already identified psychological safety as one of three conditions, together with meaningfulness and availability, under which people bring their full selves to work. In a multimethod field study of 51 work teams in a manufacturing company, A. Edmondson (1999) found that psychological safety was associated with **team learning behaviour**, which in turn was associated with team performance.

14.8.2 Mechanism

Figure 14.7 shows the model of team learning.

Psychological safety is not the same as low standards. A. C. Edmondson (2018) crosses psychological safety with performance standards to describe four zones (Figure 14.8). High safety with low standards produces a **comfort zone**. Low safety with high standards produces an **anxiety zone**, where people work hard but hide problems. The **learning zone** combines both.

14.8.3 Evidence and extensions

The construct has a large and broadly consistent evidence base. Frazier et al. (2017) meta-analysed research on psychological safety and found it positively related to information sharing, learning be-

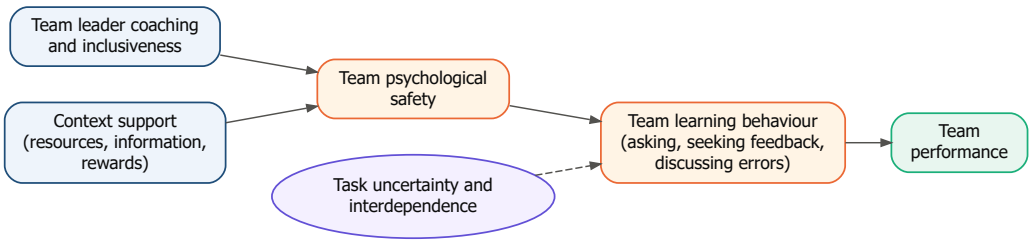


Figure 14.7. Model of team learning. Team leader coaching and context support foster psychological safety, which enables learning behaviour and, through it, team performance. Based on Edmondson (1999).

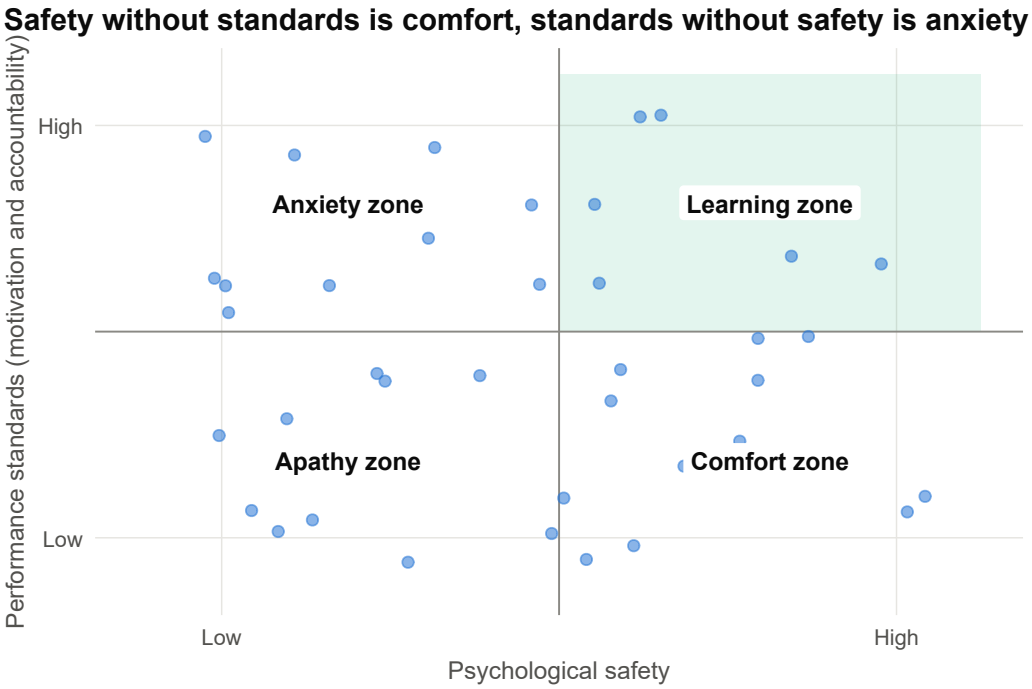


Figure 14.8. Psychological safety and performance standards: four zones. Stylised teams are plotted for illustration. Framework based on Edmondson (2018).

haviour, engagement, and task performance, with supportive leadership and peer relationships among its antecedents. In healthcare, Nembhard and Edmondson (2006) studied neonatal intensive care unit teams and found that lower-status professionals felt less psychologically safe than physicians, but that **leader inclusiveness** (words and deeds by leaders that invite and appreciate others' contributions) raised psychological safety and weakened the status gap. A. C. Edmondson, Bohmer, and Pisano (2001) showed that surgical teams that successfully adopted a new cardiac procedure were those whose leaders framed the change as a team learning challenge and made it safe for nurses and technicians to speak up. A. C. Edmondson and Lei (2014) review the construct's history and its extension from teams to organizations and to topics such as innovation, voice, and safety.

DEBATE AND EVIDENCE

Climate, trait, or fad? Three critiques recur. First, measurement: the standard seven-item scale is a perception measure, and aggregating it to the team level requires evidence of within-team agreement. Second, causality: most studies are cross-sectional, so high-performing teams may feel safe because they perform well. Third, popularization: in practitioner discourse, psychological safety is sometimes equated with comfort or the absence of criticism, which is close to the opposite of the construct. Edmondson's own framing pairs safety with high standards and candour (A. C. Edmondson 2018).

14.8.4 Applications in business

BUSINESS CASE

Hospitals and tech teams. Healthcare offers the clearest evidence that speaking up saves lives and money. Checklists, structured briefings, and practices that let any team member stop a procedure are institutional designs for psychological safety, and the research above shows that leader behaviour determines whether they are used (Nembhard and Edmondson 2006, A. C. Edmondson, Bohmer, and Pisano 2001). In technology, Google's internal research project on team effectiveness, known as Project Aristotle, became widely publicised in 2015 and 2016. Google reported that psychological safety was the most important of five dynamics that distinguished effective teams, alongside dependability, structure and clarity, meaning, and impact (Duhigg 2016). Blameless post-incident reviews, common in site reliability engineering, apply the same logic: they focus on systemic causes so that engineers disclose what actually happened.

14.8.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Team psychological safety will be positively related to the number of near misses reported, controlling for workload.
- *H2.* Leader inclusiveness will weaken the negative relationship between professional status and individual psychological safety (Nembhard and Edmondson 2006).
- *H3.* The effect of psychological safety on team performance will be stronger for teams with high task uncertainty.

Measures. Edmondson's team psychological safety scale aggregated to the team level after checking agreement indices, behavioural measures of voice and error reporting, and observational coding of meetings.

Designs. Multilevel field surveys, longitudinal panels to address reverse causality, and field experiments with leader training or structured debriefs.

COMMON PITFALL

Confusing safety with comfort. Psychological safety is about the freedom to take interpersonal risks such as admitting mistakes and challenging ideas, not about being shielded from

disagreement or accountability. Interventions that reduce candour in the name of "safety" undermine the mechanism the theory describes.

14.9 Attraction-Selection-Attrition

THEORY CARD

Core claim. People are attracted to organizations whose members resemble them, are selected by organizations that prefer people who fit, and leave when they do not fit. Over time the cycle makes organizations increasingly homogeneous, and that homogeneity, not formal structure, defines the place.
Origins. Industrial and organizational psychology. Schneider (1987), updated in Schneider, Goldstein, and Smith (1995).
Level of analysis. Individual and organization.
Key constructs. Attraction, selection, attrition, homogeneity, person-organization fit, founder influence.
Boundary conditions. Requires voluntary labour mobility and discretion in selection. Weaker in tight labour markets or under strict diversity mandates.
Related theories. Person-environment fit (Section 15.11), Organizational culture, Homophily (Section 30.2).

Schneider (1987) argued, in a paper titled "The people make the place", that organizational behaviour should be explained by the characteristics of the people in the organization rather than by structures and technologies alone. Founders' personalities shape early goals and hiring. Applicants who share those values find the organization attractive. Recruiters select people who seem to fit. Employees who do not fit leave. The **homogeneity hypothesis** follows: organizations become more homogeneous in personality and values over time. This helps explain why cultures persist (Schneider, Goldstein, and Smith 1995) and links the theory to person-organization fit research (Chatman 1989, Kristof 1996, C. A. O'Reilly, Chatman, and Caldwell 1991). Figure 14.9 simulates the process.

DEBATE AND EVIDENCE

Homogeneity: asset or liability? Homogeneity eases coordination and supports a strong culture, but Schneider himself warned that it can make organizations unable to perceive or adapt to environmental change. Empirical support for organization-level homogeneity in personality is moderate, and critics note that job and occupational requirements also produce similarity. The theory also sits uneasily with diversity goals: "cultural fit" in hiring can become a vehicle for similarity bias.

BUSINESS CASE

Hiring for culture fit or culture add. Many technology firms once screened explicitly for "culture fit", and some have publicly reframed the criterion as "culture add" or values alignment, precisely to avoid the homogenising spiral that ASA predicts. The practical design question is which attributes should be selected for fit (for example, integrity and core values) and which

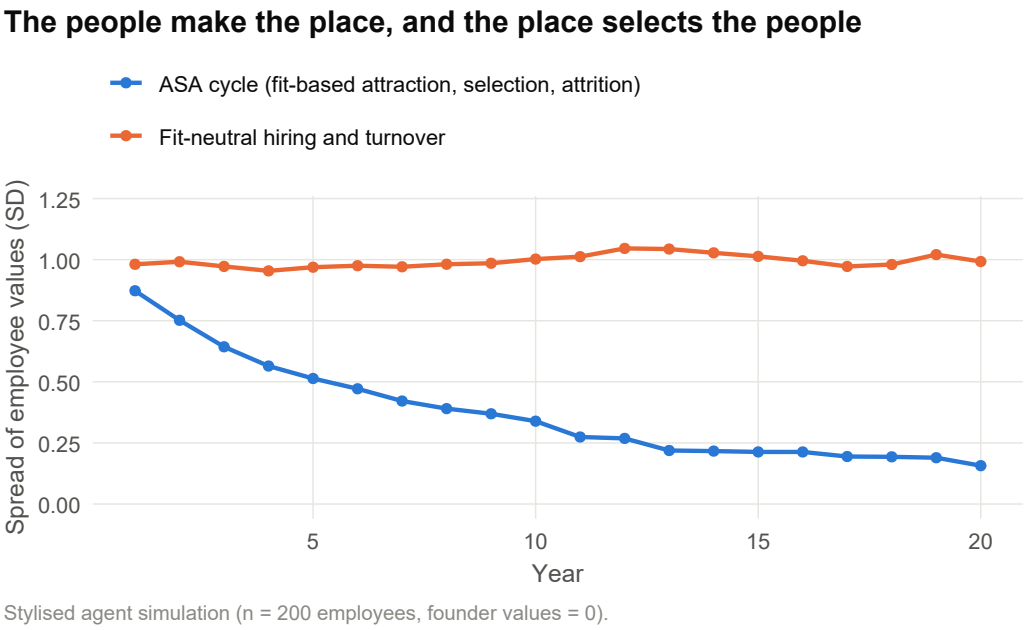


Figure 14.9. Illustrative simulation of the attraction-selection-attrition cycle. Under ASA, applicants and leavers are filtered by distance from the founder's values and the spread of values shrinks. With fit-neutral hiring, diversity is maintained. Stylised data.

should be deliberately diversified (for example, backgrounds, cognitive styles, and functional expertise).

RESEARCH TOOLKIT

Designs and measures. Multi-wave data on applicants, hires, and leavers with personality or values measures. Within-organization variance over time as the outcome. A sample hypothesis: the variance of employee values declines with organizational age, and the decline is steeper when founders retain control of hiring.

14.10 Social Information Processing

THEORY CARD

Core claim. Job attitudes and perceptions of tasks are not simple reactions to objective job features. They are constructed from social information: what coworkers say, how supervisors frame the work, and what people infer from their own past behaviour.

Origins. Organizational behaviour. Salancik and Pfeffer (1978), as a critique of need-satisfaction and job design models.

Level of analysis. Individual within a workgroup.

Key constructs. Social cues, salience of information, commitment and retrospective rationalization, socially constructed task perceptions.

Boundary conditions. Social cues matter most when tasks are ambiguous, when newcomers

are socialized, and when sources are credible and similar.
Related theories. Job characteristics model (Section 15.1), Self-perception theory (Section 21.2), Social norms (Section 22.5), Conformity (Section 22.4), Sensemaking.

Salancik and Pfeffer (1978) challenged the dominant view that workers have stable needs and evaluate jobs by how well objective characteristics satisfy those needs. They proposed that the social environment provides cues about which aspects of a job are important and how to evaluate them, and that people also infer their attitudes from their own behaviour. A job is "interesting" partly because colleagues say it is. Griffin (1983) tested the idea in a field experiment in which supervisors provided social cues about task characteristics, and found that social information as well as objective changes in tasks affected workers' perceptions of their jobs. Thomas and Griffin (1983) reviewed early evidence and concluded that social information has real effects on task perceptions, while objective task characteristics also continue to matter.

DEBATE AND EVIDENCE

Complement or competitor? The theory was framed as an alternative to the job characteristics model, but most researchers now treat the two as complementary: objective features and social cues jointly shape perceptions. Critics also note that SIP is difficult to falsify unless researchers specify which cues should dominate and when.

BUSINESS CASE

Onboarding and remote work. New hires form their view of a role in the first weeks, largely from what peers and managers communicate. SIP implies that onboarding buddies, team rituals, and managers' framing of work as meaningful shape engagement as much as the job design itself. In remote or hybrid settings, fewer informal cues are available, which may make the few explicit cues, such as how leaders talk about the work in meetings, disproportionately influential.

RESEARCH TOOLKIT

Designs. Field experiments manipulating supervisor or peer cues, network studies of attitude contagion within workgroups, and longitudinal studies of newcomers. A sample hypothesis: newcomers' job satisfaction will converge toward the average satisfaction of their immediate network contacts, more strongly when the task is ambiguous.

14.11 Sensemaking

THEORY CARD

Core claim. People in organizations do not simply perceive an objective environment. When expectations are violated, they extract cues, build plausible stories about what is going on, and act on those stories, and their actions partly create the environment they then interpret.
Origins. Social psychology of organizing. Weick (1979), Weick (1993), and Weick (1995a). Updated by Weick, Sutcliffe, and Obstfeld (2005) and reviewed by Maitlis and Christianson (2014).

Level of analysis. Individual, group, and organization.
Key constructs. Ambiguity and equivocality, triggers, extracted cues, plausibility, enactment, retrospection, identity, sensegiving, sensebreaking, cosmology episodes.
Boundary conditions. Most visible during crises, change, and novel situations, when routine interpretations fail. In stable settings, sensemaking runs in the background of habitual action.
Related theories. Social information processing, Organizational identity, Organizational learning, Attribution theory (Section 21.3).

14.11.1 Origins and core logic

Weick's organization theory starts with the claim that organizing is the process of reducing **equivocality**, the existence of multiple, conflicting interpretations (Weick 1979). **Sensemaking** is the process by which people give meaning to experience. Maitlis and Christianson (2014) define it as the process through which people work to understand issues or events that are novel, ambiguous, confusing, or in some other way violate expectations. In Weick (1995a), Weick described seven properties of sensemaking, summarised in Table 14.8.

Table 14.8. Seven properties of sensemaking, following Weick (1995).

Property	Meaning	Organizational illustration
Grounded in identity construction	Who I think I am shapes what I notice and how I interpret it	Engineers and managers read the same data through different professional identities
Retrospective	Meaning is assigned to events after they have happened	Post-mortems construct the story of a project's failure
Enactive of sensible environments	Action creates part of the environment that is then interpreted	A firm that cuts prices creates the price war it then perceives
Social	Sensemaking is shaped by actual, imagined, or implied others	Teams converge on a shared account in meetings
Ongoing	Sensemaking never starts or stops, it is interrupted	A sudden outage interrupts routine work and demands interpretation
Focused on and by extracted cues	Small, familiar cues are elaborated into larger meanings	One customer complaint is read as a sign of a broader trend
Driven by plausibility rather than accuracy	People act on stories that are good enough, not on full information	Managers act on a coherent narrative before all data are in

Weick, Sutcliffe, and Obstfeld (2005) emphasised that sensemaking is about the interplay of action and interpretation. It answers two questions: what is going on here, and what should I do next? It begins with **noticing and bracketing** cues from the flow of experience, proceeds through **labelling** and categorizing, and is completed in action, which generates new cues. Leaders also engage in **sensegiving**, attempts to influence how others make sense. In their study of a new university president launching strategic change, Gioia and Chittipeddi (1991) showed how the president alternated between understanding the institution (sensemaking) and articulating a vision for it (sensegiving), while stakeholders in turn made sense of his vision and gave feedback (Figure 14.10).

14.11.2 Evidence: Mann Gulch and the collapse of sensemaking

Weick (1993) reanalysed the 1949 Mann Gulch fire in Montana, drawing on Norman Maclean's account, in which thirteen smokejumpers died. The crew had been told to expect a routine fire. As

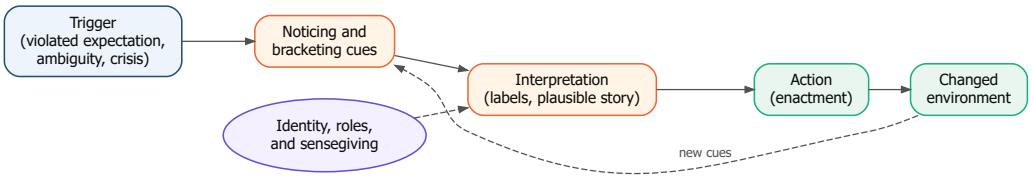


Figure 14.10. The sensemaking cycle. A violated expectation triggers noticing and bracketing of cues, which are interpreted into a plausible account and enacted. Action changes the environment and produces new cues. Based on Weick (1995) and Weick, Sutcliffe and Obstfeld (2005).

the fire behaved in ways that contradicted that expectation, their shared interpretation and their role structure collapsed together. The foreman told the crew to drop their heavy tools, then lit an escape fire and called on them to join him in the burned area, actions that made little sense within the crew's framework and that the crew did not follow. Weick called this a **cosmology episode**: a moment when people suddenly feel that the universe is no longer a rational, orderly system, and lose both the sense of what is happening and the means to rebuild it. He identified sources of **resilience**, including improvisation, virtual role systems (each person able to mentally simulate the whole group), respectful interaction, and an attitude of wisdom that combines confidence and doubt. Related work on aircraft carrier flight decks introduced **collective mind**, the heedful interrelating of individual actions that allows complex systems to operate reliably (Weick and Roberts 1993).

Maitlis and Christianson (2014) review a large literature that applies sensemaking to strategic change, crises, innovation, and identity. Their review notes three recurrent features: sensemaking is triggered by violated expectations, it is social and often contested, and it shapes action that alters the environment.

DEBATE AND EVIDENCE

Process richness versus predictive power. Sensemaking studies are mostly qualitative and retrospective, and critics argue that the perspective explains events well after the fact but rarely generates testable predictions. The emphasis on plausibility over accuracy also invites a normative question: when is a plausible but wrong story dangerous? Proponents respond that the theory's value lies in directing attention to interpretive processes that variance models miss, and recent work has developed more specific constructs (for example, triggers, sensegiving, and prospective sensemaking) that are measurable in field settings.

14.11.3 Applications in business

BUSINESS CASE

NASA's *Challenger* and *Columbia* accidents. The two space shuttle losses are among the most thoroughly documented organizational failures, and they illustrate several theories in this chapter at once. In her study of the *Challenger* launch decision, Vaughan (1996) argued that the disaster resulted not from rule breaking but from the **normalization of deviance**: over many flights, evidence of erosion in the solid rocket booster O-ring seals was interpreted, step by step and through accepted engineering procedures, as an acceptable risk. Each successful flight with some erosion became evidence that erosion was tolerable, a sensemaking

process in which plausibility and precedent outweighed alarm. Starbuck and Milliken (1988) described the same pattern as fine-tuning: past successes encourage organizations to reduce safety margins until something breaks, a form of the superstitious and myopic learning discussed under organizational learning.

After foam debris struck *Columbia*'s left wing during its January 2003 launch, engineers' requests for imagery of the wing did not lead to action, and foam strikes, which had occurred on earlier flights without catastrophe, were treated as a maintenance issue rather than a safety-of-flight issue. The Columbia Accident Investigation Board concluded that NASA's organizational causes were as important as the physical cause and drew explicit parallels with *Challenger*, including schedule pressure, a culture in which dissenting engineering views struggled to reach decision makers, and the reinterpretation of repeated anomalies as normal (Columbia Accident Investigation Board 2003). Read through this chapter's theories, the accidents combine normalized interpretations (sensemaking), learning from success (organizational learning), barriers to raising concerns (psychological safety), schedule commitments that made delay costly (escalation), and a strong culture whose assumptions went unexamined (organizational culture).

14.11.4 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Following an unexpected performance shock, top management teams whose members have more heterogeneous functional backgrounds will generate more distinct interpretations of the shock in internal communications.
- *H2.* Leader sensegiving that links a change initiative to existing identity labels will increase middle managers' support for the change.

Designs and data. Longitudinal case studies with real-time observation, crisis communications, meeting transcripts and earnings call text, interviews using critical incident techniques, and experimental manipulations of ambiguous cues.

Measures. Coded interpretations (threat or opportunity framing), number and diversity of accounts, timing of convergence on a shared account, and actions taken.

COMMON PITFALL

Using sensemaking as a label for any interpretation. Sensemaking is a specific process triggered by interruptions and resolved through action. Studies that simply describe how people interpret events, without tracing triggers, cues, plausibility, and enactment, use the vocabulary without the theory.

14.12 Chapter Summary

KEY TAKEAWAYS

- Organizational decisions often depart from the rational sequence. In **organized anarchies**, choices emerge from the timing of problems, solutions, and participants (M. D. Cohen, March, and Olsen 1972), and once committed, decision makers tend to **escalate** investments in failing courses of action (Staw 1976, Sleesman et al. 2012).
- **Organizational learning** encodes experience into routines. Single-loop learning refines them, double-loop learning questions their assumptions, and learning curves vary, depreciate, and transfer imperfectly (Argyris and Schön 1978, B. Levitt and March 1988, Argote and Epple 1990).
- **Knowledge creation** depends on converting tacit and explicit knowledge through socialization, externalization, combination, and internalization (Nonaka 1994).
- **Identity** and **culture** are deep, shared understandings that filter interpretation and action. Identity answers "who are we?", culture answers "how do we do things here and why?" (Albert and Whetten 1985, Schein 2010).
- **Paradox theory** argues that persistent tensions should be accepted and managed dynamically rather than resolved once (W. K. Smith and Lewis 2011).
- **Psychological safety** enables team learning, especially in uncertain, interdependent work, and is distinct from comfort (A. Edmondson 1999).
- **ASA** and **social information processing** explain how the composition and social context of the workforce shape attitudes and cultures (Schneider 1987, Salancik and Pfeffer 1978).
- **Sensemaking** shows how people construct the situations they act on, and how plausible but wrong interpretations can contribute to disaster (Weick 1993, 1995a).

Table 14.9 presents the comparison of the chapter's theories.

Table 14.9. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Typical business use
Garbage can model	Temporal coincidence of streams	Many choices made by oversight or flight under load	Agenda design, policy windows
Escalation of commitment	Self-justification, sunk cost reasoning	Responsible decision makers reinvest after losses	Stage gates, kill criteria, project reviews
Organizational learning	Encoding experience into routines	Unit costs fall with experience at varying rates	Continuous improvement, after-action reviews
Knowledge creation	Tacit and explicit conversion	Innovation follows the SECI spiral	Product development, knowledge management
Organizational identity	Central, enduring, distinctive self-beliefs	Identity-image gaps trigger issue reframing	Strategic change, reputation management

Table 14.9. Comparison of the chapter's theories. *(continued)*

Theory	Key mechanism	Signature prediction	Typical business use
Organizational culture	Shared basic assumptions	Strong cultures give reliable performance in stable settings	Merger integration, culture diagnosis
Paradox theory	Acceptance plus dynamic resolution	Virtuous cycles sustain contradictory goals	Hybrid organizations, ambidexterity
Psychological safety	Freedom to take interpersonal risks	Safety predicts learning behaviour and error reporting	Team leadership, healthcare, incident reviews
Attraction-selection-attrition	Fit-based entry and exit	Organizations become homogeneous over time	Hiring strategy, diversity
Social information processing	Social cues shape attitudes	Attitudes converge within workgroups	Onboarding, remote work
Sensemaking	Plausible stories enacted in action	Interruptions trigger interpretation and action cycles	Crisis management, change leadership

14.13 Discussion Questions

1. Choose a recent major decision in an organization you know. Reconstruct it once as a rational sequence and once as a garbage can process. Which account fits the evidence better, and what evidence would discriminate between them?
2. Design a governance process for a multi-year IT programme that minimizes escalation of commitment without causing premature abandonment of promising but slow projects.
3. Single-loop learning made Toyota's shop floor exceptionally productive. Under what conditions might the same strength impede double-loop learning?
4. How would you distinguish empirically between organizational identity and organizational culture in a study of a post-merger integration?
5. A hospital reports a sharp rise in recorded medication errors after a new nurse manager arrives. Using psychological safety theory, propose two competing explanations and a research design to test them.
6. Revisit the *Challenger* and *Columbia* accidents. Which theory in this chapter best explains each accident, and what single organizational change would each theory recommend?

14.14 Further Reading

- M. D. Cohen, March, and Olsen (1972): the original garbage can article, still the most provocative statement of decision making under ambiguity.
- Staw (1981) and Sleesman et al. (2012): the classic theoretical framing of escalation and the meta-analytic synthesis of its determinants.
- B. Levitt and March (1988) and Argote and Miron-Spektor (2011): two complementary reviews of organizational learning, one behavioural, one focused on experience and knowledge.
- Nonaka (1994): the canonical statement of knowledge-creation theory and the SECI model.
- Schein (2010): the standard treatment of culture's levels and of leaders' roles in creating and

changing culture.

- A. Edmondson (1999) and A. C. Edmondson and Lei (2014): the founding study and the authoritative review of psychological safety.
- Weick (1993): a short, vivid analysis of how sensemaking and organization collapse together, and how resilience can be built.
- Vaughan (1996): a landmark organizational analysis of the *Challenger* decision and the normalization of deviance.

Chapter 15

Work, Jobs, and Leadership

Beginning in late 2010, a large Chinese travel agency ran an unusual experiment. Call-centre employees who volunteered to work from home were randomly assigned either to do so or to stay in the office. The home workers became more productive and quit less often, yet a large share of them later chose to come back to the office, citing isolation (Bloom et al. 2015). Around the same time, managers everywhere were asking related questions. Why does one team thrive under a demanding manager while another burns out? Why do employees accept a pay cut with grace in one plant and retaliate in another? Why is the chief executive paid several times more than the vice presidents who report to her, even when their jobs look similar?

These puzzles share a theme: the *design* of work and of the relationships around it shapes what people do, feel, and contribute. This chapter assembles the theories that organizational psychologists and personnel economists use to explain those effects. We start with the **job** itself (the job characteristics model, job demands-resources theory, conservation of resources theory, and affective events theory). We then turn to **leadership** (leader-member exchange, transformational leadership, and path-goal theory) and to the **employment relationship** (psychological contracts, organizational justice, organizational commitment, and person-environment fit). We close with two economic theories of **investment and rewards**: human capital theory and tournament theory. Together they form the analytical core of human resource management, people analytics, and leadership development (Table 15.1).

Table 15.1. Chapter at a glance: theories of work, jobs, and leadership.

Theory	Core question	Level of analysis	Key source
Job characteristics model	Which features of a job make it motivating?	Individual / job	Hackman & Oldham (1976)
Job demands-resources	What drives burnout and engagement?	Individual / job	Demerouti et al. (2001)
Conservation of resources	Why is resource loss so stressful?	Individual	Hobfoll (1989)
Affective events theory	How do daily events shape feelings and behaviour at work?	Individual (within person)	Weiss & Cropanzano (1996)
Leader-member exchange	Why do leaders build different relationships with different followers?	Dyad	Graen & Uhl-Bien (1995)
Transformational leadership	How do leaders inspire performance beyond expectations?	Individual / group	Bass (1985)
Path-goal theory	Which leader behaviour fits which situation?	Dyad / group	House (1971)

Table 15.1. Chapter at a glance: theories of work, jobs, and leadership. *(continued)*

Theory	Core question	Level of analysis	Key source
Psychological contract	What do employees believe they were promised?	Individual / dyad	Rousseau (1989)
Organizational justice	When do people judge decisions and treatment as fair?	Individual / organization	Colquitt et al. (2001)
Organizational commitment	What binds employees to an organization?	Individual	Meyer & Allen (1991)
Person-environment fit	Does the match between person and setting matter?	Individual / organization	Kristof (1996)
Human capital theory	Who should pay for skills, and which skills?	Individual / firm	Becker (1964)
Tournament theory	Why do pay gaps widen up the hierarchy?	Firm / labour market	Lazear & Rosen (1981)

15.1 Job Characteristics Model

THEORY CARD

Core claim. Jobs that offer skill variety, task identity, task significance, autonomy, and feedback produce three critical psychological states, and these states raise internal work motivation, satisfaction, and performance, especially for people who want to grow at work.
Origins. Organizational psychology. Hackman and Oldham (1975) developed the Job Diagnostic Survey, Hackman and Oldham (1976) stated and tested the model, and Hackman and Oldham (1980) applied it to the practice of work redesign.
Level of analysis. The individual job holder and the job.
Key constructs. Core job dimensions, experienced meaningfulness, experienced responsibility, knowledge of results, internal work motivation, growth need strength, motivating potential score.
Boundary conditions. Strongest for employees with high growth need strength and adequate knowledge and skill. Weaker when pay, security, or supervision are so poor that they dominate attention.
Related theories. Self-determination theory (Section 19.6), Job demands-resources, Goal-setting theory (Section 19.4), Social information processing (Section 14.10).

15.1.1 Origins and intellectual history

Twentieth-century industrial engineering had a clear answer to the question of how to design a job: break it into the smallest possible units, standardise them, and remove discretion. The assembly line embodied this logic and delivered large productivity gains, but it also delivered boredom, absenteeism, and turnover. Herzberg's two-factor theory (Herzberg, Mausner, and Snyderman 1959) and the human relations movement had already argued that the content of work matters for motivation. What was missing was a precise, measurable account of *which* job features matter and *why*.

Hackman and Oldham (1975) supplied the measure, the Job Diagnostic Survey (JDS), and Hackman and Oldham (1976) supplied the theory. Their model specified a small set of objective job features, a set of psychological states through which those features work, and an individual difference that

conditions the whole chain. The resulting **job characteristics model** (JCM) became the dominant framework for job design for three decades and remains the starting point for modern work design research (S. K. Parker 2014).

15.1.2 Core logic and propositions

The model links five **core job dimensions** to three **critical psychological states** (Table 15.2). **Skill variety**, **task identity** (completing a whole, identifiable piece of work), and **task significance** (impact on other people) jointly create **experienced meaningfulness**. **Autonomy** creates **experienced responsibility** for outcomes. **Feedback** from the job itself creates **knowledge of results**. When all three states are present, people feel good when they perform well, and this self-generated reward sustains **internal work motivation** (Hackman and Oldham 1976).

Table 15.2. The five core job dimensions of the job characteristics model and how they can be redesigned.

Core dimension	Definition	Psychological state	Low-scoring example	Redesign lever
Skill variety	Range of activities and skills the job requires	Meaningfulness	Single repetitive assembly step	Job enlargement, rotation
Task identity	Completing a whole, visible piece of work	Meaningfulness	Processing one field of a form	Natural work units, case ownership
Task significance	Impact of the work on other people	Meaningfulness	Back-office task with no visible user	Contact with beneficiaries
Autonomy	Discretion over scheduling and methods	Responsibility	Scripted call handling	Vertical loading, self-managing teams
Feedback	Direct information about performance from the work itself	Knowledge of results	Quality checked by a later department	Open feedback channels, dashboards

The original propositions can be stated as follows:

- **P1 (Dimensions to states).** Variety, identity, and significance raise experienced meaningfulness. Autonomy raises experienced responsibility. Feedback raises knowledge of results.
- **P2 (States to outcomes).** The three states jointly raise internal motivation, growth satisfaction, general satisfaction, and work quality, and reduce absenteeism and turnover.
- **P3 (Moderation).** These relationships are stronger for employees high in **growth need strength**, the desire for challenge and personal development. In the later formulation, knowledge and skill and satisfaction with the work context were added as moderators (Hackman and Oldham 1980).

Because the model argues that all three psychological states must be present, Hackman and Oldham proposed a **motivating potential score** (MPS) that combines the dimensions multiplicatively:

$$\text{MPS} = \frac{\text{Skill variety} + \text{Task identity} + \text{Task significance}}{3} \times \text{Autonomy} \times \text{Feedback}$$

The formula averages the three meaningfulness dimensions, so a weakness in one can be offset by strength in another, but it multiplies autonomy and feedback, so a job with no autonomy or no feedback has a motivating potential close to zero no matter how varied it is.

15.1.3 Mechanism

Figure 15.1 shows the job characteristics model.

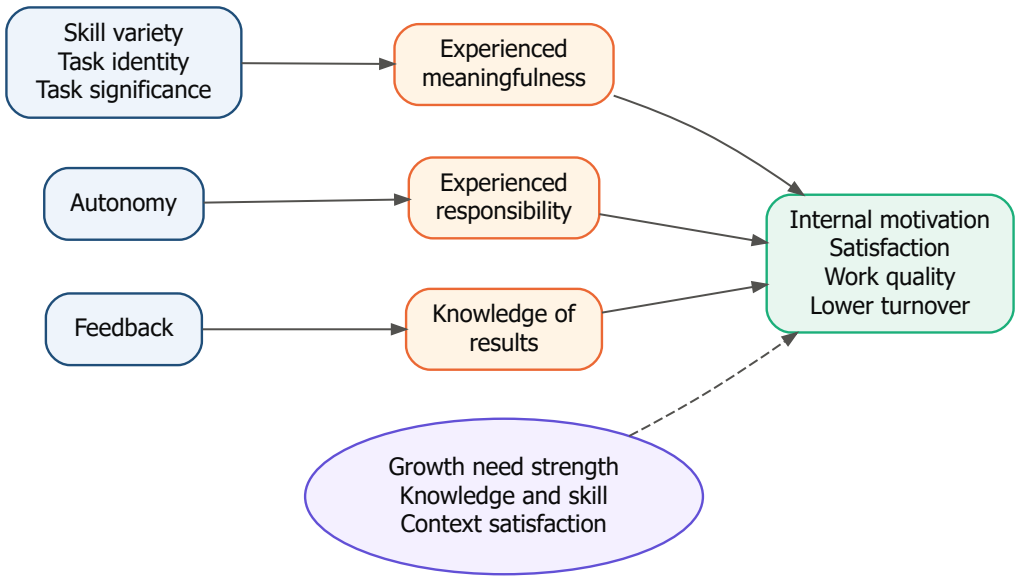


Figure 15.1. The job characteristics model. Core job dimensions create critical psychological states that drive motivation and performance. Growth need strength, knowledge and skill, and context satisfaction moderate the chain. Adapted from Hackman and Oldham (1976, 1980).

Figure 15.2 shows why the multiplicative formula matters. Consider four stylised jobs. The additive index (the simple mean of the five dimensions) ranks the scripted call-centre job as moderately enriched because it scores well on feedback and significance. The MPS ranks it near the bottom because autonomy is almost absent. Whether the multiplicative form is correct is an empirical question, and the evidence on it is discussed below.

15.1.4 Evidence and extensions

The JCM generated hundreds of studies. An early meta-analysis found reasonable support for links between the core dimensions and psychological outcomes such as satisfaction and internal motivation, weaker links to behavioural outcomes such as performance and absenteeism, and little advantage for the multiplicative MPS over simpler additive combinations (Fried and Ferris 1987). The psychological states received partial support as mediators.

The most comprehensive later synthesis, by Humphrey, Nahrgang, and Morgeson (2007), broadened the model in two ways. First, it confirmed that the motivational characteristics of the JCM relate to satisfaction, motivation, and performance, with experienced meaningfulness carrying much of the effect. Second, it showed that **social characteristics** (interdependence, social support, feedback from others) and **work context** explain additional variance in attitudes beyond the classic five dimensions. The Work Design Questionnaire was developed to measure this wider set of task, knowledge, social, and contextual features (Morgeson and Humphrey 2006).

Two further extensions changed how the field thinks about jobs. A. M. Grant (2008) showed experimentally that connecting employees with the people who benefit from their work, for instance

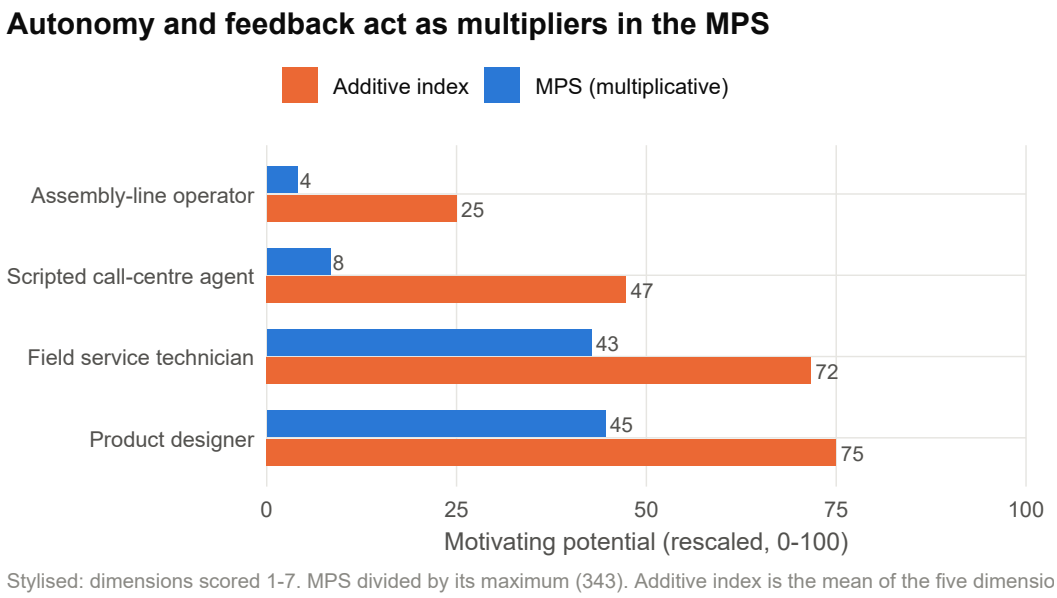


Figure 15.2. Illustrative comparison of the multiplicative motivating potential score (MPS) and a simple additive index for four stylised jobs. Both indices are rescaled to a 0-100 range. Jobs that lack autonomy are penalised far more heavily by the MPS. Stylised scores, not empirical data.

letting fundraising callers learn how scholarships changed recipients' lives, raised persistence and performance. This made **task significance** a relational, designable feature rather than a fixed property of a job. Wrzesniewski and Dutton (2001) introduced **job crafting**, the idea that employees actively reshape the task, relational, and cognitive boundaries of their own jobs. Work design is therefore not only something managers do to employees. It is also something employees do for themselves (S. K. Parker 2014).

DEBATE AND EVIDENCE

How objective are job characteristics? The JCM treats job characteristics as features of the work, but most studies measure them with incumbents' perceptions gathered in the same survey as satisfaction and motivation. Critics in the social information processing (Section 14.10) tradition argued that perceptions of a job are partly constructed from coworkers' cues, and that satisfied workers may simply describe their jobs more favourably. Common-method variance therefore inflates many correlations. Studies that use independent raters, objective job data (such as occupational databases), or field experiments that change jobs provide more credible estimates, and those effects are generally smaller, especially for performance. A second dispute concerns the MPS itself: the evidence offers little reason to prefer the multiplicative formula to a simple average (Fried and Ferris 1987).

15.1.5 Applications in business

BUSINESS CASE

Job redesign on the assembly line. In 1974 Volvo opened an assembly plant at Kalmar, Sweden, organised around work teams that assembled larger portions of each car instead of performing short, repetitive cycles on a conventional moving line. The design applied JCM principles directly: larger work units raised task identity and variety, and team responsibility for scheduling and quality raised autonomy. Volvo later went further at its Uddevalla plant. Both plants were closed in the 1990s, and analysts still disagree about whether their productivity record vindicated or undermined the approach. The lasting lesson is that enrichment interacts with technology and volume: the same design that motivates in low-volume, high-variety production may be uneconomic on a high-volume line.

BUSINESS CASE

Illustrative scenario: redesigning a claims-processing unit. An insurer organises claims processing by function: one clerk verifies policy data, another assesses damage, a third approves payment. Customers wait, and clerks report low motivation. A JCM-based redesign assigns each adjuster a portfolio of whole claims (task identity), gives authority to approve payments below a threshold (autonomy), and shows each adjuster weekly data on cycle time and customer ratings for their own cases (feedback). The model predicts the largest gains among adjusters high in growth need strength, and it warns that adjusters lacking assessment skills will experience the redesign as overload. Training must therefore precede enrichment.

15.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Autonomy is positively related to proactive behaviour, and the relationship is mediated by experienced responsibility.
- *H2.* The positive effect of task significance on performance is stronger when employees have direct contact with beneficiaries.
- *H3.* Job enrichment raises internal motivation more for employees high in growth need strength than for those low in growth need strength.

Validated scales. Job Diagnostic Survey (Hackman and Oldham 1975). Work Design Questionnaire covering task, knowledge, social, and contextual characteristics (Morgeson and Humphrey 2006).

Designs. Field experiments or quasi-experiments around job redesign programmes. Multisource designs pairing incumbent reports with supervisor or expert ratings of job features. Linking occupational databases to employee survey or administrative outcome data. Diary studies of job crafting.

Contexts. Call centres, healthcare, manufacturing teams, platform and gig work, algorithmically managed jobs.

COMMON PITFALL

Treating perceived job characteristics as objective causes. When job features, motivation, and satisfaction come from the same respondent at the same time, a correlation between them says little about the effect of redesign. Use independent sources for job characteristics, exploit changes in jobs over time, and be cautious about the MPS, which adds measurement error without clear predictive gain.

15.2 Job Demands-Resources Theory

THEORY CARD

Core claim. Every job has demands that cost energy and resources that help achieve goals. High demands drive a health-impairment process that leads to burnout, and plentiful resources drive a motivational process that leads to engagement. Resources also buffer the strain caused by demands.
Origins. Occupational health psychology. Demerouti et al. (2001) introduced the model as an account of burnout. Bakker and Demerouti (2007) extended it to engagement, and Bakker and Demerouti (2017) restated it as a theory.
Level of analysis. Individual employee within a job, increasingly studied at team and day level.
Key constructs. Job demands, job resources, personal resources, exhaustion, disengagement, burnout, work engagement, buffering, job crafting.
Boundary conditions. Demands and resources are defined by their function, so the same feature (for example, time pressure) may act as a challenge in one setting and a hindrance in another.
Related theories. Job characteristics model, Conservation of resources, Self-determination theory (Section 19.6), Broaden-and-build theory (Section 25.1).

15.2.1 Origins and intellectual history

Before JD-R, two research traditions dominated the study of job stress. The **demand-control model** argued that strain arises when high job demands meet low decision latitude (Karasek 1979). Burnout research, led by Christina Maslach, described a syndrome of exhaustion, cynicism, and reduced efficacy common in human-service jobs (Maslach, Schaufeli, and Leiter 2001). Both traditions were productive but narrow: they named only a few demands and a single resource (control), and they focused on ill health rather than on positive states.

Demerouti et al. (2001) proposed a more general model. Any job, they argued, can be described with two broad categories. **Job demands** are physical, psychological, social, or organizational aspects of the job that require sustained effort and therefore carry physiological or psychological costs. **Job resources** are aspects that help achieve work goals, reduce demands and their costs, or stimulate growth. In their study, demands were primarily related to exhaustion, while a lack of resources was primarily related to disengagement. The generality of the model, which lets researchers choose the demands and resources relevant to a given occupation, is a major reason for its spread.

15.2.2 Core logic and propositions

Bakker and Demerouti (2007) and Bakker and Demerouti (2017) summarise the theory in a small set of propositions:

- **P1 (Two categories).** Job characteristics can be classified as demands or resources.
- **P2 (Health impairment).** Chronic high demands exhaust mental and physical energy and lead to burnout and health problems.
- **P3 (Motivation).** Job resources foster work engagement and, through it, performance and retention.
- **P4 (Buffering).** Resources weaken the effect of demands on strain.
- **P5 (Boosting).** Resources matter most for motivation when demands are high, because resources become salient when they are needed to cope.
- **P6 (Personal resources and crafting).** Optimism, self-efficacy (Section 21.5), and similar personal resources work like job resources. Engaged employees craft their jobs to gain resources, while exhausted employees may undermine themselves and create more demands, producing reciprocal spirals.

Schaufeli and Bakker (2004) tested the dual-process (Section 23.6) structure with structural equation models in several samples: burnout mediated the link between demands and health problems, and engagement mediated the link between resources and turnover intentions (Figure 15.3).

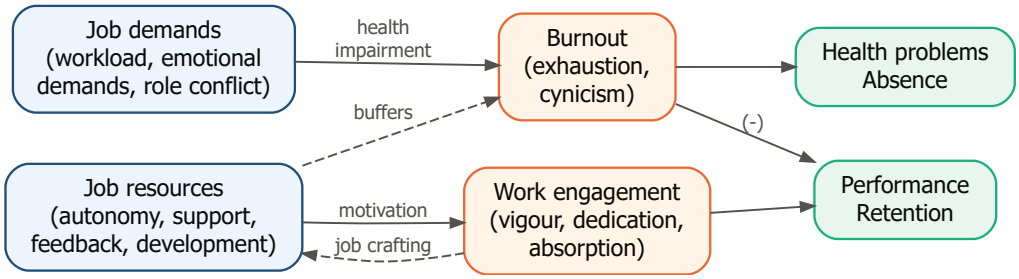


Figure 15.3. The dual processes of job demands-resources theory. Demands drive burnout through a health-impairment process. Resources drive engagement through a motivational process. Dashed arrows show buffering and the reciprocal effect of engagement on resources through job crafting. Adapted from Bakker and Demerouti (2007, 2017).

Figure 15.4 illustrates the buffering proposition with simulated data. Burnout rises with demands for everyone, but the slope is much steeper for employees with few resources.

Table 15.3 lists typical demands and resources studied in different occupations, which shows how the model's open-ended categories are adapted to context.

Table 15.3. Examples of job demands and resources across occupations. Categories are illustrative and must be tailored in each study.

Occupation	Typical job demands	Typical job resources	Likely outcome if unbalanced
Nursing	Emotional demands, shift work, patient load	Team support, autonomy in care decisions	Exhaustion, absence, exit from profession
Software engineering	Deadline pressure, on-call interruptions	Autonomy, learning opportunities, clear priorities	Cynicism, reduced code quality

Table 15.3. Examples of job demands and resources across occupations. Categories are illustrative and must be tailored in each study. *(continued)*

Occupation	Typical job demands	Typical job resources	Likely outcome if unbalanced
Retail and hospitality	Customer hostility, emotional labour, irregular hours	Supervisor support, predictable scheduling	Turnover, service failures
Teaching	Classroom disruption, administrative load	Collegial support, professional discretion	Burnout, early career exit
Remote knowledge work	Blurred boundaries, communication overload	Schedule control, manager check-ins, peer contact	Isolation, overwork

15.2.3 Evidence and extensions

The dual-process structure is well supported across countries and occupations (Schaufeli and Bakker 2004, Bakker and Demerouti 2017). The most important refinement concerns demands. LePine, Podsakoff, and LePine (2005) distinguished **challenge stressors** (workload, responsibility, time urgency) from **hindrance stressors** (red tape, role ambiguity, politics) and showed meta-analytically that challenge stressors relate positively to motivation and performance while hindrance stressors relate negatively. E. R. Crawford, LePine, and Rich (2010) integrated this distinction into JD-R: challenge demands were positively related to engagement, hindrance demands negatively, and both were positively related to burnout. A job can therefore be both energising and exhausting.

Evidence on the interaction propositions is more mixed. Buffering and boosting effects appear in many studies, but they are typically small and are sensitive to how demands and resources are matched. Day-level diary studies show that engagement and exhaustion fluctuate within persons, which has moved JD-R toward dynamic, within-person tests (Bakker and Demerouti 2017).

DEBATE AND EVIDENCE

Is JD-R a theory or a heuristic? Because almost any job feature can be labelled a demand or a resource, critics argue that JD-R is too flexible to be falsified: a feature that predicts burnout becomes a demand, one that predicts engagement becomes a resource. The challenge-hindrance distinction sharpens predictions but also concedes that the same feature can play either role. Defenders reply that the theory specifies distinct processes, mediators, and interactions that can fail empirically, and that Bakker and Demerouti (2017) added reciprocal and personal-resource propositions that are testable. Researchers should therefore classify demands and resources *before* seeing the data and justify the classification from the occupation's work processes.

15.2.4 Applications in business

BUSINESS CASE

Burnout and remote work (conceptual application). Remote and hybrid work change both sides of the JD-R ledger. They typically add resources, notably schedule control and fewer commuting demands, and meta-analytic evidence links telecommuting to higher perceived auton-

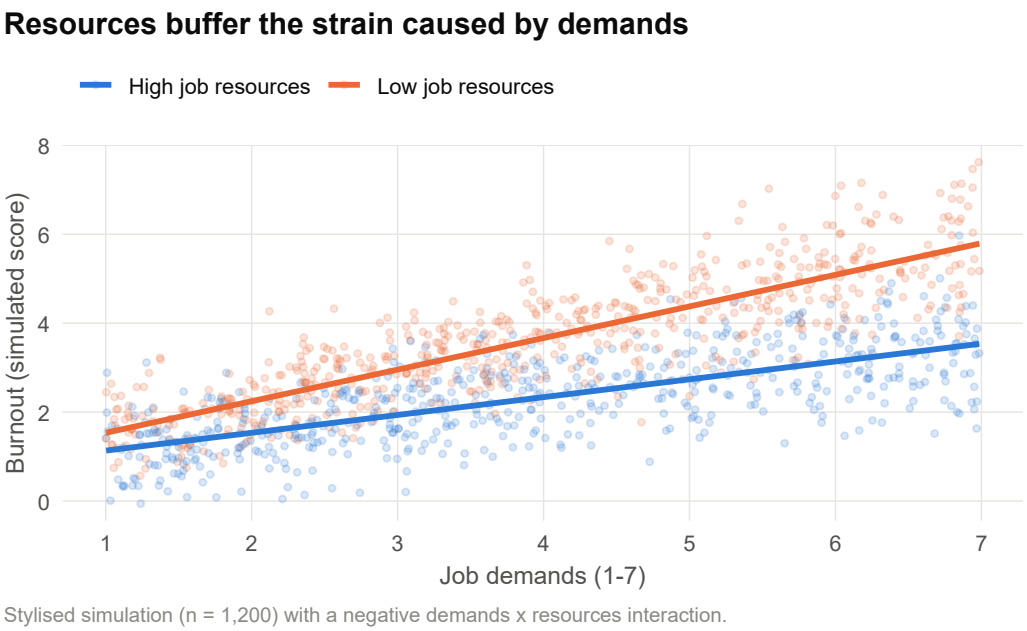


Figure 15.4. Illustrative simulation of the buffering hypothesis. Burnout rises with job demands, but the slope is flatter for employees with high job resources. Simulated data, not empirical estimates.

omy and lower work-family conflict (Gajendran and Harrison 2007). They can also remove resources (informal peer support, spontaneous feedback) and add demands (communication overload, blurred boundaries between work and home). The randomized experiment at the travel agency described in the chapter opening illustrates both sides: home work raised performance and reduced quits, but isolation led many employees to return (Bloom et al. 2015). A JD-R audit of a hybrid policy therefore asks which resources are lost when people are apart and how to replace them (structured check-ins, team days, clear availability norms), rather than treating remote work as uniformly good or bad.

BUSINESS CASE

Illustrative scenario: an engagement survey that measures the wrong thing. A retailer's annual engagement survey shows falling scores in distribution centres. Leadership responds with recognition programmes, a pure resource intervention. A JD-R diagnosis separates demands from resources and finds the real driver: a new scheduling algorithm that raised unpredictability, a hindrance demand. Resources cannot fully offset a hindrance demand that is cheap to remove. The more effective intervention restores advance notice of shifts and only then invests in recognition.

15.2.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Workload is positively related to emotional exhaustion, and the relationship is weaker for employees who report high supervisor support.
- *H2.* Hindrance demands are negatively related to engagement, whereas challenge demands are positively related to engagement.
- *H3.* Day-level engagement predicts next-day job crafting, which in turn increases job resources.

Validated scales. Maslach Burnout Inventory (Maslach, Schaufeli, and Leiter 2001). Oldenburg Burnout Inventory, which measures exhaustion and disengagement (Demerouti et al. 2001). Utrecht Work Engagement Scale, short form (Schaufeli, Bakker, and Salanova 2006).

Designs. Multi-wave panels with cross-lagged models. Daily diary studies with multilevel models that separate within-person from between-person effects. Intervention studies (workload reduction, job crafting training). Linking survey data to sickness absence and turnover records.

Contexts. Healthcare, education, call centres, hybrid work, platform work.

COMMON PITFALL

Labelling variables after the fact. Post hoc classification of a feature as a demand or resource, chosen to fit the results, makes JD-R unfalsifiable. Specify classifications in advance, distinguish challenge from hindrance demands, and test the interaction propositions with adequately powered designs rather than reporting only main effects.

15.3 Conservation of Resources Theory

THEORY CARD

Core claim. People strive to obtain, retain, and protect what they value. Stress occurs when resources are threatened, lost, or not regained after investment, and resource loss weighs more heavily than equivalent gain.

Origins. Stress psychology. Hobfoll (1989), reviewed for organizational research by Halbesleben et al. (2014) and Hobfoll et al. (2018).

Level of analysis. Individual, with extensions to teams and organizations.

Key constructs. Resources (objects, conditions, personal characteristics, energies), resource loss, resource investment, loss spirals, gain spirals, resource caravans.

Boundary conditions. What counts as a resource depends on a person's goals and culture, so predictions require careful specification of the resource in question.

Related theories. Job demands-resources, Loss aversion (Section 24.4), Prospect theory (Section 24.2).

Hobfoll (1989) proposed conservation of resources (COR) theory as an alternative to appraisal-based models of stress. He defined resources broadly: **objects** (a home, tools), **conditions** (tenure, a good marriage), **personal characteristics** (self-esteem, skills), and **energies** (time, money,

knowledge). Stress follows when these are threatened or lost, or when people invest resources and fail to obtain a return.

Two principles give the theory its distinctive predictions. The **primacy of loss** states that resource loss is disproportionately more salient than resource gain, a parallel to loss aversion (Section 24.4) in decision research. The **resource investment** principle states that people must invest resources to protect against loss, recover from it, and gain new resources. From these follow the corollaries that people with more resources are less vulnerable to loss and better able to gain, and that loss and gain both tend to cumulate in **spirals** (Hobfoll et al. 2018). The later formulation also adds a **desperation** principle: when resources are outstretched or exhausted, people may switch to defensive, sometimes seemingly irrational, behaviour to preserve what is left (Figure 15.5).

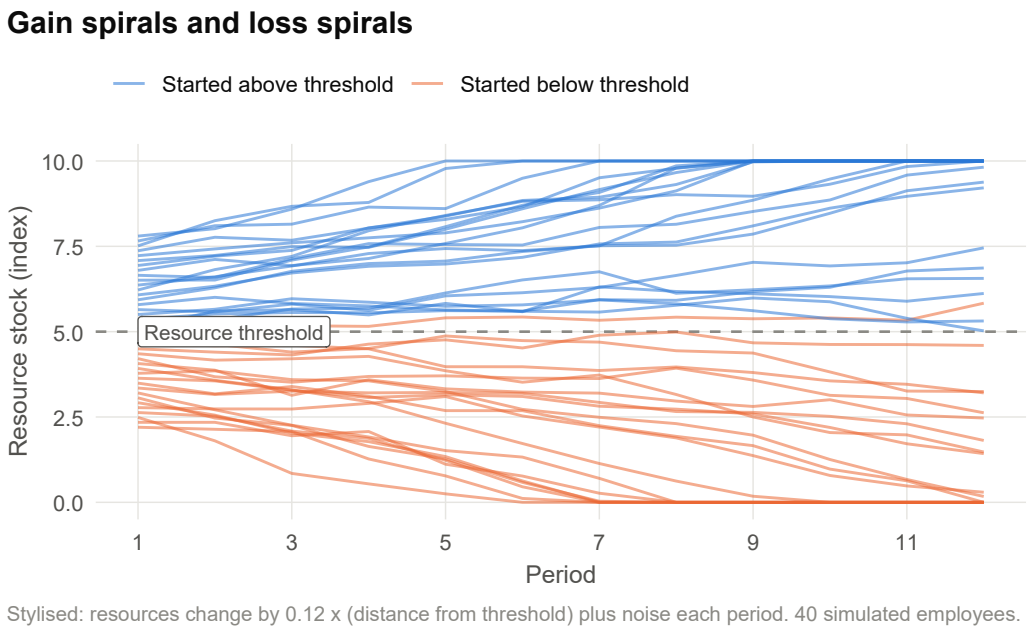


Figure 15.5. Illustrative simulation of resource spirals. Employees above a resource threshold tend to accumulate resources (gain spirals), while those below it tend to lose them (loss spirals), because both loss and gain are self-reinforcing. Simulated trajectories, not empirical data.

COR has become one of the most cited theories in organizational behaviour, used to explain burnout, work-family conflict, the effects of abusive supervision, and why employees withhold discretionary effort when they feel depleted (Halbesleben et al. 2014). Its breadth is also its main weakness. Halbesleben et al. (2014) observe that almost anything has been called a resource, and they propose defining a resource as anything an individual perceives as helping attain his or her goals. That definition improves coherence but makes resources subjective and goal-dependent, which complicates measurement.

BUSINESS CASE

Illustrative scenario: protecting resources during a reorganization. A bank merges two operations teams. COR predicts that the threat of losing valued conditions (a known role, a trusted manager, seniority) will produce stress before any loss occurs, and that employees who already have fewer resources (new hires, those caring for family members) will be most

vulnerable to a loss spiral. Practical responses follow from the theory: announce what will *not* change, protect key resources such as schedule predictability during the transition, and target support at the employees with the thinnest resource reserves.

15.4 Affective Events Theory

THEORY CARD

Core claim. Discrete events at work trigger emotional reactions. These reactions drive immediate, affect-driven behaviours and, accumulated over time, shape attitudes such as job satisfaction, which in turn drive considered, judgment-driven behaviours.
Origins. Organizational psychology. Weiss and Cropanzano (1996).
Level of analysis. Individual, with emphasis on within-person variation over time.
Key constructs. Work environment features, work events (hassles and uplifts), affective reactions (moods and emotions), affective dispositions, work attitudes, affect-driven behaviour, judgment-driven behaviour.
Boundary conditions. Most useful for explaining short-term fluctuations and spontaneous behaviours. Less informative for stable, between-person differences in attitudes.
Related theories. Appraisal theory (Section 25.5), Affect circumplex (Section 25.3), Broaden-and-build theory (Section 25.1).

For decades, job satisfaction research treated satisfaction as an emotional state and studied it as a stable, between-person attitude explained by stable job features. Weiss and Cropanzano (1996) argued that this conflated two different things. Job satisfaction is an **evaluative judgment** about one's job. Emotions and moods are **affective experiences** that occur in response to events, fluctuate within a day, and have their own consequences. Affective events theory (AET) places the **event** at the centre of the analysis (Figure 15.6).

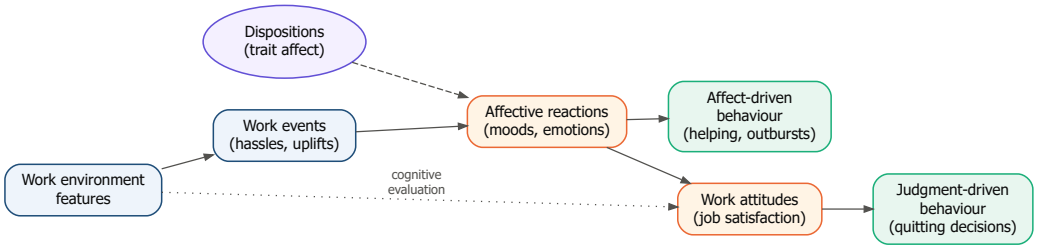


Figure 15.6. Affective events theory. Features of the work environment make some events more likely. Events trigger affective reactions, moderated by dispositions. Affect drives spontaneous behaviour directly and, through attitudes, judgment-driven behaviour. Adapted from Weiss and Cropanzano (1996).

AET makes three contributions. First, it distinguishes **affect-driven behaviours**, such as helping a colleague in a good mood or snapping at a customer after an insult, from **judgment-driven behaviours**, such as deciding to quit, which follow from considered attitudes. Second, it implies that the *pattern* of events over time, not only the average job, shapes satisfaction. Third, it requires methods that capture within-person dynamics. Weiss, Nicholas, and Daus (1999) showed that mood and beliefs about the job independently predict satisfaction, and experience sampling studies found that a substantial share of the variance in mood and job satisfaction lies within persons over time (Ilies

and Judge 2002).

BUSINESS CASE

Service work and the management of events. In front-line service jobs, a small number of hostile customer interactions can dominate an employee's day. AET suggests that managers can influence affect by changing the *frequency* of negative events (better routing of difficult calls, clear escalation paths) and by providing recovery opportunities after them (short breaks, supervisor debriefs), not only by improving average pay or job content. Because affect-driven behaviour is immediate, the timing of support matters: help offered shortly after an incident prevents spillover into the next customer interaction.

15.5 Leader-Member Exchange

THEORY CARD

Core claim. Leaders develop relationships of different quality with different followers. High-quality exchanges, built on trust, respect, and mutual obligation, lead to better performance, attitudes, and retention than low-quality exchanges limited to the formal employment contract.
Origins. Organizational psychology. The vertical dyad linkage studies of Dansereau, Graen, and Haga (1975), consolidated as leader-member exchange (LMX) theory by Graen and Uhl-Bien (1995).
Level of analysis. The leader-follower dyad, with extensions to the differentiation of relationships within teams.
Key constructs. LMX quality, role making, leadership making (stranger, acquaintance, partnership), LMX differentiation, in-group and out-group.
Boundary conditions. Assumes leaders have discretion over resources, assignments, and latitude. Effects of differentiation depend on whether team members see it as fair.
Related theories. Social exchange theory (Section 9.1), Transformational leadership, Organizational justice, Psychological safety (Section 14.8).

15.5.1 Origins and intellectual history

Most early leadership theories assumed an **average leadership style**: a leader behaves in roughly the same way toward everyone, and that style can be described by averaging subordinates' ratings. Dansereau, Graen, and Haga (1975) challenged this assumption with a longitudinal study of managers and their direct reports. They found that managers formed distinct relationships with different subordinates within the first months: some became trusted assistants with latitude and influence, others were managed through formal authority. They called this the **vertical dyad linkage** approach, and it implied that the proper unit of analysis is the dyad, not the leader.

Graen and Uhl-Bien (1995) reviewed twenty-five years of subsequent work and described the theory's evolution in four stages: discovering differentiated dyads, linking exchange quality to outcomes, describing how partnerships are built (**leadership making**), and aggregating dyads to the group and network levels. They argued that leaders should aim to offer every follower the opportunity to develop a high-quality relationship rather than accept a fixed in-group and out-group.

15.5.2 Core logic and propositions

LMX is an application of social exchange theory (Section 9.1) to the leader-follower relationship (Cropanzano and Mitchell 2005). Leaders control resources that followers value: interesting assignments, information, support, discretion, and advocacy. Followers can offer effort beyond the formal role, loyalty, and expertise. When early exchanges go well, each side extends more, and the relationship moves through three stages: **stranger** (formal, contractual exchange), **acquaintance** (testing and limited reciprocity), and **mature partnership** (high trust, mutual influence, and obligation) (Graen and Uhl-Bien 1995).

- **P1 (Differentiation).** Leaders form relationships of varying quality with members of the same unit.
- **P2 (Outcomes).** Higher LMX quality is associated with higher performance, satisfaction, commitment, and lower turnover intentions.
- **P3 (Reciprocity).** LMX develops through reciprocal exchange, with early interactions and perceived similarity shaping its trajectory.
- **P4 (Team context).** Within a team, large differences in LMX quality can harm cooperation and fairness perceptions unless differentiation is seen as based on contribution.

15.5.3 Evidence and extensions

C. R. Gerstner and Day (1997) meta-analysed the early literature. LMX was associated with supervisor-rated performance, satisfaction with the supervisor, overall satisfaction, organizational commitment, role clarity, and lower turnover intentions, but not reliably with actual turnover. Two construct issues stood out: leaders' and members' ratings of the same relationship agreed only moderately, and correlations were stronger when outcomes were reported by the same person who rated LMX. Dulebohn et al. (2012) extended the synthesis to antecedents: follower characteristics (competence, conscientiousness), leader characteristics (expectations of the follower, transformational behaviour), and interpersonal factors (perceived similarity, liking) all predict LMX quality.

Measurement has evolved with the theory. The seven-item LMX-7 treats exchange quality as a single dimension (Graen and Uhl-Bien 1995), whereas the multidimensional LMX-MDM separates affect, loyalty, contribution, and professional respect (Liden and Maslyn 1998). More recent work studies **LMX differentiation** at the team level and **relative LMX**, a member's standing compared with teammates, which links the theory to social comparison (Section 22.1).

> **DEBATE AND EVIDENCE**
>
> **Is LMX a cause or a consequence of performance?** Leaders form better relationships with followers who are already competent, and high performers receive more resources. Cross-sectional correlations between LMX and performance therefore confound the effect of the relationship with the selection of high performers into good relationships. The disagreement between leader and member ratings adds a further problem: if two parties see the same relationship differently, which view is "the" relationship? Longitudinal designs, dyadic models that use both reports, and instruments for early relationship formation are needed to separate the mechanisms (Antonakis et al. 2010).

15.5.4 Applications in business

BUSINESS CASE

Illustrative scenario: LMX in team management. A newly promoted engineering manager inherits eight engineers. Within weeks, she relies heavily on two senior engineers for design decisions and status updates, while the others receive assignments by ticket. LMX theory predicts that the two will become more committed and influential and that the remaining six may disengage, especially if they see the split as favouritism. Practical steps drawn from the leadership-making stage model include offering every team member a stretch assignment early (an opening move in the exchange), holding regular one-to-one meetings with all reports, and making the criteria for access to high-visibility work explicit so that differentiation is seen as earned. Monitoring the *spread* of relationship quality, not just its average, becomes part of the manager's dashboard.

15.5.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* LMX quality is positively related to organizational citizenship behaviour, and the relationship is mediated by felt obligation.
- *H2.* LMX differentiation within teams is negatively related to team coordination when team members perceive differentiation as unfair.
- *H3.* Relative LMX (a member's LMX compared with the team mean) predicts self-efficacy beyond the member's absolute LMX.

Validated scales. LMX-7 (Graen and Uhl-Bien 1995). LMX-MDM (Liden and Maslyn 1998).
Designs. Dyadic data from both leaders and members analysed with dyadic or multilevel models. Longitudinal studies of newly formed dyads. Team-level designs with polynomial regression for relative standing. Field experiments in leader training that encourage equal opportunities for exchange.
Contexts. New-hire onboarding, matrix organizations, remote teams, sales territories.

COMMON PITFALL

Ignoring reverse causality and same-source bias. A survey that asks employees to rate both their relationship with the manager and their own engagement will almost always find a strong correlation. Use supervisor- or record-based outcomes, measure LMX before outcomes, and model the possibility that performance shapes the relationship.

15.6 Transformational Leadership

THEORY CARD

Core claim. Transformational leaders raise followers' motivation and performance beyond what exchange-based incentives achieve by articulating an inspiring vision, serving as role models, stimulating new thinking, and attending to individual needs.

Origins. Political science and organizational psychology. J. M. Burns (1978) distinguished transforming from transactional political leadership. B. M. Bass (1985) operationalised the distinction for organizations.
Level of analysis. Individual leader and followers, with group-level extensions.
Key constructs. Idealised influence (charisma), inspirational motivation, intellectual stimulation, individualised consideration, contingent reward, management by exception, laissez-faire leadership.
Boundary conditions. Effects are thought to be stronger in uncertain or crisis conditions and when followers identify with the collective. Measurement problems limit what current evidence can show (see the debate below).
Related theories. Leader-member exchange, Path-goal theory, Social identity theory (Section 22.2), Upper echelons theory (Section 6.4).

15.6.1 Origins and intellectual history

J. M. Burns (1978), a political historian, contrasted two kinds of leaders. **Transactional** leaders exchange one thing for another: jobs for votes, subsidies for contributions. **Transforming** leaders engage followers so that both are raised to higher levels of motivation and morality. B. M. Bass (1985) brought the idea into organizational psychology and, unlike Burns, treated transformational and transactional leadership as complementary rather than opposite ends of a single continuum. A leader can do both, and transformational behaviour *adds* to the effect of transactional leadership, an idea known as the **augmentation effect**. Elaborated by Bass and colleagues into what they called the "full range" model of leadership, this framework became the most researched approach to leadership of the following three decades.

15.6.2 Core logic and propositions

The full-range model distinguishes three families of behaviour (Table 15.4 compares them with the other leadership theories in this chapter):

- **Transformational**: *idealised influence* (acting as a role model who inspires trust and identification), *inspirational motivation* (articulating an appealing vision), *intellectual stimulation* (challenging assumptions and encouraging creativity), and *individualised consideration* (coaching and attending to each follower's needs).
- **Transactional**: *contingent reward* (clarifying expectations and rewarding their achievement) and *management by exception* (monitoring and correcting deviations, either actively or passively).
- **Laissez-faire**: the avoidance of leadership responsibilities.

The theory's central proposition is that transformational behaviours change followers' values, self-concepts, and identification with the collective, so that they pursue collective goals for intrinsic reasons rather than for rewards alone.

Table 15.4. Three leadership theories compared.

Feature	Leader-member exchange	Transformational leadership	Path-goal theory
Focal unit	Leader-follower dyad	Leader and followers as a group	Leader, follower, and task

Table 15.4. Three leadership theories compared. *(continued)*

Feature	Leader-member exchange	Transformational leadership	Path-goal theory
Core mechanism	Reciprocal social exchange	Identification and value change	Expectancy: clarifying paths and raising valences
Key behaviours	Offering latitude, support, and challenging assignments	Vision, role modelling, stimulation, consideration	Directive, supportive, participative, achievement-oriented
Contingency view	Weak: quality varies by relationship	Weak: broadly effective	Strong: best behaviour depends on task and follower
Main measure	LMX-7, LMX-MDM	Multifactor Leadership Questionnaire	Leader behaviour scales
Main critique	Reverse causality, rater disagreement	Conflates behaviour with effects	Mixed tests, many untested propositions

15.6.3 Evidence and extensions

The quantitative evidence is large. Judge and Piccolo (2004) meta-analysed studies of the full range model and found that transformational leadership was strongly related to follower satisfaction, motivation, and rated leader effectiveness. Contingent reward, a transactional behaviour, showed overall validities of a similar size and was in some settings as predictive as transformational leadership, while laissez-faire leadership was negatively related to outcomes. The authors also found that the relative validity of the dimensions differed across settings and that transformational and contingent reward leadership were highly correlated, which made their unique effects hard to separate.

Experimental evidence is thinner but exists. In a small field experiment, Barling, Weber, and Kelloway (1996) trained bank branch managers in transformational leadership and found improvements in subordinates' perceptions of their managers, in subordinates' organizational commitment, and in branch-level financial outcomes compared with an untrained control group. The study is often cited because it combines random assignment with objective outcomes, but its sample was small.

DEBATE AND EVIDENCE

Back to the drawing board? Knippenberg and Sitkin (2013) published a sweeping critique of charismatic-transformational leadership research. They argued, first, that the construct lacks an independent definition: it is defined largely by its effects (followers who are inspired and perform beyond expectations), which makes the link between leadership and outcomes partly tautological. Second, the theory does not explain why these four particular dimensions belong together or how each produces its effects. Third, the dominant measures, especially the Multifactor Leadership Questionnaire, ask followers to rate behaviours in terms that already include positive outcomes, and the proposed dimensional structure is often not reproduced empirically. They recommended abandoning the umbrella construct and studying specific, well-defined leader behaviours and their mechanisms. A related methodological critique notes that follower ratings of leadership are endogenous to performance, liking, and attributions, so most observational estimates are biased (Antonakis et al. 2010). Defenders point to experimental studies like Barling, Weber, and Kelloway (1996) and argue that refinement, not abandonment, is the appropriate response. The debate is not settled, but it has raised the standard of evidence expected of leadership research.

15.6.4 Applications in business

BUSINESS CASE

Leadership development programmes. Transformational leadership is the intellectual backbone of many corporate leadership development curricula, which teach managers to communicate a vision, coach individuals, and encourage challenge to the status quo. The evidence suggests three design principles. First, do not neglect transactional basics: clear goals and contingent recognition are consistently effective (Judge and Piccolo 2004). Second, evaluate programmes with outcomes that are not rated by the same followers who rate the leader, as Barling, Weber, and Kelloway (1996) did with branch financial data. Third, teach specific behaviours (how to give developmental feedback, how to frame change around shared values) rather than an abstract "inspirational style", in line with Knippenberg and Sitkin (2013).

15.6.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Leaders' intellectual stimulation is positively related to team creativity, and the relationship is mediated by psychological safety.
- *H2.* Contingent reward leadership explains variance in follower performance beyond transformational leadership in routine production settings.
- *H3.* Training in vision communication increases objectively measured unit performance relative to a control group.

Validated scales. The Multifactor Leadership Questionnaire, developed from B. M. Bass (1985), remains the most common instrument, but users should report evidence on its factor structure in their sample and consider behaviour-specific alternatives.

Designs. Randomised leadership training with objective outcomes. Laboratory experiments that manipulate a specific leader behaviour (for example, a vision speech). Designs that address endogeneity, such as instrumental variables or exogenous leader turnover (Antonakis et al. 2010).

Contexts. Organizational change and turnaround, start-ups, military and public-service organizations, healthcare teams.

COMMON PITFALL

Measuring leadership with outcomes. Items such as "my leader motivates me to do more than I expected" fold the dependent variable into the independent variable. Use behaviour-specific measures, separate sources, and designs that can support causal claims before concluding that a leader's style caused performance.

15.7 Path-Goal Theory

THEORY CARD

Core claim. Effective leaders increase followers' motivation by clarifying the paths to work goals, removing obstacles, and making goal attainment more rewarding. The behaviour that does this best depends on the task and the follower.
Origins. Organizational psychology. House (1971), reformulated in House (1996).
Level of analysis. Leader-follower dyad and work group.
Key constructs. Directive, supportive, participative, and achievement-oriented leader behaviour, task structure, follower ability, locus of control, expectancies, valences.
Boundary conditions. Leader behaviour matters only where the situation does not already provide clarity and rewards, that is, where there are few **substitutes for leadership** (S. Kerr and Jermier 1978).
Related theories. Expectancy theory (Section 19.2), Transformational leadership, Goal-setting theory (Section 19.4).

House (1971) built path-goal theory directly on expectancy theory (Section 19.2) (Vroom 1964). Followers are motivated when they believe effort leads to performance and performance leads to valued rewards. The leader's job is to strengthen both links: clarify what effective performance requires, remove roadblocks, provide support, and tie rewards to goal attainment. Because these needs vary, the theory is explicitly **contingent**. Directive leadership helps when tasks are ambiguous and followers inexperienced, but is redundant or even irritating when tasks are routine and well understood. Supportive leadership helps most when tasks are stressful, boring, or dangerous. Participative leadership suits followers who want control, and achievement-oriented leadership suits complex tasks where high standards and confidence in followers raise expectancies (Figure 15.7).

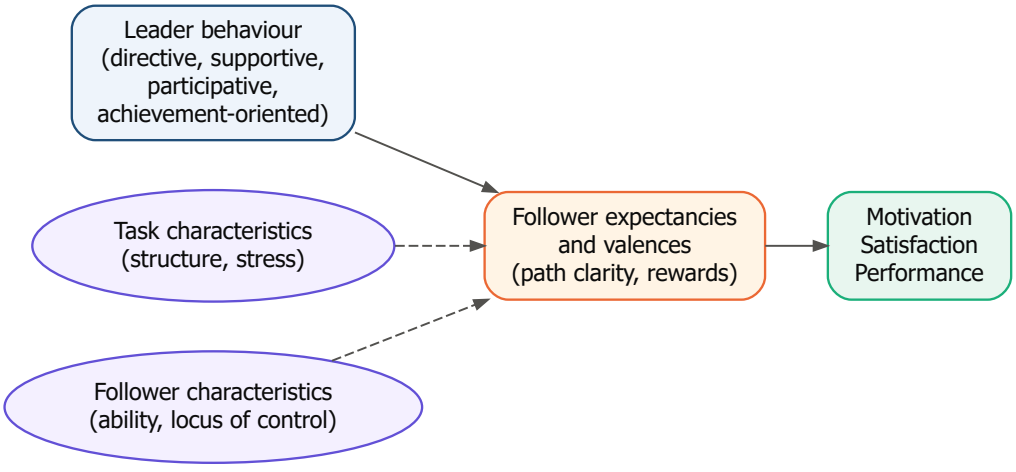


Figure 15.7. Path-goal theory. Leader behaviour affects follower expectancies and valences, which drive motivation and performance. Task and follower characteristics determine which behaviour is effective. Adapted from House (1971, 1996).

The evidence is mixed. A meta-analysis by Wofford and Liska (1993) found support for some propo-

sitions and not for others, and noted that many tests used weak designs and measures. House (1996) acknowledged these limitations and reformulated the theory around a broader set of leader behaviours, including work facilitation, interaction facilitation, group-oriented decision processes, representation and networking, and value-based behaviour, while retaining the core idea that leaders compensate for what the environment does not provide. A related line of work on **substitutes for leadership** argued that professional training, intrinsically satisfying tasks, and formalised procedures can make some leader behaviours unnecessary (S. Kerr and Jermier 1978). Path-goal theory's lasting contribution is less any single prediction than the principle that leadership style should be matched to the motivational deficits of a specific situation.

BUSINESS CASE

Illustrative scenario: one manager, two teams. A regional operations manager oversees a newly formed analytics team working on ill-defined problems and a mature payroll team running a stable, highly regulated process. Path-goal theory predicts that directive behaviour (clear problem definitions, milestones, and methods) will raise the analytics team's expectancies, while the same behaviour would be redundant for payroll. The payroll team, facing repetitive work and deadline stress at month end, benefits more from supportive behaviour and from removing obstacles such as late data from other departments. The prescription is not a single best style but a deliberate diagnosis of what each team lacks.

15.8 Psychological Contract Theory

THEORY CARD

Core claim. Employees hold beliefs about the reciprocal promises that bind them and their employer. When they perceive that the organization has broken those promises, they react with feelings of violation and reduced trust, commitment, and effort, even if no written agreement was breached.
Origins. Organizational behaviour. Rousseau (1989) reconceptualised the psychological contract as an individual-level belief, extended in Rousseau (1995). S. L. Robinson and Rousseau (1994) and Morrison and Robinson (1997) developed the theory of breach and violation.
Level of analysis. Individual employee, in relation to the organization and its agents.
Key constructs. Perceived promises, transactional and relational contracts, breach, violation, reneging, incongruence.
Boundary conditions. Applies where employment involves promises that are implicit, long-term, or incompletely specified. Contract content varies with career stage, labour market, and national employment norms.
Related theories. Social exchange theory (Section 9.1), Organizational justice, Organizational commitment.

15.8.1 Core logic

Earlier writers used "psychological contract" loosely to describe unwritten expectations between employees and employers. Rousseau (1989) gave the concept a precise meaning: an individual's **belief** about the terms of a reciprocal exchange with another party, grounded in a perceived **promise**. Because it is a belief, two employees in the same job can hold different contracts, and

the organization, which consists of many agents (recruiters, managers, HR policies), cannot simply read off what it has promised. Rousseau (1995) distinguished **transactional** contracts, which are specific, short-term, and largely monetary, from **relational** contracts, which are open-ended and include loyalty, support, and development. Later typologies add balanced and transitional forms (Table 15.5).

Table 15.5. Transactional and relational psychological contracts, following Rousseau (1995).

Feature	Transactional contract	Relational contract
Time frame	Short-term, specified duration	Open-ended, long-term
Content	Pay for defined performance	Pay plus loyalty, development, security
Specificity	Explicit and narrow	Implicit and broad
Typical setting	Contract and gig work, fixed-term projects	Career employment, internal labour markets
Typical reaction to breach	Renegotiation or exit	Felt betrayal, withdrawal of discretionary effort
Employer lever	Clear terms and prompt payment	Consistent signals from managers and policies

Morrison and Robinson (1997) separated two reactions that earlier work had merged. **Breach** is the cognitive perception that the organization has failed to fulfil one or more obligations. **Violation** is the emotional experience of anger and betrayal that may follow. Breach can arise from **reneging**, when agents knowingly fail to keep a promise, or from **incongruence**, when the two parties understood the promise differently. Whether breach turns into violation depends on how the employee interprets it, especially on attributions of intent and perceptions of fairness.

15.8.2 Evidence

The landmark longitudinal study by S. L. Robinson and Rousseau (1994) surveyed graduates of a master of business administration programme when they took up new jobs and again two years later. More than half reported that their employer had violated the psychological contract, which led the authors to describe violation as the norm rather than the exception. Violation was associated with lower trust, satisfaction, and intentions to remain. A follow-up longitudinal study confirmed that breach and violation are distinct, and that the step from one to the other depends on attributions of reneging and perceived unfairness (S. L. Robinson and Morrison 2000). The meta-analysis by H. Zhao et al. (2007) found that breach is related to feelings of violation and mistrust, to lower job satisfaction and organizational commitment, to higher turnover intentions, and to reduced citizenship behaviour and in-role performance, with affective reactions mediating many of these links.

DEBATE AND EVIDENCE

Whose contract? Because the psychological contract exists in the employee's mind, critics question whether "contract" is the right metaphor: a contract normally requires mutual agreement, and the organization is not a single party that can hold beliefs. Measurement adds a second problem. Many studies ask employees directly how well promises have been fulfilled, which mixes perceived promises, perceived delivery, and evaluation in a single item. Designs that measure promises at entry and fulfilment later, or that compare employee and manager views, are more informative.

BUSINESS CASE

Illustrative scenario: return-to-office policies as contract change. A software firm hired many employees during a period when remote work was described in recruiting conversations as permanent. Two years later, it mandates four office days per week. Legally, nothing has been breached, since the employment contracts never mentioned location. Psychological contract theory predicts that employees who perceived a promise will experience breach, and those who attribute it to reneging will experience violation and withdraw discretionary effort or leave. The theory suggests three mitigations: acknowledge the change explicitly as a change, explain the reasons (to shift attributions from reneging to necessity), and offer something in exchange, such as transition time or flexible days, so that the reciprocal balance is restored.

15.9 Organizational Justice

THEORY CARD

Core claim. Employees evaluate not only what they receive (distributive justice) but also how decisions are made (procedural justice) and how they are treated and informed (interactional justice). Each form of fairness has distinct effects on attitudes and behaviour.
Origins. Social psychology and organizational behaviour. Equity theory (Section 19.3) (J. S. Adams 1965), procedural justice in legal settings (J. Thibaut and Walker 1975), allocation procedures (Leventhal 1980), and interactional justice (Bies and Moag 1986). Integrated by Colquitt (2001) and Colquitt et al. (2001).
Level of analysis. Individual perceptions of organizational decisions, with extensions to team justice climate.
Key constructs. Distributive, procedural, interpersonal, and informational justice, voice, fair process effect.
Boundary conditions. The weight of procedural justice is greatest when outcomes are unfavourable or uncertain. Reactions to voice vary with cultural values such as power distance (Brockner et al. 2001).
Related theories. Equity theory (Section 19.3), Psychological contract, Social exchange theory (Section 9.1), Leader-member exchange.

15.9.1 Origins and intellectual history

Justice research began with outcomes. Equity theory (Section 19.3) proposed that people compare their ratio of outcomes to inputs with that of others and feel distress when the ratios differ (J. S. Adams 1965). This is **distributive justice**. J. Thibaut and Walker (1975) shifted attention to process. Studying legal dispute resolution, they found that people's satisfaction depended heavily on **process control**, the opportunity to present their case, even when they had no control over the final decision. This became known as the **voice** effect. Leventhal (1980) extended procedural justice to organizational allocation decisions and proposed six rules that make procedures fair. Bies and Moag (1986) added **interactional justice**: the quality of interpersonal treatment people receive when procedures are enacted, including truthfulness, justification, respect, and propriety.

15.9.2 Core logic and propositions

Table 15.6 summarises the four dimensions that most researchers now use, following the construct validation by Colquitt (2001), which showed that separating interactional justice into interpersonal and informational components fit the data better than three-factor alternatives.

Table 15.6. Four dimensions of organizational justice, following Colquitt (2001).

Dimension	Focal question	Fairness criteria	Key source
Distributive	Are outcomes fair given contributions?	Equity (also equality or need in some settings)	Adams (1965)
Procedural	Is the decision process fair?	Voice, consistency, bias suppression, accuracy, correctability, representation, ethicality	Thibaut & Walker (1975), Leventhal (1980)
Interpersonal	Was I treated with dignity and respect?	Respect, propriety	Bies & Moag (1986)
Informational	Was the decision explained truthfully and adequately?	Truthfulness, justification, timeliness	Bies & Moag (1986)

The theory's main propositions are:

- **P1 (Distinctiveness).** The justice dimensions are related but distinct, and each explains unique variance in outcomes (Colquitt et al. 2001).
- **P2 (Targets).** Procedural justice is most strongly tied to evaluations of the organization and its systems, while interpersonal and informational justice are most strongly tied to evaluations of the decision agent, typically the supervisor.
- **P3 (Fair process effect).** Fair procedures and treatment soften negative reactions to unfavourable outcomes. Procedural justice matters most when outcomes are poor (Brockner and Wiesenfeld 1996).
- **P4 (Behaviour).** Perceived injustice leads to withdrawal, reduced citizenship, and retaliation, while justice fosters trust, commitment, and cooperation.

Figure 15.8 sketches the interaction behind P3.

15.9.3 Evidence and extensions

The meta-analysis by Colquitt et al. (2001) integrated twenty-five years of research. It found that the justice dimensions are strongly correlated but distinguishable, that procedural justice predicts a wide range of attitudes and behaviours, and that interpersonal and informational justice add explanatory power beyond procedural and distributive justice. Two field studies became classics. In a manufacturing company that imposed a temporary pay cut, Greenberg (1990) found that employee theft rose in the affected plants, and that it rose much less in the plant where management explained the cut thoroughly and with expressed regret. In a study of terminated employees, Lind et al. (2000) found that how fairly people felt they had been treated at the time of termination was a strong predictor of whether they filed wrongful-termination claims. Both studies show that the *manner* of an unfavourable decision has economic consequences.

DEBATE AND EVIDENCE

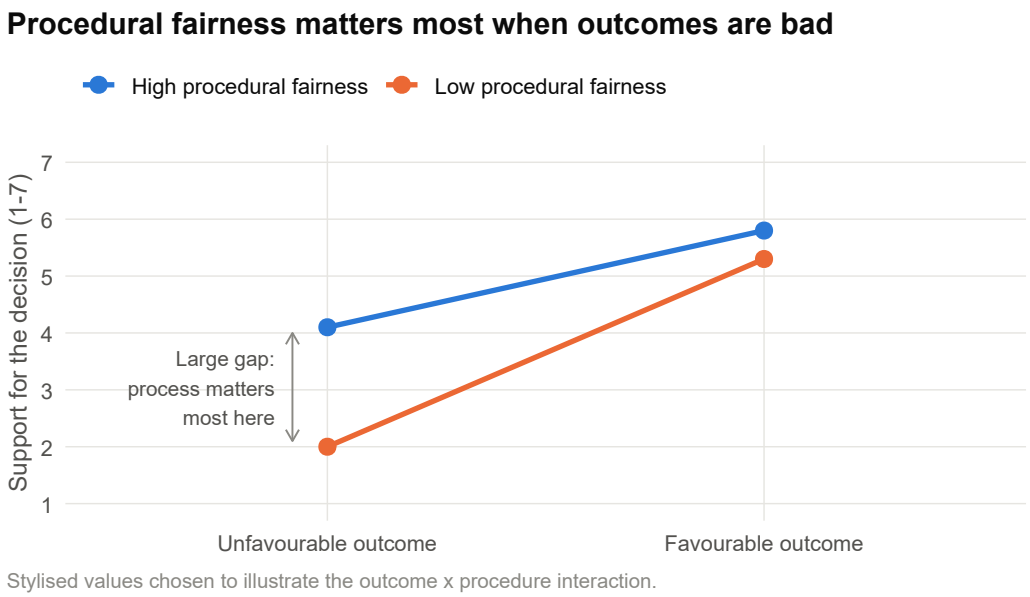


Figure 15.8. Stylised fair process effect. When outcomes are favourable, procedures matter little. When outcomes are unfavourable, fair procedures and treatment greatly reduce negative reactions. Pattern consistent with the interaction reviewed by Brockner and Wiesenfeld (1996). Not empirical data.

How many dimensions, and is fairness one judgment? The four-factor structure is well supported psychometrically, but the dimensions are highly correlated and often have similar effects. Some researchers therefore argue for an **overall justice** judgment that people form quickly and then use as a heuristic, with the specific dimensions as its inputs. A second debate concerns the boundary between interactional and procedural justice: if a respectful explanation is part of the procedure, is it a separate dimension? Cultural research adds a third: the voice effect is weaker in high power-distance settings, where employees expect less say in decisions (Brockner et al. 2001).

15.9.4 Applications in business

BUSINESS CASE

Layoffs and procedural justice. Layoffs are the clearest test of the fair process effect: the outcome is unfavourable by definition, for victims and often for survivors. The research above suggests a practical checklist. Use consistent, pre-announced criteria (consistency and bias suppression). Give employees a way to correct errors in the data used for selection (correctability). Explain the business reasons candidly and early (informational justice). Deliver the news in person, privately, and with respect, never by mass e-mail (interpersonal justice). Provide fair severance and outplacement support (distributive justice). The evidence that poor treatment at termination predicts legal claims (Lind et al. 2000) and that explanations reduce retaliation (Greenberg 1990) means these steps reduce legal and operational risk as well as human cost. Survivors watch how departing colleagues are treated and update their own trust and commitment accordingly (Brockner and Wiesenfeld 1996).

15.9.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Informational justice during an organizational change is positively related to change acceptance, and the effect is stronger when the change is unfavourable to the employee.
- *H2.* Interpersonal justice from the supervisor predicts supervisor-directed citizenship behaviour more strongly than procedural justice does.
- *H3.* Team justice climate moderates the relationship between individual justice perceptions and team commitment.

Validated scales. Colquitt's four-dimension organizational justice measure (Colquitt 2001).
Designs. Vignette and laboratory experiments manipulating outcome and process independently. Field quasi-experiments around layoffs, pay changes, or appraisal system redesign. Multilevel designs for justice climate. Archival data on grievances and legal claims.
Contexts. Performance appraisal, pay decisions, downsizing, algorithmic management, promotion systems.

COMMON PITFALL

Treating fairness as an outcome only. Many organizations evaluate decisions solely by asking whether outcomes are defensible. Justice research shows that process and treatment often matter as much, especially when the outcome is bad. In research, the parallel mistake is to measure a single global fairness item and attribute effects to a specific dimension.

15.10 Organizational Commitment

THEORY CARD

Core claim. Commitment to an organization is a psychological state with three components: a desire to stay (affective commitment), a perceived need to stay because of the costs of leaving (continuance commitment), and a felt obligation to stay (normative commitment). Each has different antecedents and consequences.
Origins. Organizational psychology. J. P. Meyer and Allen (1991), with measures in N. J. Allen and Meyer (1990).
Level of analysis. Individual employee.
Key constructs. Affective, continuance, and normative commitment, commitment profiles, side bets, perceived alternatives.
Boundary conditions. Developed in North American samples with conventional employment. Normative commitment may operate differently in more collectivist cultures.
Related theories. Psychological contract, Social identity theory (Section 22.2), Escalation of commitment (Section 14.2).

Earlier research on commitment had followed separate threads: some scholars saw commitment as emotional attachment, others as the accumulation of investments (side bets) that would be lost on leaving, and still others as a moral duty. J. P. Meyer and Allen (1991) argued that these were not competing definitions but distinct **components** of commitment that can coexist in the same person

in different strengths. All three tie an employee to the organization, but they do so for different reasons and should therefore have different effects on behaviour beyond staying (Table 15.7).

Table 15.7. The three-component model of organizational commitment, following Meyer and Allen (1991) and Meyer et al. (2002).

Component	Mindset	Typical antecedents	Relation to performance and citizenship
Affective	I want to stay	Supportive leadership, justice, meaningful work	Positive
Continuance	I need to stay	Investments that would be lost, few alternatives	Weak or negative
Normative	I ought to stay	Socialization, reciprocity for benefits received	Positive, weaker than affective

The meta-analysis by J. P. Meyer et al. (2002) supported this differentiation. All three components were negatively related to turnover intentions and turnover. Affective commitment had the strongest and most favourable correlations with attendance, performance, and citizenship behaviour, and the strongest negative correlation with stress and work-family conflict. Continuance commitment was unrelated or negatively related to these desirable outcomes. Normative commitment correlated highly with affective commitment, which has fuelled a long debate about whether it is a separate construct. Critics also note that continuance commitment appears to combine two facets, high personal sacrifice and lack of alternatives, which behave differently.

BUSINESS CASE

Illustrative scenario: the golden handcuffs problem. A technology firm relies heavily on deferred equity that vests over four years to retain engineers. Retention is high, but engagement surveys show low enthusiasm among long-tenured staff. The three-component model diagnoses the problem: deferred pay builds continuance commitment, which keeps people in place without raising effort or citizenship. Building affective commitment requires different levers, such as meaningful work, fair treatment, and development, and the firm may prefer moderate voluntary turnover among people who stay only because they need to.

15.11 Person-Environment Fit

THEORY CARD

Core claim. Attitudes and behaviour depend on the compatibility between characteristics of the person and characteristics of the work environment. Good fit promotes satisfaction, commitment, and retention.

Origins. Interactional psychology, systematised for organizations by Kristof (1996) and synthesised by Kristof-Brown, Zimmerman, and Johnson (2005).

Level of analysis. Individual in relation to job, group, supervisor, and organization.

Key constructs. Person-job, person-organization, person-group, and person-supervisor fit, supplementary and complementary fit, needs-supplies and demands-abilities fit, perceived and actual fit.

Boundary conditions. Fit effects are stronger for attitudes than for performance. Strong

emphasis on fit can reduce diversity and adaptability.
Related theories. Attraction-selection-attrition (Section 14.9), Organizational culture (Section 14.6), Job characteristics model.

Fit is an old idea, but Kristof (1996) brought conceptual order to it. She distinguished **supplementary fit**, when a person is similar to others in the environment (for example, sharing the organization's values), from **complementary fit**, when a person or environment supplies what the other lacks. Complementary fit has two forms: **needs-supplies** fit, when the job provides what the person wants, and **demands-abilities** fit, when the person has what the job requires. Fit can also be assessed at different levels of the environment, as Table 15.8 shows. Fit theory connects naturally to the attraction-selection-attrition (Section 14.9) framework, which argues that organizations become more homogeneous as similar people are attracted, selected, and retained (Schneider 1987).

Table 15.8. Types of person-environment fit, with outcomes following Kristof-Brown, Zimmerman and Johnson (2005).

Type of fit	Compatibility between	Strongest outcomes in meta-analysis	HR practice
Person-job	Abilities and job demands, needs and job supplies	Job satisfaction, organizational commitment	Job analysis, skills testing
Person-organization	Values and organizational culture	Organizational commitment, turnover intentions	Values-based interviewing, realistic previews
Person-group	Person and coworkers	Coworker satisfaction, group cohesion	Team-based selection
Person-supervisor	Person and supervisor values or personality	Satisfaction with supervisor	Manager involvement in hiring

The meta-analysis by Kristof-Brown, Zimmerman, and Johnson (2005) found that person-job fit was strongly related to job satisfaction and that person-organization fit was strongly related to organizational commitment, with weaker relationships to performance. These findings justify the widespread use of fit in hiring, but they come with two caveats. First, perceived fit (asking people how well they fit) produces much stronger correlations than fit calculated from separate measures of person and environment, partly because perceptions are contaminated by satisfaction (Cable and DeRue 2002). Second, fit computed as a difference score conceals which component drives the effect. Edwards and Parry (1993) showed how polynomial regression and response surface analysis can test fit hypotheses without these problems.

BUSINESS CASE

"Culture fit" in hiring. Many firms screen candidates for culture fit, and the evidence that person-organization fit predicts commitment supports the practice. Yet fit theory also explains its risks. When interviewers judge fit intuitively, they tend to reward similarity to themselves, which can exclude candidates who would add complementary skills and can narrow the diversity of perspectives in the firm, an outcome predicted by the attraction-selection-attrition cycle (Schneider 1987). A disciplined application specifies the values that matter, assesses them with structured questions, and balances supplementary fit on core values with complementary fit on skills.

15.12 Human Capital Theory

THEORY CARD

Core claim. Skills and knowledge are a form of capital: people and firms invest in them, bearing costs today for returns later. Who pays for training depends on whether the skills are general (valuable to many employers) or firm-specific (valuable mainly to the current employer).
Origins. Labour economics. G. S. Becker (1962) and G. S. Becker (1964), with the empirical earnings function of Mincer (1974).
Level of analysis. Individual worker and firm, with extensions to the firm's stock of human capital as a strategic resource.
Key constructs. General human capital, firm-specific human capital, returns to schooling and experience, training investment, hold-up.
Boundary conditions. The classic predictions assume competitive labour markets. Frictions, wage compression, and information asymmetries change who pays for training (Acemoglu and Pischke 1999).
Related theories. Signaling theory (Section 8.2), Resource-based theory (Section 4.1), Knowledge-based view (Section 4.3), Efficiency wages (Section 17.5).

15.12.1 Origins and core logic

G. S. Becker (1962) and G. S. Becker (1964) treated education and training as investments that raise future productivity and earnings. The theory's most influential organizational prediction concerns on-the-job training. **General** training raises a worker's productivity equally at many firms. In a competitive labour market, a firm that paid for it would see the worker poached or would have to pay the worker their higher market value, so the firm cannot recover its investment. The worker therefore pays, typically by accepting lower wages during training. **Firm-specific** training raises productivity only at the current employer. The worker's outside wage does not rise, so neither party can fully appropriate the returns alone, and the efficient arrangement is for firm and worker to **share** the costs and the returns. Sharing gives both parties a stake in continuing the relationship, which predicts lower quit and layoff rates for workers with firm-specific skills.

Mincer (1974) turned the theory into an empirical workhorse, the **Mincer earnings equation**, which regresses log earnings on years of schooling and a quadratic in labour market experience:

$$\ln w_i = \beta_0 + \beta_1 S_i + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i$$

where S_i is schooling and X_i is experience. The coefficient β_1 is interpreted as an approximate rate of return to a year of schooling (Table 15.9).

Table 15.9. General versus firm-specific human capital, following Becker (1964).

Dimension	General human capital	Firm-specific human capital
Definition	Raises productivity at many employers	Raises productivity mainly at the current employer
Examples	Literacy, programming languages, accounting standards	Proprietary systems, internal routines, client relationships
Who pays (competitive market)	Worker, through lower training wages	Shared between firm and worker

Table 15.9. General versus firm-specific human capital, following Becker (1964). *(continued)*

Dimension	General human capital	Firm-specific human capital
Who captures returns	Worker	Shared, which binds both parties
Strategic value to the firm	Low: available in the labour market	High: hard to imitate, but creates hold-up risk
Typical HR practice	Tuition support with repayment clauses	Internal academies, mentoring, job rotation

15.12.2 Evidence and extensions

The theory has been refined in three directions. First, Acemoglu and Pischke (1999) showed that when labour markets have frictions, such as search costs, information asymmetries about worker skill, or institutions that compress wages, firms can profit from paying for *general* training, because they capture part of the returns. This explains why many employers fund training that is partly general, as in firm-sponsored apprenticeship systems. Second, Lazear (2009) proposed a **skill-weights** view: every individual skill may be general, but each firm uses a particular *combination* of skills, which makes a worker's skill bundle partly specific to the firm. Third, strategy scholars linked firm-specific human capital to sustainable advantage, while warning that human assets can walk out the door and bargain for the rents they create (Coff 1997).

The strategic payoff has been tested directly. The meta-analysis by Crook et al. (2011) found a positive relationship between human capital and firm performance, which was stronger when the human capital was firm-specific and therefore harder for rivals to acquire. The theory also faces a long-standing rival: signaling theory (Section 8.2) argues that education may raise wages because it reveals pre-existing ability rather than because it creates productivity (Spence 1973).

DEBATE AND EVIDENCE

Human capital or signaling? Both theories predict that more educated workers earn more, so the earnings premium alone cannot discriminate between them. Human capital theory predicts that each year of schooling adds productivity, while signaling theory predicts extra returns to completing a credential ("sheepskin effects") and declining returns as employers learn workers' true ability over time. Research using compulsory schooling laws, twins, and employer learning models generally finds that a substantial part of the return reflects real productivity gains, but that signaling also plays a role. For managers, the practical implication is that credentials are informative but incomplete evidence of skill.

BUSINESS CASE

Illustrative scenario: a training investment decision. A logistics company is deciding whether to fund a certification in supply-chain analytics for 50 planners. The certification is general: competitors value it. Human capital theory predicts that, without safeguards, the company will lose some trained planners to rivals and capture little return. The company has three options consistent with the theory. It can share costs (employees contribute time or accept a repayment agreement if they leave within a defined period). It can bundle the general certification with firm-specific training on its own planning systems, raising the specific component and the value of staying. Or it can rely on labour market frictions, such as a strong employer reputation and internal promotion prospects, that make trained employees

less likely to leave (Acemoglu and Pischke 1999). A simple decision rule follows: fund general training fully only when expected retention multiplied by the productivity gain exceeds the cost.

15.13 Tournament Theory

THEORY CARD

Core claim. When pay depends on relative rank rather than absolute output, the size of the prize gap between ranks motivates effort. Large pay jumps at the top of a hierarchy can therefore be efficient incentives for those competing for promotion, not rewards for the winner's own marginal product.
Origins. Personnel economics. Lazear and Rosen (1981), reviewed for management by Connelly et al. (2014).
Level of analysis. Contestants within a firm or market, and the firm's pay structure.
Key constructs. Rank-order pay, prize spread, noise (luck), number of contestants, sabotage, pay dispersion.
Boundary conditions. Works best when relative performance is easier to observe than absolute output and when contestants are similar in ability. Weakens or backfires when cooperation is essential or sabotage is possible.
Related theories. Agency theory (Section 6.1), Equity theory (Section 19.3), Upper echelons theory (Section 6.4), Matthew effect (Section 30.8).

15.13.1 Core logic and propositions

Lazear and Rosen (1981) asked why compensation is often based on rank: vice presidents compete for one chief executive position, associates for partnership, salespeople for a single President's Club. Their answer is that rank-order contracts can provide efficient incentives when output is costly to measure precisely but ordinal comparisons are cheap. In a tournament, the winner receives a prize fixed in advance, so the incentive to work depends on the **spread** between the winner's and loser's prizes and on how much effort raises the probability of winning.

In a simple two-player tournament, output equals effort plus random noise, and each contestant chooses effort e to maximise expected prize minus cost $c(e)$. In a symmetric equilibrium, the first-order condition is

$$\Delta W \cdot g(0) = c'(e^*)$$

where ΔW is the prize spread and $g(0)$ is the density of the difference in the two players' noise terms at zero. Effort rises with the spread and falls as noise rises, because a noisier contest makes effort matter less for winning. Figure 15.9 plots this relationship.

- **P1 (Spread).** Effort increases with the spread between prizes.
- **P2 (Noise).** For a given spread, effort falls as luck becomes more important, so firms must offer larger spreads in noisier settings.
- **P3 (Hierarchy).** Because the top prize must motivate all who compete for it, pay jumps grow larger at higher levels of the hierarchy.

- **P4 (Costs).** Large spreads encourage sabotage, collusion, and excessive risk taking, and they can undermine cooperation (Lazear 1989).

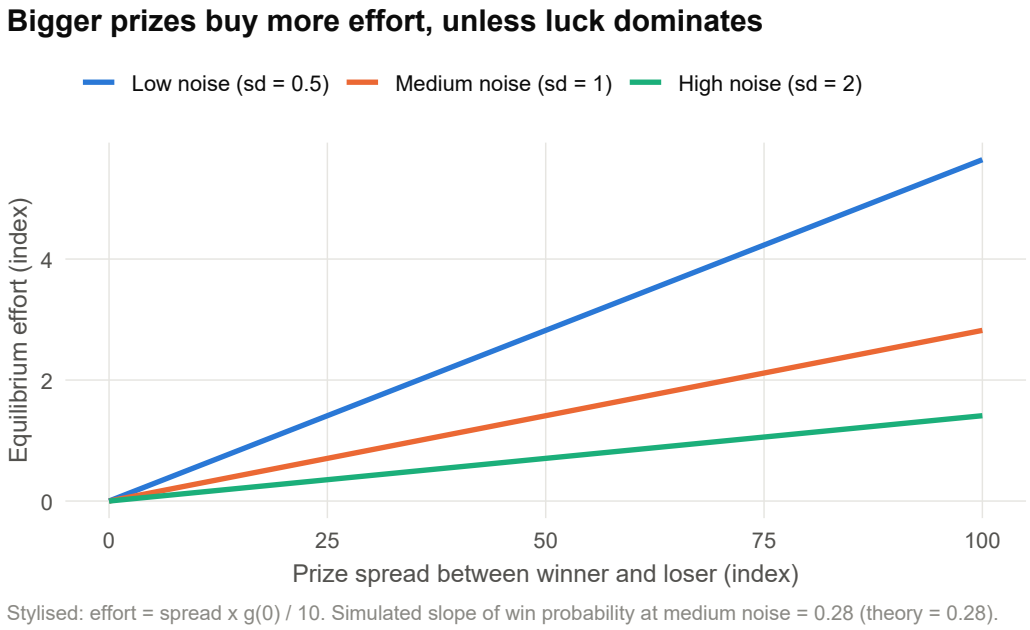


Figure 15.9. Stylised tournament model. Equilibrium effort rises linearly with the prize spread, but the slope is flatter when outcomes are noisier. Contestants with quadratic effort costs and normally distributed noise. Analytical solution of the Lazear-Rosen first-order condition, not empirical data.

15.13.2 Evidence and extensions

Sport offers clean tests because prize structures are public and performance is measured precisely. R. G. Ehrenberg and Bognanno (1990) found that professional golfers' scores were better in tournaments with larger total prizes and when the marginal return to improving one's rank was higher. B. E. Becker and Huselid (1992) found that larger prize spreads in professional auto racing were associated with better individual performance and also with riskier driving, illustrating both the incentive effect and the risk-taking cost.

In corporations, Main, O'Reilly, and Wade (1993) found that executive pay rises sharply with hierarchical level, consistent with tournament logic, and Kale, Reis, and Venkateswaran (2009) found that the pay gap between a chief executive and the next tier of executives was positively related to firm performance, as the theory predicts if the gap motivates the executives who compete for the top job. The broader review by Connelly et al. (2014) documents support for incentive effects alongside substantial evidence of costs: sabotage, reduced helping, unethical behaviour, and lower collaboration in tasks that require interdependence. These costs echo equity theory (Section 19.3) and justice research, which predict that large, unexplained pay gaps can be perceived as unfair by those who lose.

DEBATE AND EVIDENCE

Tournament incentives or managerial power? A competing explanation for large chief executive pay premiums is that powerful executives influence their own pay, so the gap reflects

bargaining power rather than optimal incentives for those below. The two explanations predict different patterns: tournament theory predicts that the gap should matter for the *vice presidents'* effort and should weaken when the next chief executive is likely to be hired externally, while the power explanation predicts that the gap should track governance weakness. Studies that exploit succession events and board composition help to separate the two (Connelly et al. 2014).

BUSINESS CASE

CEO pay tournaments and the pay-ratio disclosure. Since 2018, US public companies have disclosed the ratio of chief executive pay to the pay of the median employee, under a rule the Securities and Exchange Commission adopted to implement the Dodd-Frank Act. Tournament theory helps boards interpret their numbers. A large gap between the chief executive and the senior team can motivate senior executives *if* they realistically expect to compete for the top job, which is more likely when the firm has a credible internal succession process. If the board routinely hires chief executives from outside, the prize is out of reach for insiders and the gap buys little effort while signalling inequity. The same logic applies lower in the organization: forced-ranking systems and winner-take-all sales contests raise effort when tasks are independent and lucky outcomes are rare, but they damage cooperation in interdependent teams (Connelly et al. 2014).

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The pay gap between the chief executive and the next tier of executives is positively related to firm performance when internal succession is likely, and unrelated when external succession is likely.
- *H2.* Larger prize spreads in internal sales contests increase individual sales but reduce help given to colleagues.
- *H3.* The incentive effect of prize spread is weaker in contests with higher outcome variance.

Measures. Pay gaps between hierarchical levels from proxy statements. Number of contestants (for example, the number of vice presidents). Succession type. Performance variance.
Designs. Panel data on executive compensation with firm fixed effects. Natural experiments in sport, sales contests, and innovation contests. Laboratory tournaments that vary spread, noise, and the opportunity to sabotage.
Contexts. Executive compensation, professional services partnerships, sales organizations, innovation challenges, promotion systems.

COMMON PITFALL

Ignoring the losers. Tournament effects are often evaluated by looking at the winner. The theory's incentive effect operates on the contestants who do not win, and its costs (sabotage, exit, reduced cooperation) also fall mostly on them. Evaluate a tournament by the behaviour of the whole field.

15.14 Chapter Summary

KEY TAKEAWAYS

- Job design shapes motivation: autonomy, feedback, and meaningful work raise internal motivation, although the multiplicative MPS adds little over a simple average (Hackman and Oldham 1976, Humphrey, Nahrgang, and Morgeson 2007).
- **Job demands-resources theory** explains burnout and engagement as two processes, one driven by demands and one by resources, with resources buffering strain (Bakker and Demerouti 2007). **Conservation of resources theory** adds that loss weighs more than gain and that resources spiral (Hobfoll 1989).
- **Affective events theory** shifts attention to daily events and within-person emotional dynamics, separating affect-driven from judgment-driven behaviour (Weiss and Cropanzano 1996).
- Leadership theories differ in their unit and mechanism: **LMX** focuses on dyadic exchange (Graen and Uhl-Bien 1995), **transformational leadership** on inspiration and identification, with serious measurement critiques (Knippenberg and Sitkin 2013), and **path-goal theory** on matching behaviour to the situation (House 1971).
- The employment relationship rests on beliefs about promises (**psychological contracts**), perceptions of **justice**, and the **commitment** and **fit** that result (Rousseau 1989, Colquitt et al. 2001, J. P. Meyer and Allen 1991, Kristof 1996).
- **Human capital theory** explains who should pay for training and why firm-specific skills bind employer and employee (G. S. Becker 1964). **Tournament theory** explains why pay jumps grow up the hierarchy and when they backfire (Lazear and Rosen 1981).

Table 15.10 presents the comparison of the chapter's theories.

Table 15.10. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Typical business use
Job characteristics model	Psychological states from job features	Enriched jobs raise internal motivation	Job redesign
Job demands-resources	Health impairment and motivation processes	Resources buffer the strain of demands	Burnout prevention, engagement surveys
Conservation of resources	Primacy of loss, resource spirals	Loss of resources is disproportionately stressful	Change management, well-being support
Affective events theory	Events trigger emotions that drive behaviour	Daily events predict daily behaviour	Service work design, recovery breaks
Leader-member exchange	Dyadic social exchange	High-quality relationships raise performance	Manager training, onboarding
Transformational leadership	Identification and value change	Inspiration adds to transactional effects	Leadership development
Path-goal theory	Clarifying paths, raising valences	Best style depends on task and follower	Situational coaching
Psychological contract	Beliefs about reciprocal promises	Perceived breach reduces trust and effort	Change communication, employer branding
Organizational justice	Fairness of outcomes, process, treatment	Fair process softens bad outcomes	Layoffs, appraisal, pay decisions
Organizational commitment	Desire, need, obligation	Affective commitment predicts effort, not just staying	Retention strategy

Table 15.10. Comparison of the chapter's theories. *(continued)*

Theory	Key mechanism	Signature prediction	Typical business use
Person-environment fit	Congruence of person and setting	Fit predicts satisfaction and commitment	Selection, culture fit
Human capital theory	Investment in skills	Workers pay for general, share firm-specific training	Training investment
Tournament theory	Rank-order prizes	Effort rises with prize spread	Executive pay, promotion systems

15.15 Discussion Questions

1. The job characteristics model and job demands-resources theory both treat autonomy as beneficial. Under what conditions might autonomy act as a demand rather than a resource? Formulate a testable hypothesis.
2. Design a study that could separate the effect of leader-member exchange on performance from the effect of performance on leader-member exchange.
3. Evaluate the critique of transformational leadership by van Knippenberg and Sitkin (2013). Which of its arguments would a well-designed field experiment answer, and which would it not?
4. A firm must close one of three plants. Using organizational justice and psychological contract theory, draft a communication and implementation plan, and explain how you would measure its effects on survivors.
5. Human capital theory predicts that workers pay for general training. Why do many employers nonetheless fund degrees and certifications? Which extension of the theory best explains your answer?
6. A board is considering widening the pay gap between its chief executive and the senior team. Using tournament theory, list the conditions under which this will raise performance and the evidence you would need to see first.

15.16 Further Reading

- Hackman and Oldham (1980): the classic book-length treatment of job design, still useful for its practical diagnosis of jobs.
- Humphrey, Nahrgang, and Morgeson (2007): the best quantitative synthesis of work design research, extending the job characteristics model to social and contextual features.
- Bakker and Demerouti (2017): a concise statement of job demands-resources theory, including reciprocal and personal-resource extensions.
- Halbesleben et al. (2014): a critical review of what counts as a resource in conservation of resources theory.
- Graen and Uhl-Bien (1995): the authors' own account of how LMX theory developed and where it should go.
- Knippenberg and Sitkin (2013): an essential critique for anyone planning research on charismatic or transformational leadership.
- Colquitt et al. (2001): the meta-analysis that organised the organizational justice literature around four dimensions.

- Connelly et al. (2014): a thorough review of tournament theory across economics, management, and sport.

Part III

Markets, Finance, and Economics

Chapter 16

Finance and Corporate Policy

In 1976 a small fund company in Pennsylvania launched a mutual fund that promised never to beat the market. It would simply hold the stocks in the S&P 500 in proportion to their market value and charge very little for doing so. The offering raised far less than its sponsors had hoped, and many in the industry regarded it as a curiosity. Half a century later, index funds are among the largest pools of capital in the world, and the firm that launched that first retail index fund, Vanguard, is one of the largest asset managers anywhere. Why would investors deliberately give up the chance to outperform? Why do cash-rich companies borrow billions of dollars they do not seem to need? Why did a hedge fund whose partners included two Nobel laureates nearly collapse in 1998, and why did the shares of a struggling video-game retailer rise more than tenfold in a few weeks of January 2021?

Each puzzle sits on a different branch of financial theory, but they share a root. Finance studies how people and firms trade claims on uncertain future cash flows, and almost every theory in the field is an answer to one of three questions. **How should risk be priced?** (portfolio theory and the capital asset pricing model). **How much do prices reveal?** (the efficient market hypothesis, the limits to arbitrage, and behavioural finance). **How should a firm finance itself and what should it do with its cash?** (the Modigliani-Miller theorems, the trade-off and pecking order theories, and free cash flow theory). This chapter develops all eight, with formal logic where it is standard and applications to investing and corporate policy. They build on Agency theory (Section 6.1), Information asymmetry (Section 8.1), and Signaling theory (Section 8.2), and the behavioural theories connect to Prospect theory (Section 24.2) and Heuristics and biases (Section 23.2) (Table 16.1).

Table 16.1. Chapter at a glance: theories of finance and corporate policy.

Theory	Core question	Level of analysis	Key source
Portfolio theory	How should an investor combine risky assets?	Investor / portfolio	Markowitz (1952)
Capital asset pricing model	Which risks earn a premium in equilibrium?	Asset / market	Sharpe (1964), Lintner (1965)
Efficient market hypothesis	How fully do prices reflect available information?	Market	Fama (1970)
Modigliani-Miller theorems	When does financing affect firm value?	Firm / market	Modigliani & Miller (1958, 1963)
Trade-off and pecking order	Why do firms choose the leverage they do?	Firm	Kraus & Litzenberger (1973), Myers (1984)

Table 16.1. Chapter at a glance: theories of finance and corporate policy. *(continued)*

Theory	Core question	Level of analysis	Key source
Free cash flow theory	Why can surplus cash destroy value?	Firm / manager	Jensen (1986)
Limits to arbitrage	Why can mispricing persist?	Arbitrageur / market	Shleifer & Vishny (1997)
Behavioral finance	How do investor psychology and sentiment shape prices?	Investor / market	Barberis & Thaler (2003)

16.1 Portfolio Theory

THEORY CARD

Core claim. Investors should evaluate an asset not by its own risk but by its contribution to the risk of the whole portfolio. Combining assets whose returns are imperfectly correlated reduces risk without sacrificing expected return, and rational investors choose among portfolios on the **efficient frontier**.
Origins. Economics and finance. Markowitz (1952) formulated mean-variance portfolio selection. Tobin (1958) added the separation theorem. Markowitz shared the 1990 Nobel Memorial Prize in Economic Sciences with Merton Miller and William Sharpe.
Level of analysis. Individual investor or institutional portfolio.
Key constructs. Expected return, variance, covariance, correlation, diversification, efficient frontier, global minimum-variance portfolio, tangency portfolio, Sharpe ratio.
Boundary conditions. Investors care only about the mean and variance of portfolio returns (returns are approximately normal or utility is quadratic), inputs are known or well estimated, and the investment horizon is a single period.
Related theories. CAPM, Myopic loss aversion (Section 24.7), Mental accounting (Section 23.5), Behavioral finance.

16.1.1 Origins and intellectual history

Before 1952, investment advice focused on picking good securities: identify the asset with the highest expected value and buy it. Markowitz (1952) observed that this rule cannot explain why prudent investors hold many securities rather than one. Diversification is rational only if investors also care about risk, and once they do, the right unit of analysis is the portfolio. Markowitz measured risk as the **variance** of portfolio returns, which depends on the **covariances** among assets. Tobin (1958) then showed that with a riskless asset, every investor should hold the same risky portfolio and adjust overall risk by mixing it with the riskless asset. This **separation theorem** bridged portfolio choice and equilibrium asset pricing.

16.1.2 Core logic and propositions

For a portfolio with weights w_i on N assets, expected return is a weighted average, but variance is not:

$$E[R_p] = \sum_{i=1}^{N} w_i \, E[R_i], \qquad \sigma_p^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_i w_j \sigma_{ij}.$$

With two assets the variance becomes $\sigma_p^2 = w^2\sigma_1^2 + (1-w)^2\sigma_2^2 + 2w(1-w)\rho_{12}\sigma_1\sigma_2$. Whenever the correlation ρ_{12} is below one, the portfolio's standard deviation is less than the weighted average of the two standard deviations. That gap is the gain from diversification. For an equally weighted portfolio of N assets with average variance $\bar{\sigma}^2$ and average covariance $\overline{\text{cov}}$,

$$\sigma_p^2 = \frac{1}{N}\,\bar{\sigma}^2 + \left(1 - \frac{1}{N}\right)\overline{\text{cov}}.$$

As N grows, the first term vanishes but the second does not. Risk that is specific to individual assets (**idiosyncratic** risk) can be diversified away. Risk that assets share (**systematic** risk) cannot. Figure 16.1 plots this decomposition.

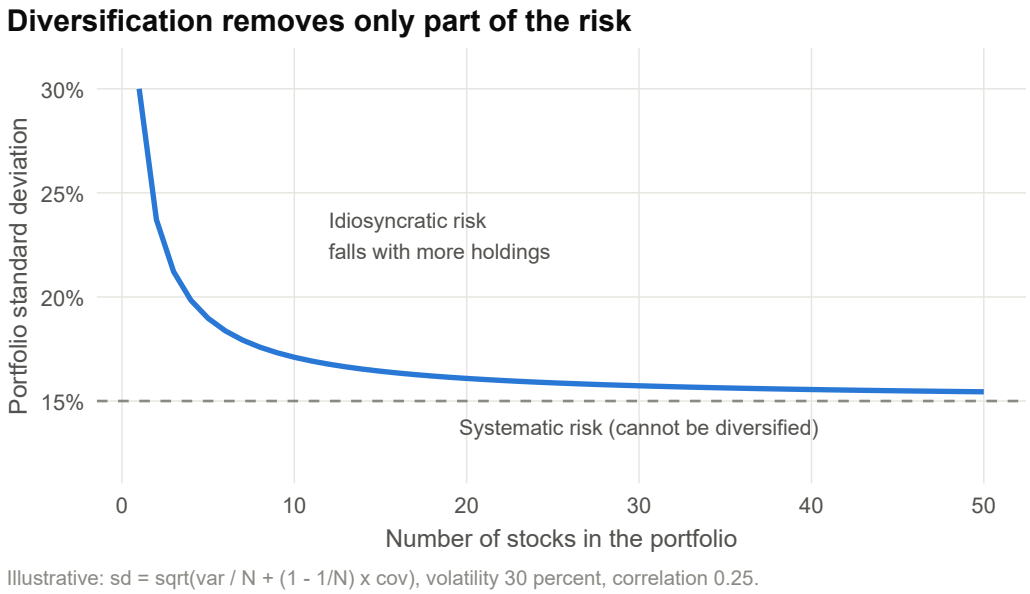


Figure 16.1. Illustrative diversification curve for an equally weighted portfolio. Each stock has 30 percent volatility and pairwise correlation 0.25. Idiosyncratic risk falls quickly as stocks are added, but systematic risk sets a floor.

The theory's central propositions can be stated as follows.

- **P1 (Risk is a portfolio property).** The risk an asset adds to a portfolio depends on its covariance with the other holdings, not on its stand-alone variance.
- **P2 (Diversification).** Combining imperfectly correlated assets lowers portfolio variance for a given expected return. Only systematic risk remains in a large, well-diversified portfolio.
- **P3 (Efficiency).** A rational mean-variance investor holds a portfolio on the upper boundary of the feasible set: no other portfolio offers a higher expected return for the same variance.

- **P4 (Separation).** With a riskless asset, every mean-variance investor holds the same risky portfolio, the **tangency portfolio** with the highest Sharpe ratio $(E[R_p]-r_f)/\sigma_p$, and expresses risk tolerance only through the share invested in it (Tobin 1958).

16.1.3 Mechanism

Figure 16.2 builds the frontier from five simulated asset classes. The blue curve is the mean-variance frontier, whose upper arm is the efficient frontier. The orange **capital allocation line** runs from the riskless rate to the tangency portfolio and dominates every other point on the frontier.

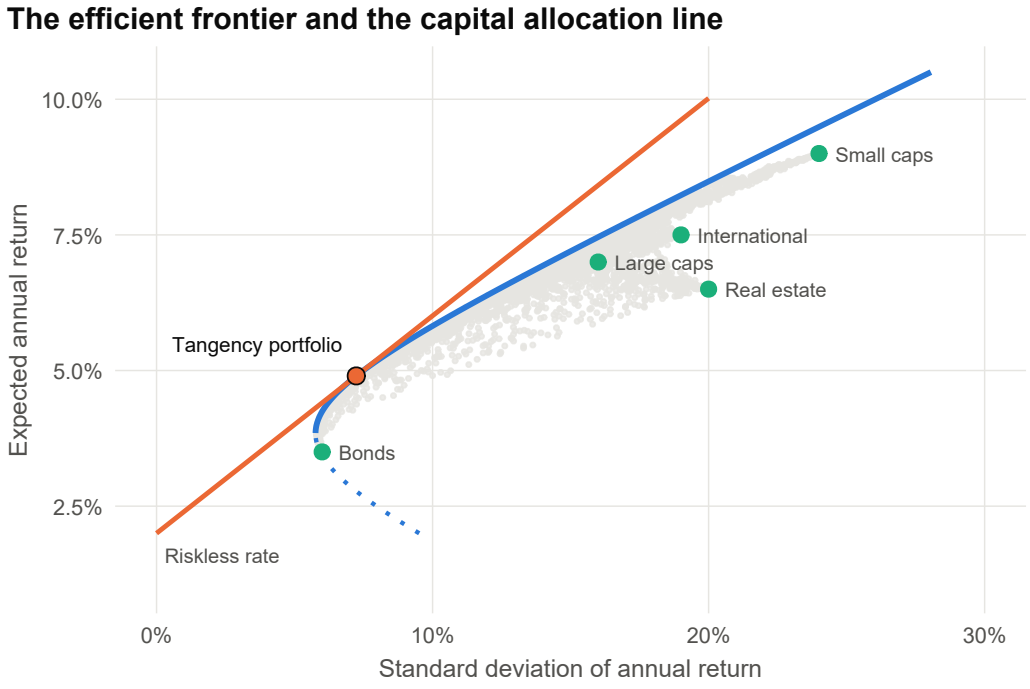


Figure 16.2. Illustrative efficient frontier for five simulated asset classes. Grey points are random long-only portfolios, the blue curve is the mean-variance frontier (short sales allowed), and the orange line combines the riskless asset with the tangency portfolio.

The lowest point on the frontier, the **global minimum-variance portfolio**, has a closed-form solution that requires no estimate of expected returns at all: $w_{\text{GMV}} = \Sigma^{-1}\mathbf{1}/(\mathbf{1}'\Sigma^{-1}\mathbf{1})$, where Σ is the covariance matrix. The short computation below applies it to three illustrative asset classes.

```
# Illustrative covariance matrix of annual returns (volatility 6, 16, 24 percent)
Sigma3 <- matrix(c(0.0036, 0.0019, 0.0029,
                   0.0019, 0.0256, 0.0243,
                   0.0029, 0.0243, 0.0576),
                 nrow = 3, byrow = TRUE,
                 dimnames = list(c("Bonds", "Large caps", "Small caps"),
                                 c("Bonds", "Large caps", "Small caps")))
x <- solve(Sigma3, rep(1, 3))                # inverse(Sigma) times a vector of ones
```

```
w_gmv <- x / sum(x)                          # normalise so the weights sum to one
round(w_gmv, 3)
```

```
#>      Bonds Large caps Small caps
#>      0.935      0.089     -0.025
```

```
sqrt(drop(t(w_gmv) %*% Sigma3 %*% w_gmv))    # volatility of the minimum-variance portfolio
```

```
#> [1] 0.0588677
```

The solution puts about 93 percent in bonds, a small long position in large caps, and a small short position in small caps. Its volatility is slightly *below* that of bonds alone, even though every other asset is riskier than bonds. That is Markowitz's insight in miniature: a risky asset can lower portfolio risk when its correlation with the rest of the portfolio is low enough.

16.1.4 Evidence and extensions

Portfolio theory is normative, so its evidence concerns whether its prescriptions work. The gains from diversification are real and large, but the optimiser is very sensitive to its inputs: it loads heavily on the assets whose expected returns are most overestimated. DeMiguel, Garlappi, and Uppal (2009) compared fourteen optimised strategies with a naive rule that puts $1/N$ in each asset and found that none consistently beat the naive rule out of sample across the datasets they studied, because estimation error offset the theoretical gains.

DEBATE AND EVIDENCE

Is variance the right measure of risk? Mean-variance analysis treats upside and downside deviations symmetrically and assumes that the first two moments capture everything investors care about. Real return distributions have fat tails, and correlations tend to rise in crises, which is exactly when diversification is most needed. Behavioural evidence adds that investors weigh losses more heavily than gains (Kahneman and Tversky 1979, N. C. Barberis 2013), see Loss aversion (Section 24.4). The pragmatic response has been shrinkage estimators, factor risk models, and weight constraints, which keep Markowitz's architecture while taming its sensitivity to inputs.

16.1.5 Applications in business

BUSINESS CASE

Concentration in employer stock: the Enron 401(k) accounts. The clearest violation of portfolio theory in household finance is holding a large share of retirement savings in one's own employer, whose wages and stock share the same firm-specific risk. When Enron filed for bankruptcy in December 2001, many employees whose 401(k) accounts were heavily invested in company stock lost much of their retirement savings along with their jobs. The episode fed into later US reforms that gave plan participants the right to diversify out of employer stock.

16.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Households with lower financial literacy hold less diversified portfolios, measured by the share of idiosyncratic variance in portfolio variance.
- *H2.* Employees hold more employer stock when the employer has recently outperformed (extrapolation), controlling for plan design.

Measures. Portfolio variance decomposition, number of effective holdings (inverse Herfindahl of weights), Sharpe ratio, tracking error, out-of-sample volatility.

Designs and data. Out-of-sample backtests with rolling estimation windows, household brokerage and pension records, administrative data on retirement plans, natural experiments from changes in default investment options.

COMMON PITFALL

Treating historical averages as expected returns. Plugging sample means into an optimiser produces portfolios that are extremely concentrated and unstable, because the optimiser chases estimation noise. Report out-of-sample performance, use shrinkage or factor models for inputs, and compare against the naive $1/N$ and minimum-variance benchmarks before claiming that an optimised portfolio adds value.

16.2 Capital Asset Pricing Model

THEORY CARD

Core claim. In equilibrium, an asset's expected return depends only on its **beta**, its sensitivity to the return on the market portfolio. Idiosyncratic risk earns no premium because investors can diversify it away.

Origins. Finance. Sharpe (1964) and Lintner (1965), building on Markowitz (1952) and Tobin (1958). Extended to multifactor models by Fama and French (1992) and Fama and French (1993).

Level of analysis. Individual asset within a market in equilibrium.

Key constructs. Market portfolio, beta, market risk premium, security market line, alpha, systematic versus idiosyncratic risk, factor premia.

Boundary conditions. Investors are mean-variance optimisers with common beliefs, can borrow and lend at the riskless rate, face no taxes or transaction costs, and hold the market portfolio of all risky assets.

Related theories. Portfolio theory, Efficient market hypothesis, Modigliani-Miller, Behavioral finance.

16.2.1 Origins and intellectual history

If every investor follows Markowitz and Tobin and all share the same beliefs, they all hold the same tangency portfolio. In equilibrium, that portfolio must be the **market portfolio**, because the supply of every asset must be held by someone. Sharpe (1964) and Lintner (1965) worked out the pricing

implication independently, and related versions were developed by Jack Treynor and Jan Mossin. The result was the first general equilibrium model of risk and return, and it gave both investors and corporate managers a usable formula for the cost of capital.

16.2.2 Core logic and propositions

The CAPM states that the expected excess return on any asset is proportional to its beta:

$$E[R_i] = r_f + \beta_i \left(E[R_m] - r_f \right), \qquad \beta_i = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{Var}(R_m)}.$$

An investor who holds the market portfolio cares about how a new asset changes the risk of that portfolio, which depends on the asset's covariance with the market. An asset that falls when the market falls must offer a higher expected return. Purely idiosyncratic risk adds almost nothing and earns only the riskless rate.

- **P1 (Only systematic risk is priced).** Idiosyncratic volatility should not predict expected returns.
- **P2 (Linearity).** Expected returns are linear in beta, with intercept r_f and slope equal to the market risk premium.
- **P3 (Zero alpha).** In equilibrium the intercept α_i in the regression of excess asset returns on excess market returns is zero for every asset.
- **P4 (Market efficiency of the market portfolio).** The market portfolio is mean-variance efficient. The CAPM is equivalent to this statement (Roll 1977).

16.2.3 Mechanism

The **security market line** (SML) in Figure 16.3 is the CAPM's testable prediction. Assets above the line (positive alpha) would be bid up and assets below it sold until both return to the line. The orange points illustrate the pattern that empirical tests have repeatedly found: a line that is **too flat**, with low-beta portfolios earning more and high-beta portfolios earning less than the CAPM predicts (Fama and French 2004).

16.2.4 Evidence and extensions

Early tests in the late 1960s and 1970s found that average returns rose with beta, but less steeply than the model predicts. Roll (1977) argued that the CAPM is not testable in practice, because the true market portfolio includes every risky asset and no index measures it. The decisive challenge came from Fama and French (1992): for US stocks from 1963 to 1990, beta had little power to explain average returns once firm size and the book-to-market ratio were included. Small firms and "value" firms with high book-to-market ratios earned more than their betas implied. Fama and French (1993) responded with a three-factor model in which returns load on the market, a size factor (SMB, small minus big), and a value factor (HML, high minus low book-to-market):

$$R_i - r_f = \alpha_i + b_i(R_m - r_f) + s_i\,\mathrm{SMB} + h_i\,\mathrm{HML} + \varepsilon_i.$$

Jegadeesh and Titman (1993) documented **momentum**, the tendency of recent winners to keep outperforming recent losers over three to twelve months, which the three-factor model does not

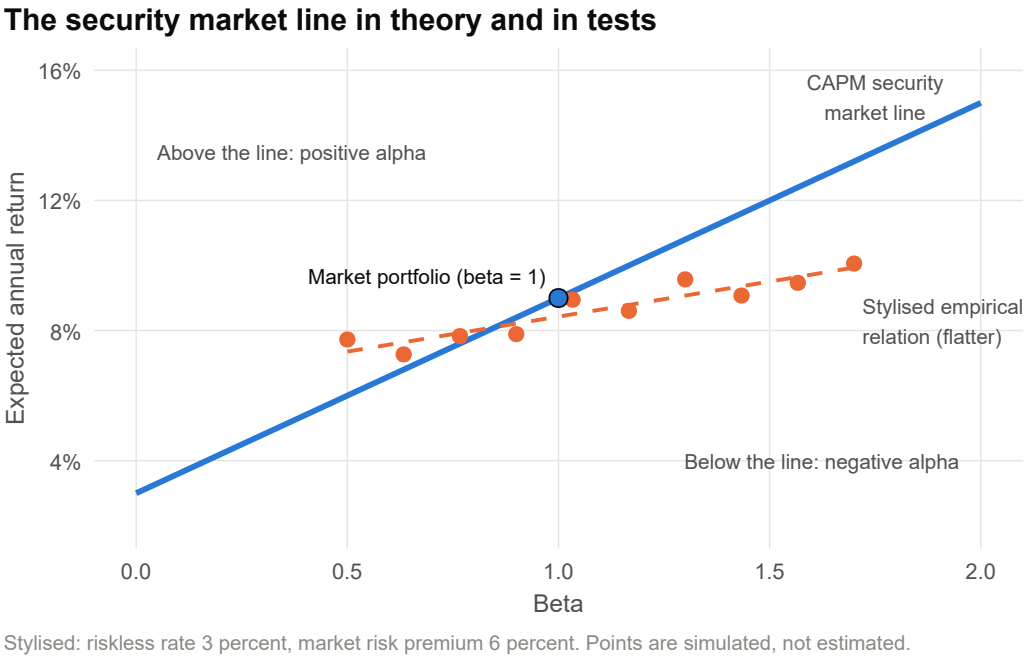


Figure 16.3. Stylised security market line. The blue line is the CAPM prediction. Orange points are stylised beta-sorted portfolios lying on a flatter line, the pattern reported in many empirical tests of the CAPM. Not empirical data.

explain. Carhart (1997) added a momentum factor for fund evaluation, and Fama and French (2015) added profitability and investment factors. Table 16.2 compares the main models.

Table 16.2. From the CAPM to multifactor models of expected returns.

Model	Risk factors	What it explains	Main limitation
CAPM (Sharpe 1964, Lintner 1965)	Market excess return	Why diversifiable risk is not rewarded	Flat empirical SML, size and value anomalies
Three-factor (Fama & French 1993)	Market, size (SMB), value (HML)	Size and value premia	Momentum, profitability
Four-factor (Carhart 1997)	Three factors plus momentum	Persistence in fund returns	Factors are empirical, not derived from theory
Five-factor (Fama & French 2015)	Market, size, value, profitability, investment	Profitability and investment patterns	Momentum, low-beta anomaly, short-term reversals

DEBATE AND EVIDENCE

Risk, mispricing, or data mining? The size and value premia admit two readings. In the risk reading, small and value firms are exposed to distress or business-cycle risk that the market beta misses, so their premia are rational compensation. In the behavioural reading, investors extrapolate past growth and overprice glamour firms, so the premia reflect correctable mispricing (N. Barberis and Thaler 2003). A third concern is statistical. Harvey, Liu, and Zhu (2016) counted hundreds of proposed return factors and argued that, given the number of researchers searching the same data, a new factor should clear a t-statistic hurdle of about 3.0 rather than the conventional 2.0. The CAPM survives because it is simple and coherent, but

few researchers regard it as an adequate empirical description of the cross-section of returns (Fama and French 2004).

16.2.5 Applications in business

In a large survey of chief financial officers, Graham and Harvey (2001) found that the CAPM was the most popular method for estimating the cost of equity, used by roughly three in four respondents. Investment managers use CAPM or multifactor alpha as the standard measure of skill, a practice that began with Jensen (1968).

BUSINESS CASE

Illustrative scenario: divisional hurdle rates. A diversified company has a regulated utility division and a biotechnology division and uses one company-wide cost of capital of 9 percent for all projects. Suppose the utility division's asset beta is about 0.4 and the biotechnology division's is about 1.5. With a riskless rate of 3 percent and a market risk premium of 6 percent, the CAPM implies hurdle rates of roughly 5.4 and 12 percent. The single company rate rejects good low-risk utility projects and accepts bad high-risk biotechnology projects. The remedy is divisional hurdle rates based on betas of comparable pure-play firms: the cost of capital belongs to the project, not the firm.

16.2.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* After controlling for market beta, firms with higher customer satisfaction earn lower subsequent returns because their cash flows are less risky.
- *H2.* The alpha of an investment strategy based on a new signal is not explained by the Fama-French five factors plus momentum.

Designs. Time-series regressions of portfolio excess returns on factors (alpha tests), Fama-MacBeth cross-sectional regressions of returns on characteristics and betas, portfolio sorts with out-of-sample periods.

Data. CRSP and Compustat for US equities, factor returns from Kenneth French's data library, international equity databases, mutual fund return databases.

COMMON PITFALL

Mistaking a factor loading for an explanation. Showing that a strategy's returns load on a factor does not explain *why* the factor earns a premium. Equally, an anomaly that disappears after controlling for factors may simply have been absorbed by a factor that is itself an unexplained anomaly. State whether a variable is meant to proxy for risk or for mispricing, and design a test that can tell the two apart.

16.3 Efficient Market Hypothesis

THEORY CARD

Core claim. In an efficient market, prices fully reflect available information, so no investor can earn returns in excess of fair compensation for risk by trading on that information.
Origins. Finance. Fama (1970) synthesised the theory and evidence, building on the random-walk literature. Fama shared the 2013 Nobel Memorial Prize with Lars Peter Hansen and Robert Shiller, one of the hypothesis's leading critics.
Level of analysis. Market.
Key constructs. Information set, weak, semi-strong, and strong form efficiency, abnormal return, joint hypothesis problem, event study, anomaly.
Boundary conditions. Information is widely available, trading costs are low, and enough well-capitalised investors trade on information to keep prices close to fundamental value.
Related theories. Limits to arbitrage, Behavioral finance, Wisdom of the crowd (Section 30.5), Information asymmetry (Section 8.1).

16.3.1 Origins and intellectual history

Statisticians had long noticed that stock price changes look like a random walk. The insight behind the EMH is that this is what a well-functioning market should produce: if prices already reflect all predictable information, only *new* information moves them, and new information is by definition unpredictable. Fama (1970) organised the evidence around three information sets, summarised in Table 16.3.

Table 16.3. The three forms of market efficiency (following Fama, 1970).

Form	Information reflected in prices	Implication if true	Typical test
Weak form	Past prices and trading volume	Technical analysis cannot earn abnormal returns	Return autocorrelation, trading-rule backtests
Semi-strong form	All publicly available information	Fundamental analysis of public data cannot earn abnormal returns	Event studies of announcements
Strong form	All information, including private	Even insiders cannot profit	Returns to insider trades and professional managers

16.3.2 Core logic and propositions

Formally, a market is efficient with respect to an information set Φ_t if the expected return conditional on that information equals the equilibrium expected return implied by some model of risk:

$$E\left[R_{i,t+1} \mid \Phi_t\right] = E^*\left[R_{i,t+1}\right], \qquad \text{so that} \quad E\left[R_{i,t+1} - E^*[R_{i,t+1}] \mid \Phi_t\right] = 0.$$

The abnormal return, the gap between realised and model-implied returns, should be unpredictable using Φ_t. This formulation exposes the **joint hypothesis problem**: every test of efficiency is also a test of the asset pricing model that defines E^*. If returns appear predictable, either the market is inefficient or the risk model is wrong, and the data alone cannot say which (Fama 1970, 1998).

- **P1 (No predictable abnormal returns).** Abnormal returns are unpredictable from information in the relevant set.
- **P2 (Rapid and unbiased adjustment).** Prices react to new public information quickly and, on average, by the right amount, with no systematic drift or reversal.
- **P3 (No persistent skill).** Professional investors do not outperform risk-adjusted benchmarks after costs on a persistent basis.

16.3.3 Mechanism

The EMH rests on competition among informed traders, who trade until the price no longer offers a profit. Grossman and Stiglitz (1980) showed that if prices fully reflected all information, no one would have an incentive to pay for information, because they could learn it for free from prices. Yet if no one gathered information, prices could not reflect it. The resolution is an equilibrium degree of *inefficiency*: prices must be noisy enough to compensate informed traders for their costs. Figure 16.4 shows this loop.

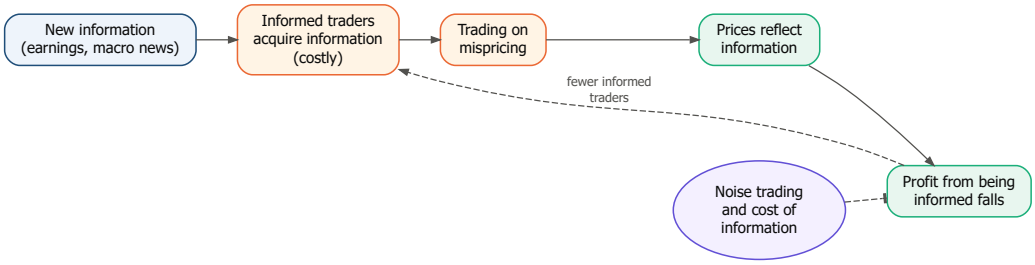


Figure 16.4. Mechanism of informational efficiency and the Grossman-Stiglitz paradox. Informed trading makes prices informative, which lowers the profit from becoming informed. In equilibrium, noise keeps prices informative but imperfect, so information gathering remains just worth its cost.

16.3.4 Evidence and extensions

Event studies, which measure abnormal returns around announcements, show that prices typically react within minutes to earnings, merger bids, and macroeconomic releases, consistent with P2. On P3, Jensen (1968) and Carhart (1997) found that mutual funds on average do not beat risk-adjusted benchmarks after fees, and that the persistence that exists is concentrated among the worst performers.

Against this, researchers have catalogued **anomalies**. The best documented is **post-earnings-announcement drift** (PEAD): after a large positive earnings surprise, returns continue to drift upward for weeks, and after a large negative surprise they drift downward. Bernard and Thomas (1989) showed that the drift is difficult to explain as compensation for risk and concluded that it reflects a delayed price response, as if investors fail to incorporate the full implications of earnings news. Figure 16.5 sketches the pattern.

A second line of evidence concerns the *level* of prices rather than the predictability of returns. If a stock's price is the expected present value of its future dividends, then the price p_t is a forecast of the ex post rational price $p_t^* = \sum_{k \geq 1} D_{t+k}/(1+r)^k$. A rational forecast must be less volatile than the quantity it forecasts, so $\mathrm{Var}(p_t) \leq \mathrm{Var}(p_t^*)$. Shiller (1981) found that US stock prices were far more volatile than this bound allowed, a result he interpreted as evidence that prices move too much to be justified by news about fundamentals.

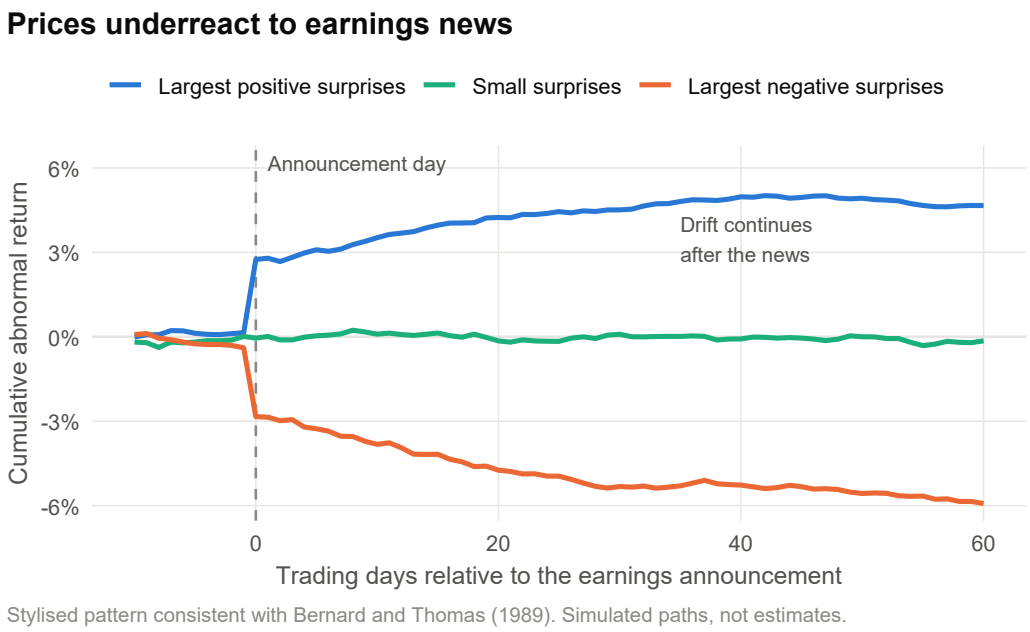


Figure 16.5. Stylised post-earnings-announcement drift. Cumulative abnormal returns jump at the announcement and then continue to drift in the direction of the surprise for about two months, a pattern inconsistent with complete and immediate adjustment. Not empirical data.

DEBATE AND EVIDENCE

Is the market efficient? The debate has matured rather than ended. Defenders emphasise that most anomalies are small, fragile, or explained by risk, that apparent overreactions are about as common as underreactions, and that active managers as a group do not beat passive benchmarks after costs (Fama 1998, Malkiel 2003). Critics reply that excess volatility, bubbles, and persistent anomalies show that prices can deviate from fundamental value for long periods (Shiller 1981, Shleifer 2000). A middle ground comes from McLean and Pontiff (2016): the returns to 97 published anomalies were about 26 percent lower out of sample and about 58 percent lower after publication. Some predictability was statistical bias, but much was real mispricing that informed traders then arbitraged away, as Grossman and Stiglitz would predict. The joint 2013 Nobel prize to Fama and Shiller captures the state of the field: markets are hard to beat, but prices are not always right.

16.3.5 Applications in business

BUSINESS CASE

Index funds and the rise of Vanguard. The most consequential application of the EMH is the index fund. P. A. Samuelson (1974) argued that few managers could be expected to beat the market and called for a low-cost fund that simply tracked a broad index. In 1976 John Bogle's newly founded Vanguard launched the First Index Investment Trust, later renamed the Vanguard 500 Index Fund, to track the S&P 500. The initial offering raised only a small fraction of its target. Sharpe (1991) later stated the logic with great economy: before costs, the average

actively managed dollar must earn the market return, so after costs it must underperform the average passive dollar. The argument requires only that outperformance be zero-sum before costs. French (2008) estimated that over 1980-2006 US investors spent about two thirds of one percent of the aggregate value of the market each year on fees, trading costs, and other expenses of active investing. Vanguard's structure reinforced the strategy: the company is owned by its funds, and through them by fund shareholders, which aligns its incentives with low costs. The later growth of passive investing raises a new question: do markets remain informationally efficient if fewer investors gather information? That is the Grossman-Stiglitz paradox in practical form.

16.3.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Post-earnings-announcement drift is stronger for firms with lower investor attention (fewer analysts, announcements on Fridays or busy news days).
- *H2.* Anomaly returns decline after the publication of the academic paper that documents them, and the decline is larger for anomalies that are cheaper to trade.

Designs. Event studies with market-model or factor-model abnormal returns, calendar-time portfolio regressions, difference-in-differences around changes in information access, tests of return predictability with out-of-sample validation.
Data. CRSP daily returns, I/B/E/S analyst forecasts and earnings surprises, news timestamps, SEC EDGAR filings.

COMMON PITFALL

Confusing "no free lunch" with "prices are right". The EMH predicts that abnormal returns are unpredictable, not that prices equal fundamental value at every moment. A market can be hard to beat and still be wrong, because mispricing may be unpredictable in timing or too risky to correct. Conversely, finding a predictable return pattern does not prove inefficiency until risk and trading costs are ruled out. Always state the asset pricing model your test assumes.

16.4 The Modigliani-Miller Theorems

THEORY CARD

Core claim. In a perfect capital market, a firm's value is determined by its real assets and investment policy, not by how it is financed. Capital structure and dividend policy are irrelevant. Once corporate taxes are introduced, debt adds value through the interest tax shield.
Origins. Economics and finance. Modigliani and Miller (1958) on capital structure, M. H. Miller and Modigliani (1961) on dividend policy, and Modigliani and Miller (1963) on corporate taxes. Modigliani received the Nobel Memorial Prize in 1985 and Miller in 1990.
Level of analysis. Firm in a capital market.
Key constructs. Firm value, leverage, cost of equity, cost of debt, weighted average cost of capital (WACC), homemade leverage, arbitrage, interest tax shield.

Boundary conditions. No taxes, bankruptcy costs, transaction costs, or information asymmetries, investors can borrow on the same terms as firms, and investment policy is held fixed. **Related theories.** Trade-off and pecking order, Free cash flow theory, Agency theory (Section 6.1), Signaling theory (Section 8.2).

16.4.1 Origins and intellectual history

Before 1958, corporate finance was largely descriptive and institutional, and practitioners believed that a judicious amount of cheap debt lowered a firm's cost of capital. Modigliani and Miller (1958) showed that this belief rests on an error. If investors can borrow and lend on the same terms as firms, any leverage a firm chooses can be undone or replicated by investors on their own account (**homemade leverage**). Two firms with identical assets but different capital structures must then have the same total value, or investors could earn a riskless profit by selling the overvalued firm and buying the undervalued one. The proof by arbitrage was itself a methodological innovation and became the template for modern asset pricing, including option pricing.

16.4.2 Core logic and propositions

The results are usually stated as three propositions and a tax correction. Let V_L and V_U be the values of levered and unlevered firms with the same assets, D and E the market values of debt and equity, r_A the return on assets, and r_D and r_E the returns on debt and equity.

- **Proposition I (Value irrelevance).** $V_L = V_U$. The value of the firm is independent of its capital structure.
- **Proposition II (Cost of equity).** The expected return on equity rises linearly with the debt-to-equity ratio, exactly offsetting the apparent gain from cheaper debt:

$$r_E = r_A + \frac{D}{E}\left(r_A - r_D\right), \qquad \text{WACC} = \frac{E}{V} r_E + \frac{D}{V} r_D = r_A.$$

- **Proposition III (Dividend irrelevance).** Holding investment fixed, the split of earnings between dividends and retention does not affect value, because a dividend financed by issuing new shares transfers value from old to new shareholders one for one (M. H. Miller and Modigliani 1961).
- **Tax correction.** With a corporate tax rate τ_c and permanent debt, interest payments are deductible and

$$V_L = V_U + \tau_c D, \qquad \text{WACC}_{\text{after-tax}} = r_A\left(1 - \tau_c \frac{D}{V}\right),$$

so value rises with leverage (Modigliani and Miller 1963).

16.4.3 Mechanism

Figure 16.6 shows the propositions graphically. Without taxes, substituting cheap debt for expensive equity leaves the WACC unchanged because equity holders demand a higher return for the extra financial risk (in CAPM terms, leverage raises the equity beta). With taxes, the WACC falls as leverage rises.

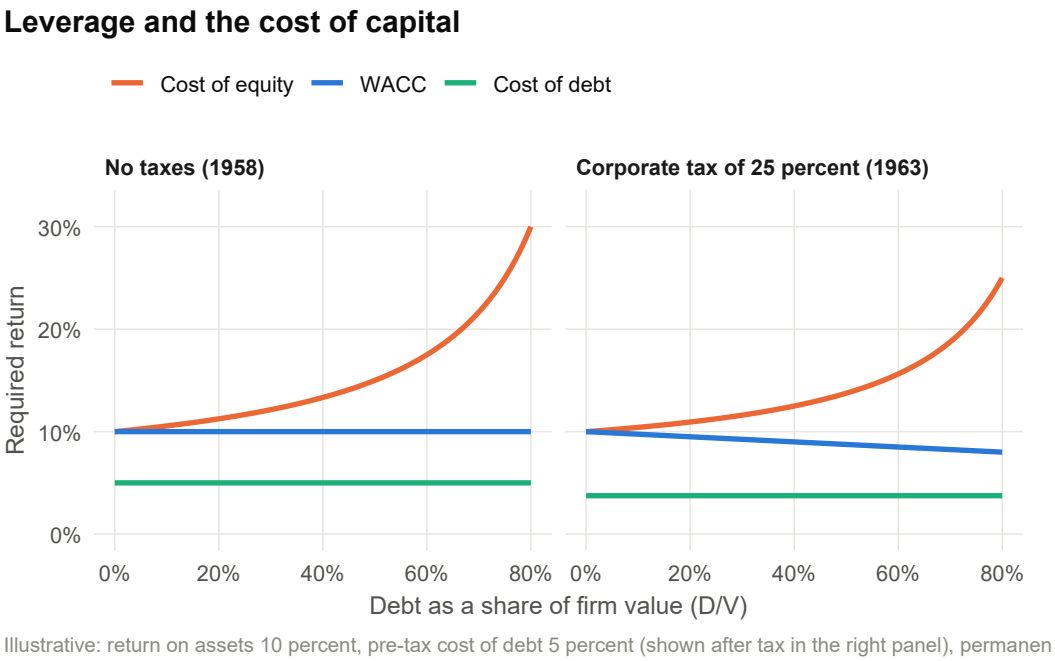


Figure 16.6. Illustrative Modigliani-Miller propositions. Without taxes, the cost of equity rises with leverage and the weighted average cost of capital stays flat. With a 25 percent corporate tax, the after-tax WACC declines as leverage rises because of the interest tax shield.

16.4.4 Evidence and extensions

The MM propositions are not empirical predictions in the usual sense. Their importance lies in what they imply about where to look. If financing matters, it must be because one of the assumptions fails. Table 16.4 maps the main research programs in corporate finance onto the assumptions they relax, which is why MM is often described as the foundation of the field rather than a description of it (Myers 2001).

Table 16.4. The Modigliani-Miller assumptions as a map of corporate finance.

MM assumption	What happens when it fails	Resulting theory	Key reference
No corporate taxes	Interest deductibility makes debt valuable	Tax shield, static trade-off	Modigliani & Miller (1963)
No costs of financial distress	Expected bankruptcy and distress costs offset tax benefits	Trade-off theory	Kraus & Litzenberger (1973)
Symmetric information	Equity issues signal overvaluation, internal funds are preferred	Pecking order theory	Myers & Majluf (1984)
No agency conflicts	Debt and payouts discipline managers with surplus cash	Free cash flow theory	Jensen (1986)
Investment policy fixed	Debt overhang causes underinvestment in good projects	Debt overhang	Myers (1977)
Prices equal fundamental value	Managers issue equity when it is overpriced	Market timing	Baker & Wurgler (2002)

DEBATE AND EVIDENCE

Irrelevance as a benchmark or as a claim? Critics sometimes dismiss MM as unrealistic. Its defenders reply that unrealistic assumptions are the point: the theorems isolate the conditions under which financing could matter, and every later theory in this chapter is a precise answer to the question "which friction?" The remaining debate is quantitative: how large are the tax benefits of debt after personal taxes, and how large are expected distress costs? Those magnitudes have proved hard to pin down, as the next section shows.

16.4.5 Applications in business

BUSINESS CASE

Taxes and the Apple bond issue of 2013. In April 2013 Apple sold 17 billion dollars of bonds, at the time one of the largest corporate bond offerings on record, even though it held well over 100 billion dollars in cash and securities. A large share of that cash was held by foreign subsidiaries, and bringing it home would have triggered US taxes under the rules then in force. Borrowing at low interest rates to fund dividends and share repurchases let Apple return cash to shareholders without paying the repatriation tax, and the interest was tax deductible. The case illustrates the MM logic in reverse: financing decisions create value precisely when a friction, here the tax code, breaks the irrelevance conditions.

16.4.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Exogenous increases in corporate tax rates lead firms to increase leverage, and the effect is larger for firms with fewer non-debt tax shields.
- *H2.* The equity beta of a firm rises with leverage as predicted by Proposition II after controlling for asset beta.

Designs. Natural experiments from tax reforms, staggered state tax changes, regression of equity betas on leverage, event studies of pure recapitalisations.

Data. Compustat fundamentals, CRSP returns, tax rate databases, bond issuance data (Mergent FISD).

COMMON PITFALL

Reading MM as "capital structure does not matter". The theorems say that capital structure is irrelevant *only* under perfect markets, which makes them a checklist of reasons why it does matter. A second misuse is to apply Proposition II's logic backward: a lower WACC from adding debt is not evidence of value creation unless the cost of equity has been adjusted for the higher financial risk.

16.5 Trade-off and Pecking Order Theories

THEORY CARD

Core claim. The **static trade-off theory** holds that firms choose a target leverage that balances the tax benefits of debt against the expected costs of financial distress. The **pecking order theory** holds that, because of information asymmetry, firms prefer internal funds to debt and debt to equity, so leverage is the cumulative result of past financing needs rather than a target.
Origins. Finance. Kraus and Litzenberger (1973) formalised the trade-off. Myers (1984) and Myers and Majluf (1984) developed the pecking order. Shyam-Sunder and Myers (1999) and Frank and Goyal (2003) set up the classic empirical contest.
Level of analysis. Firm.
Key constructs. Target leverage, interest tax shield, costs of financial distress, debt overhang, financing deficit, financial slack, adverse selection cost of equity, market timing.
Boundary conditions. Trade-off logic requires meaningful tax benefits and distress costs. Pecking order logic requires managers to know more than investors and to act in the interest of existing shareholders.
Related theories. Modigliani-Miller, Information asymmetry (Section 8.1), Signaling theory (Section 8.2), Free cash flow theory.

16.5.1 Origins and intellectual history

Once Modigliani and Miller (1963) showed that debt creates a tax shield, the obvious puzzle was why firms are not financed almost entirely with debt. Kraus and Litzenberger (1973) provided the standard answer in a state-preference model: debt creates tax savings in good states but deadweight bankruptcy costs in bad ones, so there is an interior optimum. Later work added the indirect costs of distress, such as lost customers and suppliers, and Myers (1977) identified **debt overhang**: a heavily indebted firm may forgo valuable projects because much of their payoff would go to creditors.

In his 1984 presidential address to the American Finance Association, published as "The Capital Structure Puzzle", Myers (1984) argued that the trade-off theory fits the data poorly. Many highly profitable firms, which have the most taxable income to shield, borrow the least. He proposed instead a **pecking order** grounded in the asymmetric information model of Myers and Majluf (1984). If managers know more than investors about the value of the firm, they will be most willing to issue equity when it is overvalued. Investors understand this and treat an equity issue as bad news, so issuing equity is expensive for good firms. Internal funds avoid the problem entirely, and safe debt nearly does. The result is a financing hierarchy.

16.5.2 Core logic and propositions

The trade-off theory can be written as

$$V_L = V_U + \text{PV}(\text{interest tax shields}) - \text{PV}(\text{costs of financial distress}),$$

with the optimal debt level where the marginal tax benefit equals the marginal expected distress

cost. The pecking order is usually tested with the financing deficit, the cash a firm must raise externally:

$$\mathrm{DEF}_t = \mathrm{DIV}_t + X_t + \Delta W_t + R_t - C_t, \qquad \Delta D_{it} = a + b_{\mathrm{PO}}\,\mathrm{DEF}_{it} + e_{it},$$

where DIV is dividends, X capital expenditure, ΔW the increase in working capital, R the current portion of long-term debt, and C operating cash flow after interest and taxes. Under a strict pecking order, debt fills the deficit one for one, so $a = 0$ and $b_{\mathrm{PO}} = 1$ (Shyam-Sunder and Myers 1999).

- **P1 (Trade-off: target).** Firms have target leverage ratios and adjust toward them.
- **P2 (Trade-off: cross-section).** Leverage rises with tax benefits and tangible assets and falls with distress costs, earnings volatility, and growth options.
- **P3 (Pecking order: hierarchy).** Firms use internal funds first, then debt, and issue equity only as a last resort.
- **P4 (Pecking order: profitability).** More profitable firms use less debt because they need less external finance.
- **P5 (Pecking order: announcement effects).** Announcements of equity issues are followed by negative abnormal returns, and the effect is larger when information asymmetry is greater.

16.5.3 Mechanism

Figure 16.7 shows the trade-off. The MM tax line rises without limit. Distress costs are negligible at low leverage and rise steeply as default becomes likely, so firm value peaks at an interior debt level.

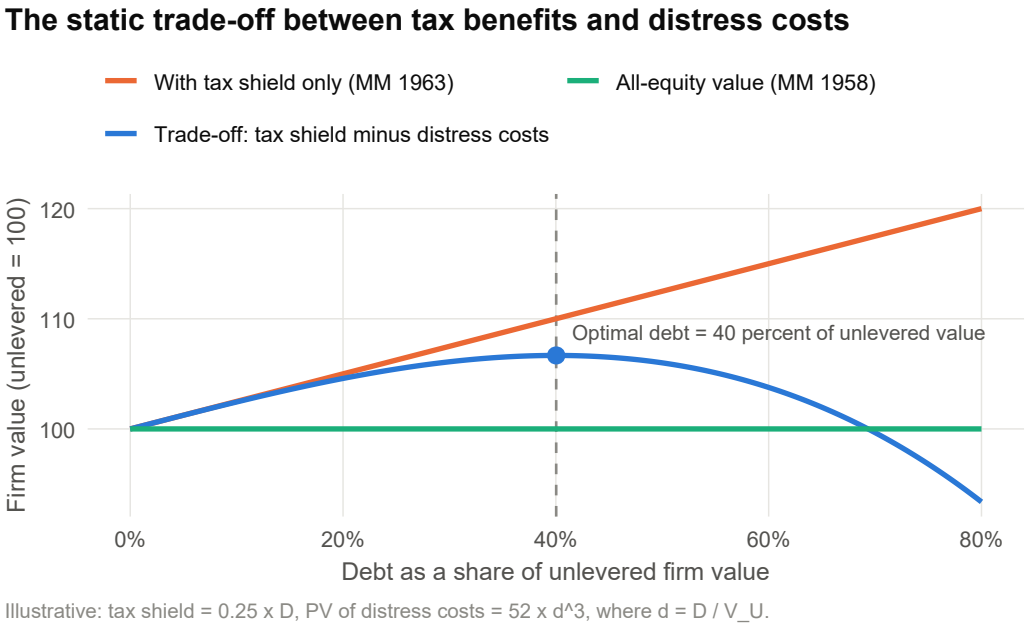


Figure 16.7. Illustrative static trade-off theory. The interest tax shield raises firm value linearly with debt, while the present value of financial distress costs rises with leverage. Firm value peaks at an interior optimum, marked by the dashed line.

Figure 16.8 summarises the pecking order mechanism. The cost of each source of funds rises with its sensitivity to managers' private information.

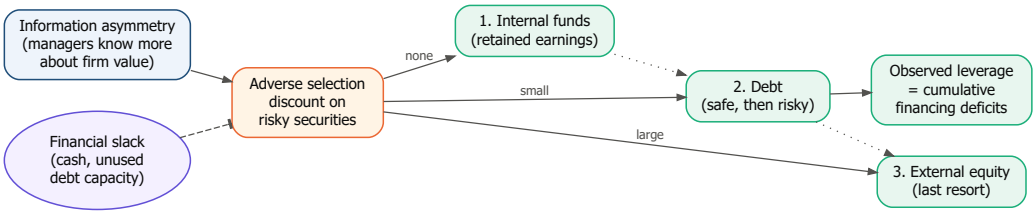


Figure 16.8. Mechanism of the pecking order theory. Because managers know more than investors, securities whose value depends more on that private information carry a larger adverse selection discount. Firms therefore finance with internal funds first, then debt, and equity last. Financial slack relaxes the constraint.

16.5.4 Evidence and extensions

The two theories make sharply different predictions, summarised in Table 16.5. The empirical contest has run for four decades without a clear winner.

Table 16.5. Contrasting predictions of the static trade-off and pecking order theories.

Issue	Static trade-off	Pecking order	What the evidence suggests
Target leverage	Yes, firms adjust toward it	No well-defined target	Mixed: adjustment is slow, leverage is highly persistent
Profitability and leverage	Positive (more taxable income to shield)	Negative (less need for external funds)	Negative, favouring the pecking order
Response to financing deficit	Partial, depends on distance from target	Debt fills the deficit one for one	Strong for large mature firms, weak for small growth firms
Equity issuance	When leverage is above target	Rare, a last resort	Small growth firms issue equity often
Market reaction to equity issues	Neutral	Negative	Negative on average
Role of stock prices	None	None directly	Firms issue equity after price run-ups (market timing)

Shyam-Sunder and Myers (1999) found that for a sample of large, mature US firms, net debt issues tracked the financing deficit closely, which favoured the pecking order, and that standard target-adjustment tests had little power to distinguish the two theories. Frank and Goyal (2003) extended the test to a much broader sample of US firms and found that net equity issues tracked the deficit more closely than net debt issues. Small, high-growth firms, where information asymmetry and hence the pecking order should be strongest, issued the most equity. The pecking order fit best for large firms.

Two further findings complicate both theories. Lemmon, Roberts, and Zender (2008) showed that leverage is remarkably persistent, with most of its variation explained by a stable firm effect that standard determinants do not capture. M. Baker and Wurgler (2002) argued that firms issue equity when their market values are high relative to book and past market values, and that these **market timing** decisions have long-lasting effects on capital structure. In survey evidence, CFOs stressed financial flexibility and credit ratings, and few described their leverage target as strict (Graham and Harvey 2001).

DEBATE AND EVIDENCE

Is there a capital structure theory at all? Myers (2001) concluded that there is no universal theory of the debt-equity choice, only conditional theories that work in different settings: trade-off logic for differences across industries, pecking order logic for the profitability effect and issue announcements, market timing for bursts of issuance in hot markets. A further critique is econometric: leverage ratios are mechanically mean-reverting, so finding "adjustment toward a target" does not by itself confirm the trade-off theory (Shyam-Sunder and Myers 1999).

16.5.5 Applications in business

BUSINESS CASE

Leverage decisions in practice. Firms in the same economy choose strikingly different capital structures, and the differences line up with trade-off logic. Regulated utilities and real estate investment trusts, whose cash flows are stable and whose assets are tangible and redeployable, routinely carry high leverage. Technology and biotechnology firms, whose value lies in growth options and intangible assets that would lose much of their value in distress, typically carry little net debt, and many hold large cash balances. Within those broad patterns, pecking order and timing behaviour are easy to observe. Announcements of seasoned equity offerings are, on average, followed by negative stock price reactions, consistent with Myers and Majluf (1984). Conversely, firms whose shares have run up issue equity opportunistically. In 2021, after the run-up in their share prices described in the Behavioral finance section, GameStop and AMC Entertainment both sold new shares through at-the-market offerings, using a high stock price to raise capital and reduce debt, which is market timing in the sense of M. Baker and Wurgler (2002).

16.5.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The negative relation between profitability and leverage is stronger for firms with greater information asymmetry (fewer analysts, more intangible assets).
- *H2.* Firms issue more equity when their market-to-book ratio is high relative to its historical average, controlling for investment opportunities.

Measures. Book and market leverage, financing deficit, market-to-book ratio, asset tangibility, marginal tax rates, distance to default, analyst forecast dispersion as a proxy for information asymmetry.

Designs. Partial adjustment models of leverage, financing deficit regressions, event studies of security issues, natural experiments from tax and bankruptcy law changes, surveys of CFOs.

COMMON PITFALL

Treating every cross-sectional correlation as a test of one theory. Profitability, size, tangibility, and growth opportunities are related to leverage under several theories at once, often

with the same sign. A negative profitability coefficient is consistent with the pecking order but also with mean reversion in leverage and with agency explanations. Specify which prediction is *unique* to the theory you are testing, and use settings where the theories diverge.

16.6 Free Cash Flow Theory

THEORY CARD

Core claim. When a firm generates more cash than it can invest in positive net present value projects, self-interested managers tend to waste the surplus on low-return growth rather than return it to shareholders. Debt, payouts, and the market for corporate control reduce these agency costs.
Origins. Finance and organisational economics. Jensen (1986), building on the agency framework of Jensen and Meckling (1976).
Level of analysis. Firm and top management.
Key constructs. Free cash flow, empire building, agency costs of equity, debt as a bonding device, payout policy, leveraged buyout, market for corporate control.
Boundary conditions. Firms have substantial cash flows and limited growth opportunities, and managerial incentives are imperfectly aligned with shareholders'. The theory has least force for young, high-growth firms with more good projects than cash.
Related theories. Agency theory (Section 6.1), Stewardship theory (Section 6.2), Upper echelons theory (Section 6.4), Modigliani-Miller.

16.6.1 Origins and intellectual history

Jensen's 1986 paper applied the agency framework of Jensen and Meckling (1976) to a specific question: what happens to cash a firm does not need? He defined **free cash flow** as cash flow in excess of that required to fund all projects with positive net present value at the relevant cost of capital. Managers, he argued, want firms to grow beyond their optimal size because size brings power, status, and pay, and free cash flow fuels that growth. Jensen used the theory to explain the takeovers and leveraged buyouts of the 1980s, and he pointed to the oil industry, where high prices in the late 1970s generated large cash flows that many companies reinvested in exploration and diversification even as the returns to such investment declined (Jensen 1986).

16.6.2 Core logic and propositions

- **P1 (Overinvestment).** Firms with high free cash flow and poor investment opportunities overinvest, for instance in diversifying acquisitions, and these investments earn low returns.
- **P2 (Debt as bonding).** Debt commits managers to pay out future cash flows, reducing the cash available for waste. Unlike a promise to pay dividends, debt payments are enforceable through bankruptcy.
- **P3 (Payouts).** Increases in dividends and repurchases by firms with free cash flow problems are good news for shareholders. Cuts are bad news.
- **P4 (Control market).** Takeovers and leveraged buyouts create value disproportionately in firms with large free cash flow and few growth opportunities.

16.6.3 Mechanism

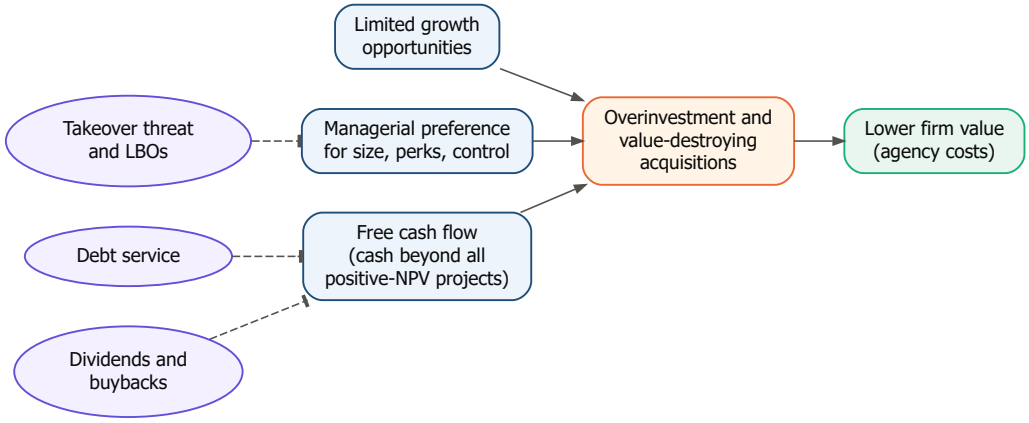


Figure 16.9. Mechanism of free cash flow theory. Surplus cash combined with weak growth opportunities and managerial empire-building incentives leads to overinvestment and value destruction. Debt, payouts, takeovers, and incentive pay reduce the cash available for waste.

The governance devices in Figure 16.9 differ in strength and cost, as Table 16.6 summarises. The trade-off between them echoes the general logic of Agency theory (Section 6.1): every disciplinary device is itself costly, and the efficient mix depends on how severe the agency problem is.

Table 16.6. Devices that discipline the use of free cash flow.

Device	How it reduces free cash flow problems	Commitment strength	Main cost or risk
Regular dividends	Pays out cash and creates an expectation of continuation	Moderate (cuts are penalised but legal)	Personal taxes, inflexibility in downturns
Share repurchases	Returns cash flexibly	Weak (no ongoing commitment)	Can be timed opportunistically or used to manage earnings per share
Debt	Contractually pre-commits future cash flows	Strong (default triggers bankruptcy)	Financial distress, debt overhang
Leveraged buyout	Combines high debt, concentrated ownership, and equity incentives	Very strong	Fragility if cash flows fall, high transaction costs
Hostile takeover threat	Replaces managers who waste cash	Depends on legal defences	Disruption, short-termism concerns
Equity-based pay	Aligns managers with shareholder value	Moderate	Risk-taking, manipulation of timing

16.6.4 Evidence and extensions

Kaplan (1989) studied management buyouts of public companies in the 1980s and found that operating income rose, capital expenditure fell, and net cash flow increased in the years after the buyout, consistent with the view that high leverage and concentrated ownership improved the use of cash. Harford (1999) found that cash-rich firms were more likely to attempt acquisitions and that those acquisitions were followed by abnormal declines in operating performance, which is the overinvestment the theory predicts. Managerial overconfidence (Malmendier and Tate 2005) and hubris (Roll 1986) predict the same overinvestment without conscious self-interest, a distinction that matters for remedies.

Payout policy is the natural testing ground. Lintner (1956) found that managers set dividends with a long-run target payout ratio in mind and adjust gradually, reluctant to raise dividends they might later have to cut. Half a century later, Brav et al. (2005) found that this reluctance remained strong and that repurchases had become the preferred flexible tool for distributing cash. Dividends can also signal private information about future cash flows (S. Bhattacharya 1979), and the signaling and free cash flow accounts are not mutually exclusive.

DEBATE AND EVIDENCE

Agency problem or precaution? Free cash flow theory reads large cash holdings as a symptom of agency problems. An alternative view is precautionary: firms with volatile cash flows, high R&D, and limited access to debt markets hold cash to avoid having to pass up good projects in bad times, a motive rooted in the pecking order. A second debate concerns buybacks. Critics argue that repurchases divert cash from investment and let managers time sales of their own shares. Defenders reply that returning cash that cannot be profitably invested is exactly what Jensen recommended. Adjudicating requires knowing whether the firm forwent positive NPV projects, which is hard to observe.

16.6.5 Applications in business

BUSINESS CASE

The RJR Nabisco buyout. The largest leveraged buyout of the 1980s, the 1989 acquisition of RJR Nabisco by Kohlberg Kravis Roberts, was widely interpreted through Jensen's lens. The company's tobacco business generated large, stable cash flows with limited reinvestment needs, which made it a classic free cash flow firm. The buyout replaced public equity with heavy debt, forcing the company to devote its cash flow to interest and principal and to sell assets. The episode, which followed a contested bidding process, became a public symbol both of the disciplinary logic of debt and of the risks of paying too much in a competitive auction.

BUSINESS CASE

Dividend and buyback policy at mature technology firms. As large technology companies matured, their free cash flow came to exceed what they could invest internally. Apple announced in 2012 that it would begin paying a dividend and repurchasing shares, and it has since returned very large amounts of cash through buybacks and dividends. Microsoft began paying dividends in 2003 and paid a large special dividend in 2004. Such policies combine free cash flow logic (return cash that cannot be invested at high returns), Signaling (Section 8.2) (a regular dividend conveys confidence in future cash flows), and flexibility (buybacks can be scaled without the penalty attached to a dividend cut). The practical checklist: estimate free cash flow after all positive NPV investment, commit a sustainable part through regular dividends, and use repurchases for the residual.

16.6.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The announcement return to a diversifying acquisition is more negative when the acquirer has high free cash flow and low Tobin's q.
- *H2.* Firms that increase leverage through recapitalisations subsequently reduce capital expenditure and improve operating margins, more so when prior free cash flow was high.

Measures. Free cash flow scaled by assets, Tobin's q as a proxy for growth opportunities, excess cash relative to a predicted benchmark, governance indices, payout ratios, acquisition announcements.

Designs. Event studies, panel regressions with firm fixed effects, natural experiments from changes in takeover law (such as state anti-takeover statutes) or dividend taxation, comparisons of buyout firms with matched controls.

COMMON PITFALL

Equating cash with waste. Not all large cash balances signal an agency problem. Before concluding that cash is "free", estimate the firm's investment opportunities, its access to external finance, and the volatility of its cash flows. The theory predicts waste only when cash exceeds what positive NPV projects and prudent precaution require.

16.7 Limits to Arbitrage

THEORY CARD

Core claim. Correcting mispricing requires capital and bears risk. Because real arbitrage is risky, capital constrained, and often delegated to professionals who answer to impatient investors, mispricing can persist and even widen before it corrects.

Origins. Finance. De Long et al. (1990) on noise trader risk and Shleifer and Vishny (1997) on performance-based arbitrage, synthesised in Shleifer (2000).

Level of analysis. Arbitrageur and market.

Key constructs. Fundamental risk, noise trader risk, implementation costs, short-sale constraints, performance-based arbitrage, forced liquidation, synchronisation risk.

Boundary conditions. Mispricing must be uncertain in duration or magnitude, arbitrage capital must be limited, and close substitutes for the mispriced asset must be imperfect or costly to trade.

Related theories. Efficient market hypothesis, Behavioral finance, Agency theory (Section 6.1), Collective action (Section 17.2).

16.7.1 Origins and intellectual history

The classical defence of efficiency, associated with Milton Friedman and with Fama (1970), holds that irrational traders cannot affect prices for long, because arbitrageurs trade against them and irrational traders lose money and disappear. De Long et al. (1990) challenged both halves of the

argument. If noise traders' beliefs are unpredictable, an arbitrageur betting against today's mispricing risks that it gets worse tomorrow. With short horizons, this **noise trader risk** limits arbitrage, so prices can diverge from fundamentals even with no fundamental risk at all. Noise traders may even earn higher average returns than rational investors because they bear more of the risk they themselves create.

Shleifer and Vishny (1997) added an organisational dimension. Arbitrage is conducted by specialised professionals managing other people's money, and investors judge them by recent returns. When a mispricing widens, the arbitrageur shows losses just as the opportunity becomes most attractive, investors withdraw capital, and the arbitrageur may be forced to liquidate at the worst moment. This is **performance-based arbitrage**, and it makes arbitrage least effective when it is most needed.

16.7.2 Core logic and propositions

In the overlapping generations model of De Long et al. (1990), a safe asset and an unsafe asset pay the same dividend, so the unsafe asset's fundamental value is 1. A fraction μ of investors are noise traders whose misperception of the expected return is ρ_t, normally distributed with mean ρ^* and variance σ_ρ^2. Rational investors have risk aversion γ, and r is the riskless rate. The equilibrium price is

$$p_t = 1 + \frac{\mu\,(\rho_t - \rho^*)}{1+r} + \frac{\mu\rho^*}{r} - \frac{2\gamma\mu^2\sigma_\rho^2}{r(1+r)^2}.$$

The second term moves the price with current noise trader sentiment, the third captures their average optimism or pessimism, and the last is a discount for the risk that noise traders' beliefs will change. Although the asset's fundamental value is certain, its price is volatile, and unless noise traders are optimistic enough on average it trades below fundamental value, because rational investors must be paid to bear the risk that noise traders create.

- **P1 (Noise trader risk).** When arbitrageurs have finite horizons, mispricing driven by investor sentiment can persist even without fundamental risk.
- **P2 (Performance-based arbitrage).** When arbitrage capital is delegated and investors withdraw after losses, arbitrageurs are least able to trade when mispricing is greatest.
- **P3 (Where anomalies live).** Mispricing is larger and more persistent in assets that are costly to short, volatile, and lacking close substitutes.
- **P4 (Crowding and liquidation).** When many arbitrageurs hold similar positions, forced selling by some can worsen mispricing for all.

16.7.3 Mechanism

Figure 16.10 simulates a stylised episode. An arbitrageur identifies an 8 percent mispricing between two securities with identical cash flows and takes a leveraged position. Noise trader sentiment widens the gap to almost 30 percent before it eventually closes. A leveraged fund whose investors withdraw after losses is forced to liquidate near the trough and never recovers, while patient capital holding exactly the same position eventually profits.

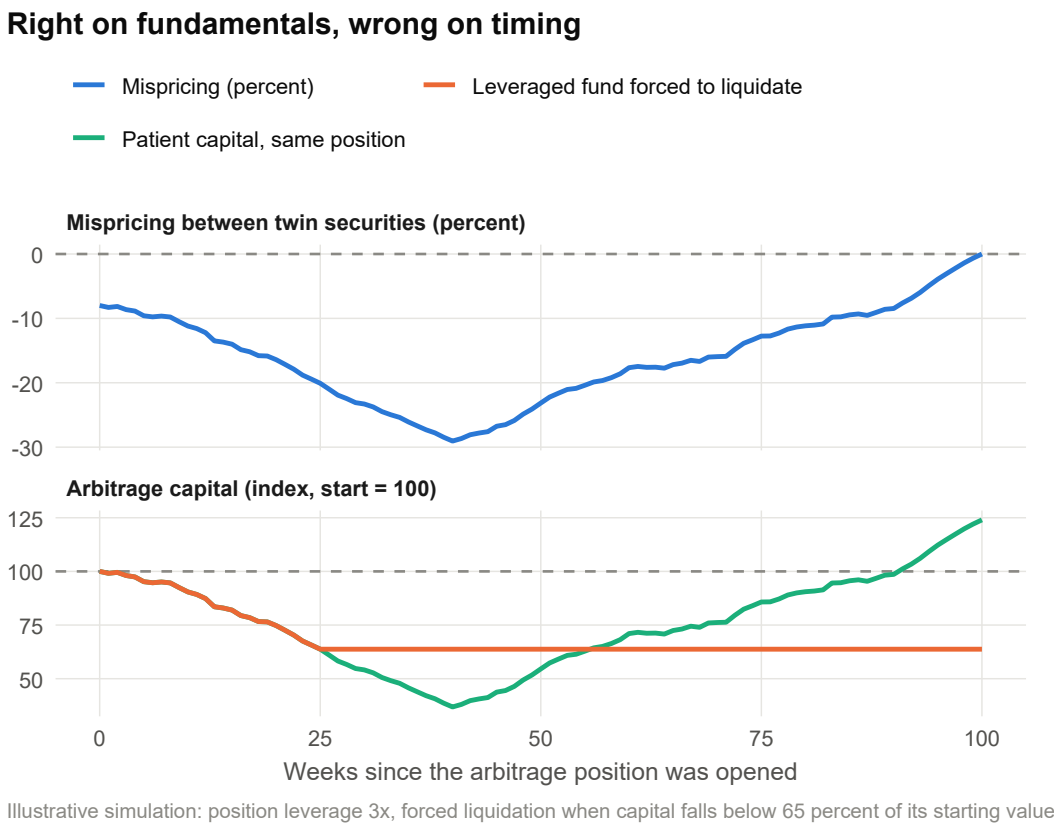


Figure 16.10. Illustrative performance-based arbitrage. Top: the mispricing between two claims on identical cash flows widens before it converges. Bottom: a leveraged arbitrage fund facing investor withdrawals is forced to liquidate near the worst point, while patient capital with the same position ends with a gain.

16.7.4 Evidence and extensions

The cleanest evidence comes from securities whose relative value is fixed by contract. Table 16.7 summarises four well-documented episodes.

Table 16.7. Episodes illustrating the limits to arbitrage.

Episode	What should have held	What happened	Limit to arbitrage
Royal Dutch and Shell (Froot & Dabora 1999)	Twin shares with a fixed split of cash flows should trade at a constant price ratio	The ratio deviated substantially from parity for years	Noise trader risk with no fixed convergence date
Palm and 3Com, 2000 (Lamont & Thaler 2003)	3Com, which still owned most of Palm, should be worth at least its Palm stake	The market implied a negative value for 3Com's other businesses	Short-sale constraints on Palm shares
Long-Term Capital Management, 1998	Convergence trades should profit as spreads narrow	Spreads widened sharply after the Russian default and the fund lost most of its capital	Leverage, crowding, and forced liquidation
GameStop, January 2021 (SEC 2021)	Prices should reflect expected cash flows	The price rose roughly twentyfold in weeks and remained volatile	Short-selling risk, margin calls, and coordinated noise trading

The Royal Dutch and Shell case is particularly instructive. Under their 1907 alliance, cash flows were split 60 to 40, so Royal Dutch should have been worth 1.5 times Shell. Froot and Dabora (1999) showed that the ratio deviated from parity by large amounts for long periods. An arbitrageur betting on convergence faced no fundamental risk, only the risk that the deviation would widen, with no date by which it had to close.

DEBATE AND EVIDENCE

Does limited arbitrage rescue or bury the EMH? Defenders of efficiency note that in each famous case the mispricing was eventually corrected and that it is hard to profit from these episodes after transaction costs. Behavioural economists reply that this is exactly the point: "no free lunch" does not imply "prices are right" (N. Barberis and Thaler 2003). If even riskless relative mispricing can persist, mispricing of whole markets, with no close substitute and no convergence date, can persist far longer. Still, limits to arbitrage explain why mispricing *can* persist, not which direction it takes. That requires a theory of investor beliefs.

16.7.5 Applications in business

BUSINESS CASE

Long-Term Capital Management, 1998. LTCM was founded in 1994 by John Meriwether, and its partners included Myron Scholes and Robert Merton, who shared the 1997 Nobel Memorial Prize for their work on option pricing. The fund specialised in convergence and relative-value trades, such as buying less liquid bonds and selling more liquid but otherwise similar bonds, betting that small price gaps would close. Because each gap was small, the strategy required very high leverage. After Russia defaulted on its domestic debt in August 1998, investors worldwide fled to the safest and most liquid securities. The spreads LTCM had bet would narrow widened instead, and because many other traders held similar positions, attempts to reduce risk pushed prices further against the fund. By September its capital had collapsed, and the Federal Reserve Bank of New York convened a group of major banks and securities firms that recapitalised the fund with about 3.6 billion dollars to avoid a disorderly liquidation (Lowenstein 2000, President's Working Group on Financial Markets 1999). The episode is the textbook illustration of Shleifer and Vishny (1997), published a year earlier: arbitrage can fail when capital is leveraged, crowded, and impatient. Risk managers should stress-test for widening spreads and crowded positions.

16.7.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Anomaly returns are larger among stocks with higher short-selling costs (lower institutional ownership, higher lending fees).
- *H2.* Hedge fund outflows following poor performance predict subsequent widening of mispricing in the stocks those funds hold.

Measures. Mispricing in contractually linked securities, idiosyncratic volatility as arbitrage risk, short interest and loan fees, fund flows, funding liquidity spreads.

Designs. Cross-sectional tests that interact anomaly signals with limits to arbitrage proxies,

natural experiments from short-sale bans or changes in lending supply, case studies of contractually linked securities.

COMMON PITFALL

Using limits to arbitrage as a universal escape hatch. Almost any failure of a mispricing to correct can be attributed, after the fact, to some limit to arbitrage. To make the theory testable, specify *in advance* which limit (short-sale costs, idiosyncratic risk, capital constraints) should matter, measure it, and predict that mispricing will be greater where that limit is tighter.

16.8 Behavioral Finance

THEORY CARD

Core claim. Investor psychology produces systematic errors in beliefs and preferences, and because arbitrage is limited, these errors affect prices, trading, and corporate decisions.
Origins. Finance, economics, and psychology. Early contributions include Shefrin and Statman (1985) and De Long et al. (1990). N. Barberis and Thaler (2003) is the standard survey. Robert Shiller received the Nobel Memorial Prize in 2013 and Richard Thaler in 2017.
Level of analysis. Individual investor, manager, and market.
Key constructs. Overconfidence, disposition effect, extrapolation, limited attention, loss aversion (Section 24.4), mental accounting, investor sentiment, market timing.
Boundary conditions. Biases must be systematic (shared across many investors) to affect prices, and arbitrage must be costly. Effects are strongest for assets that are hard to value and hard to arbitrage.
Related theories. Limits to arbitrage, Prospect theory (Section 24.2), Disposition effect (Section 24.8), Heuristics and biases (Section 23.2), Mental accounting (Section 23.5), Conformity (Section 22.4), Social transmission (Section 26.3).

16.8.1 Origins and intellectual history

Behavioural finance grew from the confluence of two ideas. From psychology came evidence that people deviate systematically from expected utility and Bayesian updating, especially the work of Kahneman and Tversky on judgment under uncertainty and Prospect theory (Section 24.2) (Kahneman and Tversky 1979). From finance came the realisation, formalised by De Long et al. (1990) and Shleifer and Vishny (1997), that rational arbitrage need not eliminate the price effects of irrational traders. N. Barberis and Thaler (2003) describe the field as resting on these two building blocks: limits to arbitrage, which explain why mispricing can survive, and psychology, which explains what form it takes.

16.8.2 Core logic and propositions

- **P1 (Overconfidence and trading).** Investors who overestimate the precision of their information trade too much and earn lower net returns.
- **P2 (Disposition effect).** Investors realise gains more readily than losses, a pattern consistent with prospect theory (Chapter 24) and mental accounting.

- **P3 (Sentiment).** Waves of investor optimism and pessimism move the prices of hard-to-value, hard-to-arbitrage stocks more than others, and high sentiment predicts low subsequent returns for those stocks.
- **P4 (Managerial response).** Managers respond to mispricing by issuing equity when it is overvalued and buying back shares when it is undervalued, and their own biases shape investment.

16.8.3 Mechanism

Figure 16.11 shows how the building blocks combine. Psychological biases generate correlated demand. Limits to arbitrage allow that demand to move prices. Mispricing then feeds back into corporate decisions through equity issuance, repurchases, and investment.

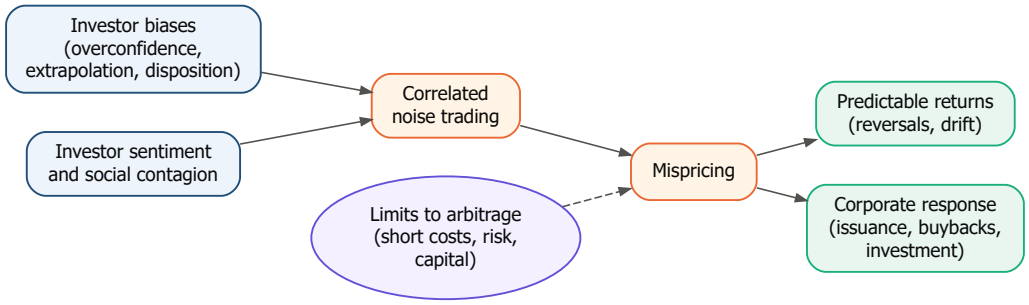


Figure 16.11. Mechanism of behavioural finance. Systematic biases create correlated noise trading. When arbitrage is limited, prices deviate from fundamentals, which affects returns and induces managers to time security issues and repurchases.

16.8.4 Evidence and extensions

Overconfidence. Using records from a large discount brokerage, Barber and Odean (2000) found that the households that traded most earned substantially lower net returns than the market, largely because of trading costs. Barber and Odean (2001) then used gender as a proxy for overconfidence, motivated by evidence that men are more overconfident than women in domains such as finance. Men traded about 45 percent more than women, and trading reduced men's net returns by more than it reduced women's.

Disposition effect. Shefrin and Statman (1985) argued that investors are predisposed to sell winners too early and ride losers too long. Odean (1998) confirmed the pattern in brokerage data: investors realised a higher proportion of their gains than of their losses, and the winners they sold went on to outperform the losers they kept. See the Disposition effect (Section 24.8) for a full treatment.

Investor sentiment. M. Baker and Wurgler (2006) built a composite sentiment index from proxies such as the closed-end fund discount, share turnover, the number and first-day returns of IPOs, the equity share of new issues, and the dividend premium. When sentiment was high, subsequent returns were relatively low for small, young, highly volatile, unprofitable, non-dividend-paying, extreme growth, and distressed stocks, and when sentiment was low the pattern reversed. These are exactly the stocks that are hardest to value and hardest to arbitrage, as P3 predicts. Figure 16.12 sketches the pattern.

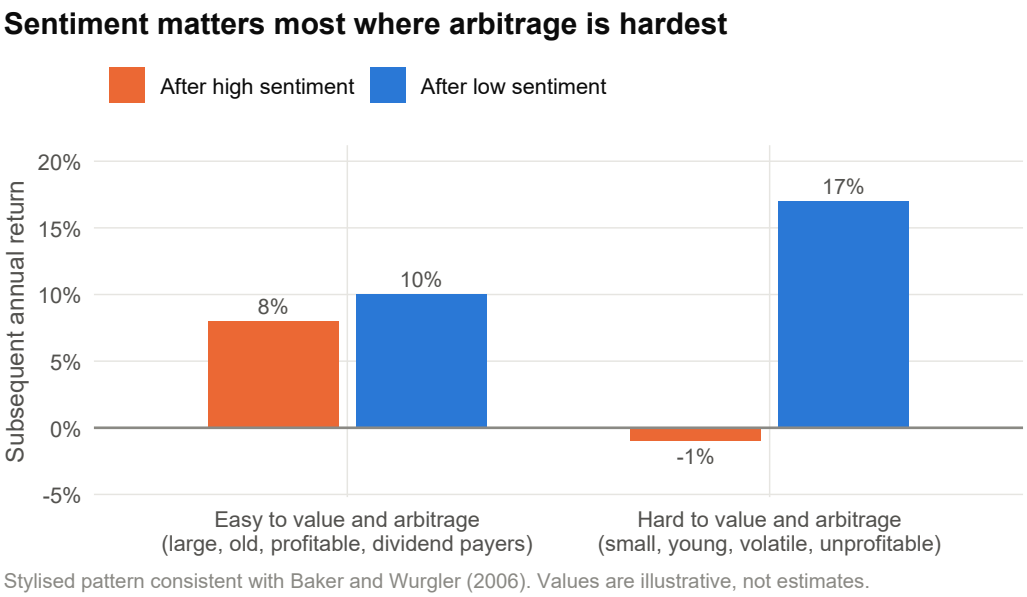


Figure 16.12. Stylised pattern of sentiment and subsequent returns. Hard-to-value stocks earn low returns after periods of high sentiment and high returns after periods of low sentiment, while easy-to-value stocks are little affected. Not empirical data.

Corporate decisions. Rational managers can exploit mispricing by timing equity issues (M. Baker and Wurgler 2002), and managers have biases too: overconfident CEOs overinvest (Malmendier and Tate 2005) and acquirers may overpay because of hubris (Roll 1986). Table 16.8 organises the main findings.

Table 16.8. Behavioural biases in finance: roots, consequences, and evidence.

Bias	Psychological root	Market or firm consequence	Key evidence	Practical implication
Overconfidence	Miscalibrated precision of private information	Excessive trading, lower net returns	Barber & Odean (2000, 2001)	Default to low-turnover strategies, cost disclosure
Disposition effect	Prospect theory, mental accounting, regret	Selling winners and holding losers, drift after news	Shefrin & Statman (1985), Odean (1998)	Rule-based selling, tax-loss harvesting
Extrapolation	Representativeness, recency	Overpricing of glamour stocks, reversals	Barberis & Thaler (2003)	Skepticism toward trend-based forecasts
Limited attention	Finite cognitive resources	Underreaction to news, earnings drift	Bernard & Thomas (1989)	Timing and salience of disclosures
Sentiment	Correlated mood and beliefs	Mispricing of hard-to-value stocks	Baker & Wurgler (2006)	Equity issuance when sentiment is high
Managerial overconfidence	Better-than-average beliefs	Overinvestment and costly acquisitions	Malmendier & Tate (2005), Roll (1986)	Independent boards, investment hurdles

DEBATE AND EVIDENCE

Anomalies or a theory? The strongest criticism of behavioural finance comes from Fama (1998), who argued that the literature had produced a list of anomalies rather than a theory: some studies find underreaction, others overreaction, and many anomalies are sensitive to methodology. Behavioural researchers have responded with models that derive both patterns from a few psychological assumptions, with out-of-sample tests such as those of sentiment indices (M. Baker and Wurgler 2006), and with individual trading data in which biases can be observed directly (Odean 1998, Barber and Odean 2001). Biases observed in trading records have generally proved sturdier than return anomalies, many of which have shrunk since publication (McLean and Pontiff 2016).

16.8.5 Applications in business

BUSINESS CASE

GameStop, January 2021. In late 2020 GameStop, a struggling video-game retailer, was one of the most heavily shorted stocks in the US market, with reported short interest exceeding the number of shares available for trading. Retail investors, many coordinating on the Reddit forum r/wallstreetbets and trading through commission-free apps, began buying the shares and call options. The price rose from under 20 dollars at the start of January 2021 to a closing high above 300 dollars on 27 January. Several hedge funds with short positions suffered large losses, and one, Melvin Capital, received a multibillion-dollar capital infusion from other investment firms. On 28 January several brokers restricted purchases of GameStop and other "meme stocks", citing sharply higher collateral requirements at their clearinghouse. The SEC staff report on the episode found that short covering accounted for only a small part of the buying and that the rise was driven mainly by investors who were not covering short positions (U.S. Securities and Exchange Commission 2021).

The episode combines the chapter's theories. Social contagion and sentiment, see Social transmission (Section 26.3) and Conformity (Section 22.4), produced correlated noise trading. Short sellers faced unlimited potential losses and margin calls, so betting against the rally was extremely risky even for investors convinced that the price exceeded fundamental value, as De Long et al. (1990) and Shleifer and Vishny (1997) predict. GameStop then sold new shares at the elevated price later in 2021, the market timing of M. Baker and Wurgler (2002).

16.8.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Retail investors who receive more push notifications trade more frequently and earn lower risk-adjusted net returns, an effect moderated by financial literacy.
- *H2.* Firms issue more equity and conduct fewer repurchases when sentiment for their industry is high, controlling for investment opportunities.

Measures. Turnover, realised gains and losses (PGR and PLR), sentiment indices, social media attention, short interest, retail order flow, issuance and repurchase activity.

Designs. Account-level brokerage data, field experiments on information presentation, natu-

ral experiments from app feature rollouts, panel regressions of issuance on sentiment, laboratory asset markets.

COMMON PITFALL

Naming a bias instead of testing it. It is easy to label any pattern in returns with a bias after the fact. A credible behavioural finance study derives a prediction from a specific psychological mechanism, shows that the mechanism operates in the population studied, and rules out risk-based and institutional explanations, ideally with a test where the behavioural and rational accounts predict different outcomes.

16.9 Chapter Summary

KEY TAKEAWAYS

- **Portfolio theory**: risk is a property of portfolios. Diversification removes idiosyncratic risk (Markowitz 1952, Tobin 1958).
- The **CAPM** prices only systematic risk through beta (Sharpe 1964, Lintner 1965). Its empirical failures led to multifactor models (Fama and French 1992, 1993).
- The **efficient market hypothesis** holds that prices reflect available information (Fama 1970). Index funds rest on its most robust implication (Sharpe 1991), while excess volatility shows that prices need not equal fundamental value (Shiller 1981).
- The **Modigliani-Miller theorems** make financing irrelevant in perfect markets (Modigliani and Miller 1958), turning every friction, starting with taxes (Modigliani and Miller 1963), into a candidate theory of capital structure.
- The **trade-off** and **pecking order** theories explain leverage through tax shields and distress costs, and through information asymmetry, respectively. Neither fully fits the data (Myers 1984, Shyam-Sunder and Myers 1999, Frank and Goyal 2003).
- **Free cash flow theory** explains why debt, payouts, and takeovers can create value in mature, cash-rich firms (Jensen 1986).
- **Limits to arbitrage** explain why mispricing persists: arbitrage is risky and capital is impatient (De Long et al. 1990, Shleifer and Vishny 1997).
- **Behavioral finance** combines limits to arbitrage with investor psychology to explain trading, anomalies, and corporate financing choices (N. Barberis and Thaler 2003).

Table 16.9 presents the comparison of the chapter's theories.

Table 16.9. Comparison of the chapter's theories.

Theory	Core mechanism	Signature prediction	Main challenge	Typical business use
Portfolio theory	Covariance and diversification	Efficient portfolios dominate concentrated ones	Estimation error in inputs	Asset allocation, retirement plans
CAPM	Only systematic risk is priced	Expected returns linear in beta	Flat SML, size, value, momentum	Cost of capital, performance evaluation
Efficient market hypothesis	Competition among informed traders	No predictable abnormal returns	Anomalies, excess volatility	Index funds, event studies

Table 16.9. Comparison of the chapter's theories. *(continued)*

Theory	Core mechanism	Signature prediction	Main challenge	Typical business use
Modigliani-Miller	Arbitrage and homemade leverage	Financing irrelevant in perfect markets	Assumptions rarely hold	Benchmark for financing decisions
Trade-off and pecking order	Taxes versus distress, information asymmetry	Target leverage or financing hierarchy	Persistence, market timing	Leverage and issuance policy
Free cash flow theory	Agency costs of surplus cash	Debt and payouts discipline managers	Precautionary cash holdings	Payout policy, buyouts
Limits to arbitrage	Risky, capital-constrained arbitrage	Mispricing persists where arbitrage is hard	Ex post flexibility	Risk management, hedge fund design
Behavioral finance	Investor biases plus limits to arbitrage	Sentiment and bias-driven mispricing	Many anomalies, few unified models	Product design, market timing

16.10 Discussion Questions

1. Many founders and executives hold most of their wealth in a single firm. Using agency theory and portfolio theory, explain when such concentration is efficient for the firm even if it is costly for the individual.
2. The CAPM fails many empirical tests yet remains the most common method for estimating the cost of equity. Is this a failure of practice or a defensible choice?
3. Design an event study of a sustainability announcement. Which form of market efficiency must hold for your estimate to be meaningful?
4. A CFO argues that because debt is cheaper than equity, more leverage lowers the cost of capital. Evaluate the argument with the Modigliani-Miller propositions and the trade-off theory.
5. Large technology firms hold very large cash balances. Develop competing hypotheses from free cash flow and pecking order theory, and describe data that could distinguish them.
6. Using the GameStop episode, explain why short sellers who believed the stock was overvalued might rationally have closed their positions. What does this imply for price efficiency during episodes of extreme sentiment?

16.11 Further Reading

- Markowitz (1952): the short paper that founded modern portfolio theory and remains a model of clear argument.
- Fama and French (2004): an accessible review of the CAPM's logic, its empirical record, and why its failures matter for practice.
- Fama (1970): the classic synthesis of market efficiency, with the three forms and the joint hypothesis problem.
- Myers (2001): a lucid survey of capital structure theories from Modigliani-Miller through the trade-off and pecking order debates.
- Jensen (1986): a compact statement of free cash flow theory that shaped two decades of research on payouts and takeovers.

- Shleifer and Vishny (1997): the paper that explains why arbitrage fails when it is most needed, vindicated by the LTCM crisis a year later.
- N. Barberis and Thaler (2003): the standard survey of behavioural finance, organised around limits to arbitrage and investor psychology.
- French (2008): a careful accounting of what investors spend trying to beat the market, essential background for the index fund debate.

Chapter 17

Economics of Markets and Institutions

In January 1914 the Ford Motor Company announced that it would pay many of its production workers five dollars a day, roughly double the prevailing wage for comparable work. A hardware store that raises the price of snow shovels the morning after a blizzard is condemned by its customers, even though the price rise is exactly what an introductory textbook would prescribe. Apple lets consumers download most apps without charge but takes a commission from developers, and payment-card networks reward cardholders while charging merchants. Governments now sell radio spectrum by auction and create markets in the right to emit carbon dioxide. None of these practices fits the frictionless model of a competitive market, in which prices clear markets, every good is privately owned, information is free, and people care only about their own consumption.

This chapter assembles the theories that economists developed when they relaxed those assumptions one at a time: costly bargaining over externalities, shared goods, incomplete contracts, costly information, unobservable effort, interdependent demand, private valuations, weak rules, and other-regarding preferences. Together they form the analytical core of modern industrial organization, labour economics, law and economics, and market design, and they supply much of the vocabulary that strategy and marketing scholars use to study platforms, pricing, and governance (Table 17.1).

Table 17.1. Chapter at a glance: theories of markets and institutions.

Theory	Core question	Level of analysis	Key source
Coase theorem	When do private bargains solve externalities, and what happens when bargaining is costly?	Dyad / market / policy	Coase (1960)
Collective action	Why do groups fail to provide goods their members want?	Group / society	Olson (1965)
Property rights theory	Who should own assets when contracts are incomplete?	Dyad / firm	Grossman & Hart (1986)
Search theory	Why do prices differ for identical goods, and why do jobs and workers fail to meet?	Individual / market	Stigler (1961)
Efficiency wages	Why do firms pay more than the market-clearing wage?	Firm / labour market	Akerlof (1982), Shapiro & Stiglitz (1984)
Network effects	Why do some markets tip to a single standard?	Market / industry	Katz & Shapiro (1985)
Two-sided markets	How should a platform price to two interdependent user groups?	Platform / market	Rochet & Tirole (2003)

Table 17.1. Chapter at a glance: theories of markets and institutions. *(continued)*

Theory	Core question	Level of analysis	Key source
Auction theory	How do the rules of a sale shape bidding, revenue, and efficiency?	Market / mechanism	Vickrey (1961), Milgrom & Weber (1982)
New institutional economics	How do the rules of the game shape economic performance?	Institution / society	North (1990), Williamson (2000)
Social preferences	When do people sacrifice money for fairness, and how does that constrain markets?	Individual / dyad	Fehr & Schmidt (1999)

17.1 Coase Theorem

THEORY CARD

Core claim. If property rights are well defined and transaction costs are zero, the parties affected by an externality will bargain to the efficient outcome whatever the initial assignment of rights. Because transaction costs are rarely zero, the assignment of rights and the design of institutions do matter, and the real task is to compare the costs of alternative arrangements.
Origins. Law and economics. Coase (1960), written in reaction to the welfare economics of Pigou (1920). Coase received the 1991 Nobel Memorial Prize in Economic Sciences, in part for this paper and for Coase (1937).
Level of analysis. Dyad (polluter and pollutee), market, and public policy.
Key constructs. Externality, entitlement (property right), transaction costs, reciprocal nature of harm, Pigouvian tax, tradable permit.
Boundary conditions. The invariance result requires few parties, clearly defined and enforceable rights, symmetric information, and negligible wealth and endowment effects.
Related theories. Transaction cost economics (Section 6.3), Collective action, Property rights theory, Tragedy of the commons (Section 30.11), Endowment effect (Section 23.4).

17.1.1 Origins and intellectual history

Pigou (1920) observed that the private net product of an activity can diverge from its social net product: a railway's sparks burn neighbouring crops without the railway paying for the damage. The Pigouvian remedy is a tax equal to the marginal damage or a liability rule that makes the injurer pay.

Coase (1960) challenged this view with a deceptively simple point: harm is **reciprocal**. If a rancher's cattle trample a farmer's crops, preventing the damage imposes a cost on the rancher. The question is not how to restrain the rancher but whether the rancher should be allowed to harm the farmer or the farmer to harm the rancher. Through examples drawn from English nuisance cases, Coase showed that if the parties can bargain without cost, they reach the arrangement that maximizes joint value whoever is initially liable. The legal rule determines who pays whom, not what is produced. A year earlier he had proposed selling radio frequencies to the highest bidder (Coase 1959), an idea that became United States policy in 1993 (see Auction theory).

17.1.2 Core logic and propositions

Let a firm's private benefit from an activity of level q be $B(q)$, and let the activity impose damage $D(q)$ on a neighbour, with B concave and D convex. The efficient level q^* maximizes joint surplus and satisfies

$$B'(q^*) \;=\; D'(q^*).$$

The Pigouvian solution imposes a tax $t = D'(q^*)$ per unit. The Coasean solution needs no tax. If the firm has the right to operate, the neighbour pays it to cut back. If the neighbour has the right to be free of harm, the firm buys permission. Both bargains end at q^*, and only the distribution of surplus differs. The theory's propositions are:

- **P1 (Invariance).** With zero transaction costs and well-defined rights, the efficient allocation is reached regardless of the initial assignment of rights.
- **P2 (Transaction costs).** As the costs of negotiating and enforcing agreements rise, the initial assignment increasingly determines the final allocation.
- **P3 (Comparative institutions).** The right response to an externality depends on the comparative costs of bargaining, integration, regulation, and taxation.
- **P4 (Markets in rights).** Where rights can be defined, monitored, and traded cheaply, a market in rights meets a given target at the lowest total cost.

Table 17.2 compares the main instruments that follow from Pigou and Coase.

Table 17.2. Policy instruments for externalities.

Instrument	What the policymaker fixes	Information required	Main strength	Main weakness
Pigouvian tax	Price of the externality	Marginal damage at the optimum	Equalizes marginal abatement cost across emitters	Quantity of emissions is uncertain
Command-and-control standard	Technology or emission rate per source	Abatement options at each source	Simple to monitor and enforce	Ignores cost differences across sources
Cap-and-trade (tradable permits)	Total quantity of emissions	Aggregate target and monitoring	Least-cost abatement with a known cap	Price volatility, allocation politics
Liability rule or private bargaining	Who holds the entitlement	Courts must assess damages	No regulator needed when parties are few	Fails with many parties and high transaction costs

17.1.3 Mechanism

Cap-and-trade is the clearest application of Coasean reasoning. A regulator creates a fixed quantity of emission rights. Firms that can abate cheaply abate more and sell permits to firms whose abatement is expensive, until marginal abatement costs are equal across firms, the condition for meeting the cap at minimum total cost (Montgomery 1972). Figure 17.1 illustrates the logic for two firms that must jointly abate 100 tonnes.

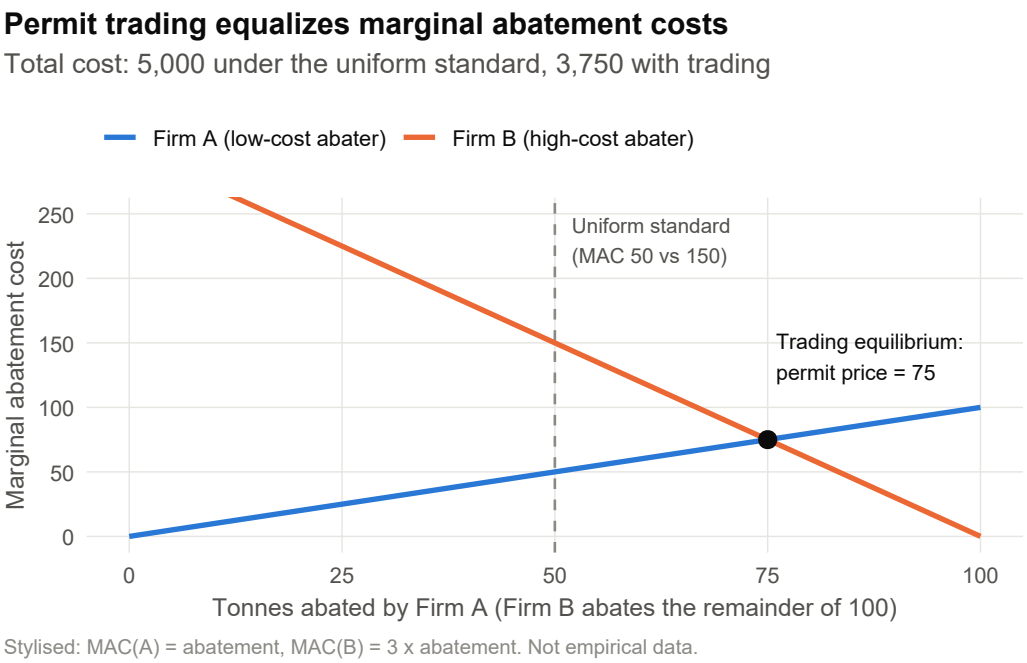


Figure 17.1. Stylised two-firm emissions market. A uniform standard (each firm abates 50 tonnes) leaves marginal abatement costs unequal. Trading moves abatement to the low-cost firm until marginal costs are equal at the permit price, cutting total abatement cost by a quarter in this example.

17.1.4 Evidence and extensions

In laboratory bargaining with full information, Hoffman and Spitzer (1982) found that pairs and small groups usually reached the joint-payoff-maximizing agreement whichever party held the legal right. In the field, Cheung (1973) documented that beekeepers and orchard owners in Washington State routinely contract for pollination, a textbook example of an uncompensated externality that turned out to be priced.

Two literatures qualify the theorem. Kahneman, Knetsch, and Thaler (1990) showed that people demand substantially more to give up a good than they would pay to acquire it, so the initial allocation of an entitlement affects its final allocation even when trading is costless (see Endowment effect (Section 23.4)). Weitzman (1974) showed that when abatement costs are uncertain, a tax and a cap are no longer equivalent: which is better depends on the relative slopes of marginal cost and marginal damage.

The largest natural experiment is the United States sulphur dioxide allowance market created by the 1990 Clean Air Act Amendments, which met its target at substantially lower cost than a comparable command-and-control approach (Stavins 1998). Later regulatory and judicial decisions layered source-specific constraints on top of the cap and drained the market of importance, a cautionary tale about the political durability of market-based instruments (Schmalensee and Stavins 2013).

DEBATE AND EVIDENCE

What does the Coase theorem actually say? Coase himself objected that economists had read the zero-transaction-cost result as the main message. On his reading, because transac-

tion costs are positive, rights assignments and institutions matter, and policy should compare real alternatives rather than idealized ones. A second debate concerns **strategic behaviour**: with private information about valuations, bargaining can break down even between two parties. A third concerns **distribution**: even when invariance holds, the assignment of rights determines who is richer, which is why permit allocation is the most contested part of every cap-and-trade programme.

17.1.5 Applications in business

BUSINESS CASE

Carbon markets. The European Union Emissions Trading System, launched in 2005, caps emissions from power generation and heavy industry and allows covered installations to trade allowances. Its first phase illustrated the importance of accurate rights definition: allowances were allocated largely free on the basis of estimated needs, and when verified emissions data showed that the market was over-allocated, prices collapsed to near zero by 2007 because first-phase allowances could not be carried into the next phase. Later reforms moved toward auctioning and introduced a Market Stability Reserve that withdraws surplus allowances. For firms, the managerial implications follow directly from Figure 17.1: treat carbon as a priced input, abate wherever internal abatement is cheaper than the allowance price, and buy allowances where it is not. Many firms also apply an internal carbon price to investment appraisal.

17.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Bilateral disputes over externalities are more likely to be resolved by private agreement when the number of affected parties is small and property rights are clearly documented.
- *H2.* Under a cap-and-trade programme, dispersion of marginal abatement cost across covered firms declines as market liquidity rises.

Designs. Laboratory bargaining experiments varying rights assignment and information. Difference-in-differences around the introduction or reform of permit markets. Event studies of allowance price shocks on covered firms' stock returns.

COMMON PITFALL

Reading Coase as an argument against all regulation. The theorem does not say that externalities take care of themselves. It says they would if transaction costs were zero, which they rarely are. When the affected parties number in the millions (as with air pollution or carbon dioxide), private bargaining is impossible, and the relevant comparison is among imperfect institutions: taxes, caps, standards, and liability. Citing Coase to oppose a carbon market inverts the theorem: tradable permits are themselves a Coasean creation of property rights.

17.2 Collective Action

THEORY CARD

Core claim. Rational, self-interested members of a large group will not contribute voluntarily to a good that benefits the whole group, because each receives the benefit whether or not he or she contributes. Large groups therefore organize only when they can offer **selective incentives** or impose coercion.
Origins. Public economics and political economy. P. A. Samuelson (1954) formalized public goods. Olson (1965) extended the logic to groups, unions, cartels, and lobbies.
Level of analysis. Group, organization, industry, and society.
Key constructs. Public good (non-rival and non-excludable), free riding, group size, selective incentives, privileged and latent groups, conditional cooperation, peer punishment.
Boundary conditions. Benefits cannot be withheld from non-contributors, contributions are costly, and members interact anonymously. Repeated interaction and sanctioning weaken the prediction.
Related theories. Tragedy of the commons (Section 30.11), Social loafing (Section 30.9), Social preferences, Social norms (Section 22.5), New institutional economics.

17.2.1 Origins and intellectual history

P. A. Samuelson (1954) defined a collective consumption good as one whose consumption by one person does not reduce the amount available to anyone else. He showed that efficient provision requires the *sum* of individuals' marginal rates of substitution between the public good G and a private good to equal the marginal rate of transformation:

$$\sum_{i=1}^{n} MRS_i^{G,x} \;=\; MRT^{G,x}.$$

Because people can free ride, decentralized markets underprovide public goods.

Olson (1965) showed that shared interests do not imply collective action. Every member of a trade association benefits from favourable legislation, but each firm's contribution barely affects the outcome and the firm benefits whether or not it pays. The argument explains why small, concentrated interests often defeat large, diffuse ones, and why unions and professional associations bundle political activity with private benefits for members. The closely related Tragedy of the commons (Section 30.11) concerns resources that are rival in consumption, and Ostrom (1990)'s work on self-governing communities, discussed there, is the most influential response to both problems.

17.2.2 Core logic and propositions

The **linear public goods game** captures the structure. Each of n players receives an endowment e and chooses a contribution c_i to a group project. Contributions are multiplied and shared equally:

$$\pi_i \;=\; e - c_i + m\sum_{j=1}^{n} c_j, \qquad \frac{1}{n} < m < 1.$$

Each unit contributed returns only $m < 1$ to the contributor, so contributing nothing is a dominant strategy. Each unit contributed returns $nm > 1$ to the group, so full contribution is socially efficient. The gap between the two is the collective action problem. Olson's propositions follow:

- **P1 (Group size).** Larger groups are less likely to provide collective goods, because each member's share of the benefit is smaller, each contribution is less noticeable, and organizing costs are higher.
- **P2 (Selective incentives).** Large ("latent") groups mobilize when participation is tied to private benefits available only to contributors, such as insurance, journals, training, or certification.
- **P3 (Privileged groups).** If one member values the good more than its full cost, the good is provided, and the large member bears a disproportionate share (Olson's "exploitation of the great by the small").
- **P4 (Sanctions and reciprocity).** Contributions are sustained when members can observe one another and punish free riders, an extension developed in experimental work (Fehr and Gächter 2000a).

17.2.3 Mechanism

In repeated public goods experiments, contributions typically start well above zero and decline as cooperative participants discover that others free ride. Many participants are **conditional cooperators** who contribute more when they believe others are contributing (Fischbacher, Gächter, and Fehr 2001). Fehr and Gächter (2000a) showed that allowing participants to pay to punish low contributors reverses the decline. Figure 17.2 sketches the qualitative pattern.

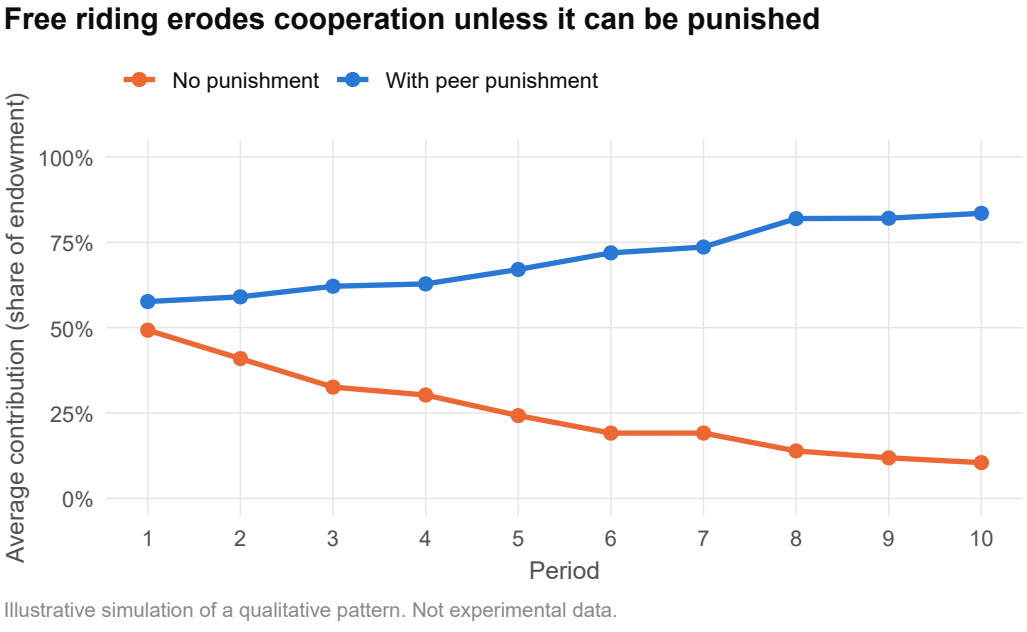


Figure 17.2. Stylised pattern of contributions in a repeated public goods game with and without peer punishment. Without punishment, contributions decay toward free riding. With costly punishment, contributions are sustained. Qualitative pattern consistent with Fehr and Gachter (2000a). Illustrative, not their data.

17.2.4 Evidence and extensions

The free-rider prediction is confirmed in its direction but not in its magnitude: people contribute more than pure self-interest predicts, especially early, and their behaviour depends on what they expect others to do. Cultural context matters. Herrmann, Thöni, and Gächter (2008) ran the same public goods game with punishment in 16 subject pools around the world and found that in some societies punished participants frequently retaliated against high contributors (**antisocial punishment**), which undermined cooperation. Ostrom (2000) reviews field and laboratory evidence that communities develop norms and graduated sanctions that sustain collective action well beyond what Olson's model implies.

Table 17.3 summarizes the main solutions identified across this literature.

Table 17.3. Solutions to collective action problems.

Solution	Logic	Business or policy example
Selective incentives	Bundle the public good with private benefits for contributors	Trade association membership that includes data, events, and certification
Coercion or mandate	Make contribution compulsory	Taxes, compulsory levies for generic product advertising
Privileged sponsor	A member large enough to gain from providing the good alone	A dominant firm funding an open industry standard
Federated structure	Organize small groups in which contributions are noticeable	Local chapters of national associations
Monitoring and sanctions	Make free riding visible and costly	Peer review of contributions in research consortia
Norms and reputation	Social approval for contributors, disapproval for free riders	Public recognition of open-source maintainers

DEBATE AND EVIDENCE

Is group size really the enemy of cooperation? If the good is purely non-rival, a larger group may include more members with high valuations, so the probability that someone provides it can rise with size, and field research on commons governance finds no simple monotonic effect (Ostrom 2000). A second critique targets the behavioural assumption: conditional cooperators and willing punishers mean that institutions which make contributions visible can unlock cooperation that Olson's model rules out (Fehr and Gächter 2000a, Fischbacher, Gächter, and Fehr 2001). A third notes that large-scale successes such as open-source software rely on selective incentives Olson anticipated: reputation, learning, and career signalling.

17.2.5 Applications in business

BUSINESS CASE

Underfunded digital infrastructure. Open-source software is a public good for the firms that build on it. In April 2014 the Heartbleed vulnerability in OpenSSL, a widely used cryptographic library, revealed that critical code relied on a very small group of mostly volunteer maintainers, although many large firms depended on it. The response followed Olson's playbook: the Linux Foundation organized the Core Infrastructure Initiative, through which large technology firms (a privileged group) jointly funded critical projects, with public recognition

of sponsors as a selective incentive.

17.2.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Contributions to an inter-firm consortium will be higher when contributions are publicly disclosed to other members.
- *H2.* The positive effect of membership benefits on association membership will be stronger in industries with many small firms than in concentrated industries.

Designs. Laboratory and online public goods games. Field experiments that vary the visibility of contributions (for example, charitable giving). Panel data on association membership and lobbying expenditure.

COMMON PITFALL

Calling every shared good a public good. Many "collective" goods are excludable: industry data, certification, and standards can be restricted to members. Such club goods face weaker free-rider problems, and the right remedy is pricing access, not subsidy or mandate.

17.3 Property Rights Theory

THEORY CARD

Core claim. When contracts are incomplete, ownership of physical assets confers **residual control rights**, the right to decide how assets are used in circumstances the contract does not cover. Ownership therefore shapes each party's bargaining position after investments are sunk, and hence each party's incentive to invest beforehand. Assets should be owned by the party whose investment matters most.

Origins. Economics of the firm. Grossman and Hart (1986), Hart and Moore (1990), synthesized in Hart (1995). Hart shared the 2016 Nobel Memorial Prize for contract theory.

Level of analysis. Dyad (buyer and supplier) and the boundary of the firm.

Key constructs. Incomplete contracts, residual control rights, relationship-specific investment, hold-up, outside option, asset complementarity, indispensability.

Boundary conditions. Some investments cannot be specified in an enforceable contract, parties can renegotiate after uncertainty resolves, and ownership of physical (or legally protectable) assets affects outside options.

Related theories. Transaction cost economics (Section 6.3), Agency theory (Section 6.1), Coase theorem, Resource-based theory (Section 4.1).

17.3.1 Origins and intellectual history

Coase (1937) asked why some transactions are organized inside firms rather than across markets. Williamson's Transaction cost economics (Section 6.3) and B. Klein, Crawford, and Alchian (1978) answered that relationship-specific investments create appropriable quasi-rents that a trad-

ing partner can try to capture by threatening to walk away, the **hold-up** problem, and that vertical integration suppresses this opportunism. Grossman and Hart (1986) accepted the hold-up premise but asked why, if integration cures opportunism, firms do not integrate everything. Their answer was that integration does not remove incentive problems. It moves them. When firm 1 buys firm 2, the former owner-manager of firm 2 becomes an employee with weaker incentives, because the assets she needs to capture the returns on her effort are now controlled by someone else. The crucial step was to define **ownership as the residual right** to decide everything the contract leaves unspecified. With complete contracts, ownership would not matter. With incomplete contracts, it determines who has the upper hand when the unexpected happens.

17.3.2 Core logic and propositions

Consider two parties who each make a non-contractible investment i_k before they trade. Trade generates joint value $v(i_1, i_2)$. If bargaining breaks down, each party receives an outside option $o_k^g(i_k)$ that depends on which assets it controls under ownership structure g (non-integration, party 1 owns, party 2 owns). With symmetric Nash bargaining, party 1's payoff is

$$u_1 \;=\; o_1^g + \tfrac{1}{2}\left[v(i_1, i_2) - o_1^g - o_2^g\right] - i_1,$$

so party 1's marginal return on investment is $\tfrac{1}{2}\,\partial v/\partial i_1 + \tfrac{1}{2}\,\partial o_1^g/\partial i_1$. Because the party captures only part of the value its investment creates, investment falls short of the first best. Owning more assets increases $\partial o_1^g/\partial i_1$ and therefore increases party 1's investment, but at the expense of party 2's. Hart and Moore (1990) and Hart (1995) derive the following propositions:

- **P1 (Importance of investment).** Ownership should go to the party whose non-contractible investment contributes more to joint value.
- **P2 (Independent assets).** Assets that are economically independent should be owned separately.
- **P3 (Complementary assets).** Assets that are strictly complementary should be owned together.
- **P4 (Indispensability).** A party whose human capital is indispensable to an asset should own that asset.
- **P5 (Costs of integration).** Integration raises the investment incentives of the acquirer and lowers those of the acquired party. It is efficient only when the first effect dominates.

Figure 17.3 summarizes the mechanism.

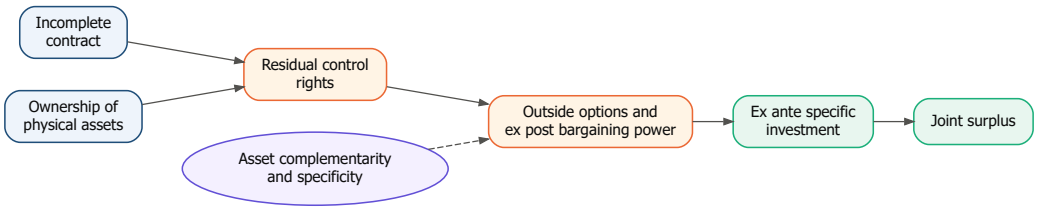


Figure 17.3. Mechanism of property rights theory: with incomplete contracts, ownership allocates residual control rights, which shape outside options and bargaining power ex post, and therefore investment incentives ex ante.

17.3.3 Evidence and extensions

Clean tests are difficult because investments are, by assumption, hard to observe. G. P. Baker and Hubbard (2003) provide one of the best: the diffusion of on-board computers in United States trucking changed what could be monitored and contracted, which in turn changed who owned the trucks. The most famous historical case is contested. B. Klein, Crawford, and Alchian (1978) described General Motors' 1926 acquisition of Fisher Body as a response to hold-up by a supplier that had refused to locate plants near GM's, but later archival work disputed that account (Coase 2000, Freeland 2000, Casadesus-Masanell and Spulber 2000), a reminder that canonical examples deserve primary-source scrutiny.

Table 17.4 contrasts property rights theory with Transaction cost economics (Section 6.3), its closest relative.

Table 17.4. Transaction cost economics and property rights theory compared.

Dimension	Transaction cost economics	Property rights theory
Key source	Williamson (1975, 1985)	Grossman & Hart (1986), Hart & Moore (1990)
Unit of analysis	The transaction	Ownership of assets
Main inefficiency	Ex post haggling and maladaptation	Ex ante underinvestment
Role of integration	Replaces haggling with authority (fiat)	Reallocates residual control rights
Cost of integration	Bureaucratic costs, weaker incentives	Lower investment incentives for the acquired party
Prediction	Integrate when asset specificity and uncertainty are high	Give ownership to the party whose investment matters most
Typical evidence	Make-or-buy regressions on asset specificity	Ownership changes after shifts in contractibility

DEBATE AND EVIDENCE

How firm are the foundations? Maskin and Tirole (1999) argued that if parties can foresee the payoffs of future contingencies, even without describing them, clever mechanisms can in principle restore efficient investment, undermining the incomplete-contracts premise. Defenders reply that such mechanisms are fragile and rarely observed. A second critique is empirical: because the two theories in Table 17.4 often predict the same make-or-buy choice, findings that asset specificity predicts integration cannot distinguish them. Discriminating tests need settings where the theories diverge.

17.3.4 Applications in business

BUSINESS CASE

Illustrative scenario: who should own the tooling? An automotive manufacturer asks a supplier to build specialized moulds for a new component. If the supplier owns the moulds, it has a strong incentive to maintain and improve them, but the manufacturer fears hold-up once switching is costly. If the manufacturer owns the moulds, it can move them if the relationship sours, which disciplines the supplier but weakens its incentive to improve the process. Property rights theory predicts manufacturer ownership when the manufacturer's investments matter more than the supplier's improvement effort.

17.3.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Following a technology that makes a supplier's effort contractible, the probability that the buyer owns the relevant asset will rise.
- *H2.* In R&D alliances, the partner whose non-contractible know-how is more important to the project is more likely to hold the resulting intellectual property.

Designs. Natural experiments in contractibility (new monitoring technologies, legal changes). Contract-level data on ownership and control clauses. Panel studies of inventors before and after acquisitions.

COMMON PITFALL

Treating integration as free. A common misuse is to cite hold-up as a reason to integrate without asking whose incentives integration destroys. Ownership is a zero-sum allocation of control: strengthening one party's investment incentives weakens another's. Ignoring this trade-off leads to systematic over-integration.

17.4 Search Theory

THEORY CARD

Core claim. When information about prices, products, or trading partners is costly to acquire, buyers and job seekers search only until the expected gain from one more search falls to its cost. Price dispersion for identical goods and the coexistence of unemployment and vacancies are therefore equilibrium outcomes, not anomalies.

Origins. Economics of information. Stigler (1961), J. J. McCall (1970), Diamond (1971), and for labour markets Mortensen and Pissarides (1994). Diamond, Mortensen, and Pissarides shared the 2010 Nobel Memorial Prize.

Level of analysis. Individual searcher, firm, and market.

Key constructs. Search cost, price dispersion, reservation price and reservation wage, sequential versus fixed-sample search, matching function, market tightness.

Boundary conditions. Offers vary across sellers or employers, finding them is costly. Free, universally used comparison tools shrink its predictions.

Related theories. Information asymmetry (Section 8.1), Bounded rationality (Section 23.1), Two-sided markets, Reference price (Section 27.11).

17.4.1 Origins and intellectual history

Stigler (1961, 213) observed that information occupies "a slum dwelling in the town of economics": price theory had assumed away the problem of finding out what prices are. Because ignorance is costly to remove, rational buyers stop searching before they find the lowest price, and sellers can sustain different prices for identical goods. The more a buyer spends, the more search pays.

J. J. McCall (1970) modelled **sequential** search for jobs, in which a worker samples one offer at a time and stops at the first offer above a **reservation wage**. Diamond (1971) showed a startling

equilibrium implication: if all consumers face even a tiny positive search cost and firms set prices, the unique equilibrium is for every firm to charge the monopoly price, because no consumer will search to find a slightly lower one (the **Diamond paradox**). Later work restored price dispersion as an equilibrium by allowing consumers to differ in search costs or in how many quotes they obtain (Varian 1980, Burdett and Judd 1983). In labour economics, Mortensen and Pissarides (1994) built a model of job creation and destruction in which a **matching function** links unemployed workers and vacancies, explaining why both exist at once and how they move over the business cycle.

17.4.2 Core logic and propositions

In sequential search with cost c per offer and wage offers drawn from distribution F, the optimal policy is to accept the first offer at or above a reservation wage w^R that equates the cost of one more draw with its expected gain:

$$c = \int_{w^R}^{\infty} \left(w - w^R\right)\, dF(w).$$

Higher search costs lower the reservation wage, and a more dispersed offer distribution raises it. In matching models, hires $m(u, v)$ increase in unemployment u and vacancies v, and **market tightness** $\theta = v/u$ governs how fast workers find jobs and firms fill vacancies. The propositions are:

- **P1 (Optimal stopping).** Searchers stop when the marginal gain of search equals its cost, so they rationally accept prices above the minimum.
- **P2 (Stakes).** Search effort and price convergence increase with the size of the expenditure (Stigler 1961).
- **P3 (Search costs and market power).** Higher search costs raise equilibrium prices. With homogeneous positive search costs, prices rise to the monopoly level (Diamond 1971).
- **P4 (Frictional unemployment).** Unemployment and vacancies coexist, and lower search frictions raise matching efficiency (Mortensen and Pissarides 1994).

17.4.3 Mechanism

Figure 17.4 illustrates Stigler's fixed-sample logic: each additional quote lowers the expected minimum price by less and less, so a business buying ten units rationally collects many more quotes than a household buying one.

17.4.4 Evidence and extensions

The internet was expected to collapse price dispersion. Brynjolfsson and Smith (2000) found that online prices for books and CDs were lower than offline prices but that online dispersion remained substantial. J. R. Brown and Goolsbee (2002) found that the growth of comparison sites for term life insurance was associated with meaningful price reductions. Ellison and Ellison (2009) showed that retailers on a price comparison site responded to highly price-elastic search by **obfuscation**: advertising a low-quality product at a low price and steering buyers to higher-margin upgrades. Search costs are partly chosen by sellers. In labour economics, Shimer (2005) showed that the standard matching model with plausible parameters generates far smaller fluctuations in unemployment and vacancies than observed, prompting a large literature on wage rigidity.

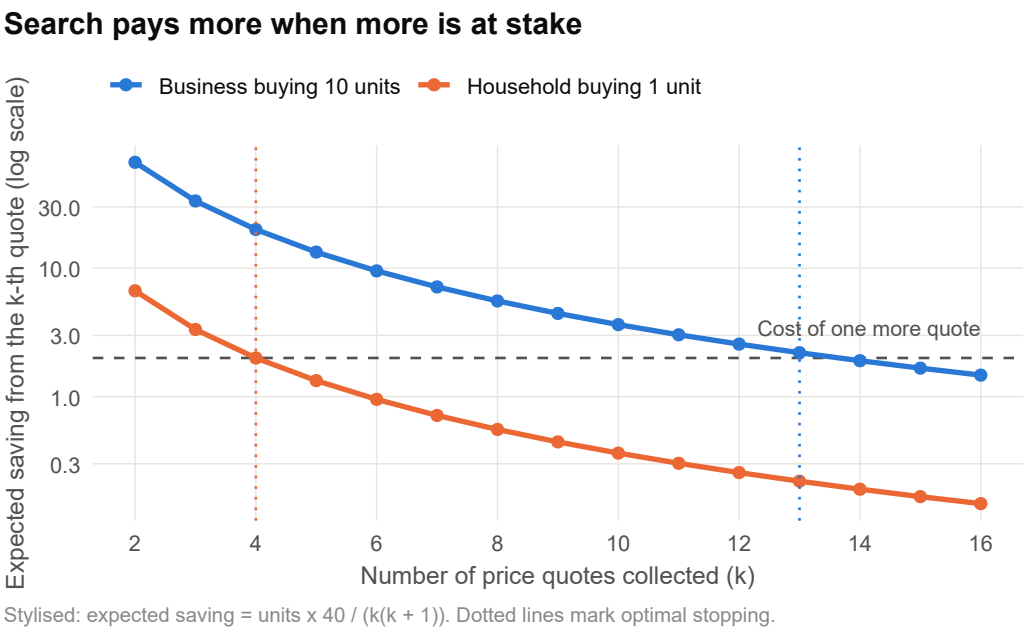


Figure 17.4. Stylised Stigler search model. Prices are uniform between 80 and 120 and each quote costs 2. The expected saving from the k-th quote falls quickly, so a buyer of one unit stops at about four quotes while a buyer of ten units collects about thirteen.

Table 17.5. Search frictions across markets.

Market	Source of search cost	Observable consequence	Managerial lever
Consumer retail	Time to visit stores or compare sites	Persistent price dispersion	Price-match guarantees, comparison listings
Insurance and finance	Complex products, hard to compare	Dispersion in premiums and fees	Standardized disclosures, aggregators
Labour	Finding and screening candidates	Unemployment and vacancies coexist	Referral programmes, job platforms
Housing	Viewing, negotiating, coordinating	Long time-to-sale, agent intermediation	Listing platforms, virtual tours
B2B procurement	Qualifying suppliers, requesting quotes	Sticky supplier relationships	Reverse auctions, supplier portals

DEBATE AND EVIDENCE

Do lower search costs always help consumers? Comparison tools should intensify competition, but firms respond by obfuscating, adding fees revealed late in the purchase, or differentiating products so that they are harder to compare (Ellison and Ellison 2009). Lower search costs can also benefit sellers by improving matches. The welfare effect of a search technology therefore depends on how sellers adapt.

17.4.5 Applications in business

BUSINESS CASE

Comparison sites and obfuscation. Price comparison platforms for insurance, flights, and consumer electronics lower search costs for consumers and push sellers toward the lowest headline price. Sellers have responded with partitioned prices, add-on fees, and stripped-down base products, a pattern consistent with Ellison and Ellison (2009). For managers, the headline price is only one dimension of search: firms can compete by being easy to compare or by making comparison harder (Table 17.5 lists levers in other markets). For regulators, disclosure rules on total price can be more effective than measures that only make search cheaper.

17.4.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Price dispersion for an identical product will be lower in categories with higher average expenditure per purchase.
- *H2.* Entry of a comparison platform into a market will reduce average prices and increase the use of add-on fees.

Designs. Web-scraped price panels, clickstream data on consumer search, field experiments that vary search costs, structural estimation of search models.

COMMON PITFALL

Calling price dispersion irrational. Paying more than the lowest available price is not evidence that consumers are irrational. It is exactly what optimal search predicts when search is costly. Before attributing dispersion to bias, measure search costs and the stakes of the purchase.

17.5 Efficiency Wages

THEORY CARD

Core claim. Firms may rationally pay more than the market-clearing wage because higher pay raises productivity: it elicits reciprocal effort, makes shirking costlier, reduces turnover, and attracts better applicants. Because firms do not cut wages even when job seekers are queuing, involuntary unemployment can persist in equilibrium.

Origins. Labour economics and macroeconomics. Akerlof (1982) (gift exchange), Shapiro and Stiglitz (1984) (shirking), with historical evidence from Raff and Summers (1987). The cost-minimizing condition goes back to Solow (1979).

Level of analysis. Firm and labour market.

Key constructs. Wage premium, effort, monitoring intensity, detection probability, unemployment as a discipline device, gift exchange, reference (fair) wage, turnover cost.

Boundary conditions. Effort is imperfectly observable or contractible, turnover is costly, and workers compare pay with a reference. Where piece rates are feasible, efficiency wage motives weaken.
Related theories. Agency theory (Section 6.1), Social preferences, Equity theory (Section 19.3), Psychological contract (Section 15.8), Organizational justice (Section 15.9), Tournament theory (Section 15.13).

17.5.1 Origins and intellectual history

Competitive theory predicts that wages fall until the labour market clears, so persistent unemployment and firms' reluctance to cut pay in recessions are puzzles. Efficiency wage theory resolves them by noting that the wage is also an incentive. If productivity depends on the wage, a firm minimizes cost per unit of effective labour, not cost per worker. With effort $e(w)$, the firm chooses the wage that minimizes $w/e(w)$, which yields the condition that the elasticity of effort with respect to the wage equals one (Solow 1979):

$$\frac{w\, e'(w)}{e(w)} \;=\; 1.$$

If this wage exceeds the market-clearing level, the firm will not lower it, and unemployment persists.

Two micro-foundations dominate. In the **shirking model** of Shapiro and Stiglitz (1984), shirkers are caught with probability q and fired, a threat that bites only if losing the job is costly. Akerlof (1982), inspired by a study of clerical workers who routinely exceeded their output standard, modelled employment as **partial gift exchange**: the firm pays above the outside option, and workers reciprocate with effort above the minimum. Other variants emphasize **turnover** costs and **adverse selection (Section 8.1)**.

17.5.2 Core logic and propositions

In the shirking model, a worker prefers working to shirking when the wage satisfies the **no-shirking condition** (NSC):

$$w \;\geq\; \bar{w} + e + \frac{e}{q}\left(r + \frac{b}{u}\right),$$

where $\bar{w}$ is the unemployment benefit, e the disutility of effort, q the detection probability, r the discount rate, b the exogenous separation rate, and u the unemployment rate. As unemployment falls, a fired worker finds a new job quickly, so the wage needed to deter shirking rises without limit. Full employment is incompatible with discipline: some unemployment is necessary to make dismissal a credible threat. The family of models yields these propositions:

- **P1 (Wage premium).** Firms pay above the market-clearing wage when effort, quality, or retention depends on the wage.
- **P2 (Monitoring substitution).** Wage premiums are higher where monitoring is costly or detection is unlikely.

- **P3 (Reciprocity).** Wage gifts perceived as generous relative to a reference wage elicit higher effort (Akerlof 1982).
- **P4 (Involuntary unemployment).** Equilibrium unemployment persists even though unemployed workers would accept the going wage (Shapiro and Stiglitz 1984).
- **P5 (Wage rigidity).** Firms resist cutting wages in downturns because cuts damage morale and effort.

17.5.3 Mechanism

Figure 17.5 plots the no-shirking condition against labour demand. Equilibrium occurs where the curves cross, to the left of full employment, and the gap is involuntary unemployment.

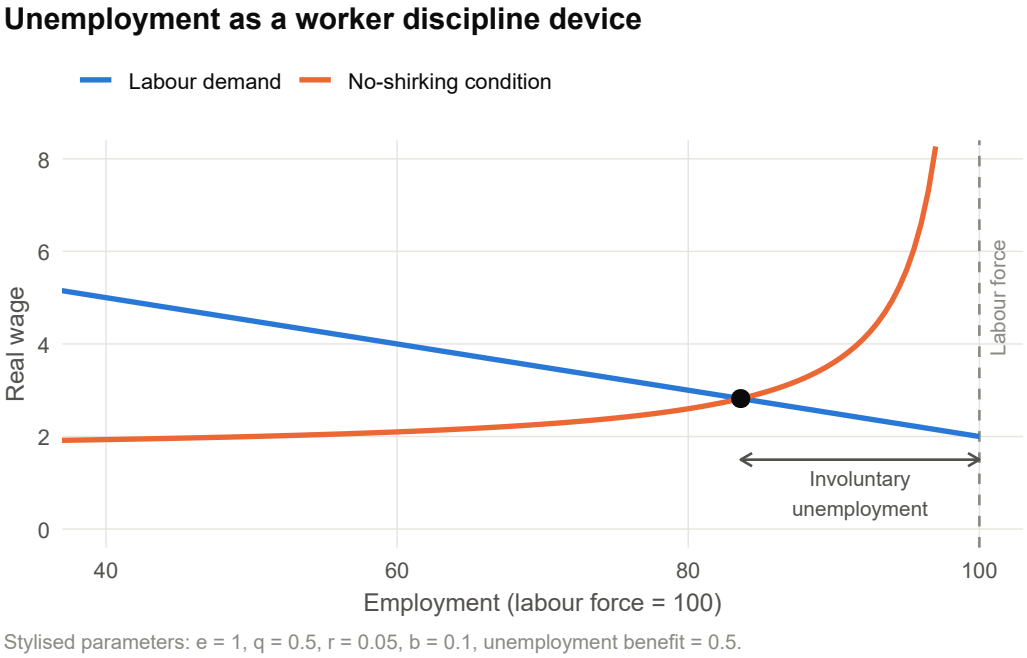


Figure 17.5. Stylised Shapiro-Stiglitz model. The no-shirking wage rises steeply as employment approaches the labour force, so equilibrium (where labour demand meets the no-shirking condition) leaves some workers involuntarily unemployed.

Table 17.6 compares the main variants of the theory.

Table 17.6. Variants of efficiency wage theory.

Variant	Why higher wages raise productivity	Key prediction	Illustrative evidence
Shirking	Job loss is costlier, so workers shirk less	Premiums higher where monitoring is hard	Cappelli & Chauvin (1991)
Gift exchange	Workers reciprocate generous pay with effort	Effort rises with wage relative to reference	Fehr, Kirchsteiger & Riedl (1993)
Turnover	Workers quit less, saving hiring and training costs	Premiums higher where training is firm-specific	Raff & Summers (1987)
Adverse selection	Higher wages attract a better applicant pool	Applicant quality rises with posted wage	Applicant pools at high-wage employers

Table 17.6. Variants of efficiency wage theory. *(continued)*

Variant	Why higher wages raise productivity	Key prediction	Illustrative evidence
Morale and wage rigidity	Pay cuts are seen as unfair and damage morale	Firms prefer layoffs to pay cuts in downturns	Bewley (1999)

17.5.4 Evidence and extensions

Raff and Summers (1987) examined the Ford Motor Company's introduction of the five-dollar day in 1914. After the change, turnover and absenteeism fell sharply, long queues of applicants formed at the plant, and the authors found evidence of productivity gains, a pattern they interpret as consistent with efficiency wages rather than a simple response to a labour shortage. Krueger and Summers (1988) documented large and persistent wage differences across industries for observably similar workers, which are hard to reconcile with competitive labour markets. Cappelli and Chauvin (1991) found that plants paying larger premiums over local wages had lower rates of disciplinary dismissal, as the shirking model predicts.

In experimental labour markets, Fehr, Kirchsteiger, and Riedl (1993) found that employers offered wages well above the market-clearing level and workers reciprocated with higher effort. Interviewing managers, Bewley (1999) concluded that firms avoid pay cuts mainly to protect morale. In the field, Gneezy and List (2006) found that an unexpectedly high wage raised effort in the first hours of a short task, but the effect faded quickly.

DEBATE AND EVIDENCE

Efficiency wages or something else? Inter-industry wage differentials may reflect unmeasured worker ability, compensating differentials for unpleasant work, or rent sharing rather than efficiency wages. Deferred compensation and seniority-based pay can in principle discipline workers without unemployment, which raises the question of why firms rely on wage premiums. The decay of gift-exchange effects in short field experiments (Gneezy and List 2006) suggests that reciprocity may be strongest in long-term relationships, precisely where it is hardest to test cleanly.

17.5.5 Applications in business

BUSINESS CASE

Ford's five-dollar day and Costco's wage policy. Ford's 1914 policy, which roughly doubled pay for eligible workers, is the canonical efficiency wage case (Raff and Summers 1987). A modern counterpart is the comparison between Costco and Walmart's Sam's Club. Cascio (2006) documented that Costco paid considerably higher wages and benefits than its warehouse-club rival and argued that it was rewarded with lower employee turnover and higher productivity. The logic is the one in Table 17.6: the relevant cost is labour cost per unit of effective output, and when turnover and service quality depend on pay, the cheapest wage per worker can be the most expensive wage per unit of output.

17.5.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Establishments that pay a larger premium over the local market wage will experience lower voluntary turnover and fewer disciplinary dismissals.
- *H2.* The effect of a wage increase on effort will be larger when the increase is framed as discretionary (a gift) than when it is mandated.

Designs. Natural experiments from minimum wage changes or firm-level wage policy changes, linked employer-employee panels, field experiments with temporary workers, laboratory gift-exchange games.

COMMON PITFALL

Assuming any wage increase pays for itself. The theory says that the cost-minimizing wage may exceed the market wage, not that higher wages always raise profits. A premium that is not noticed, not attributed to the employer's generosity, or matched by every competitor yields little extra effort.

17.6 Network Effects

THEORY CARD

Core claim. When the value of a product to each user rises with the number of other users, demand exhibits positive feedback. Markets can **tip** toward a single product or standard, expectations about who will win become self-fulfilling, and an inferior technology can become **locked in**.

Origins. Industrial organization. M. L. Katz and Shapiro (1985), Farrell and Saloner (1985), Arthur (1989), with the historical case of the typewriter keyboard in P. A. David (1985).

Level of analysis. Market and industry.

Key constructs. Direct and indirect network effects, installed base, expectations, compatibility, switching costs, tipping, lock-in, excess inertia.

Boundary conditions. Value depends substantially on network size, multi-homing is costly, and products are not highly differentiated. Local network effects and cheap multi-homing reduce tipping.

Related theories. Two-sided markets, Diffusion of innovations (Section 10.1), Dominant design (Section 12.5), First-mover advantage (Section 11.5), Bass model (Section 10.2).

17.6.1 Origins and intellectual history

Direct network effects arise when users benefit directly from others' adoption (telephones, messaging). **Indirect network effects** arise when a larger user base attracts more complementary products (consoles and games). M. L. Katz and Shapiro (1985) modelled competition among firms whose products exhibit network externalities and showed that consumer expectations about network size determine equilibrium outcomes. Farrell and Saloner (1985) analysed the coordination problem of switching to a new standard and identified **excess inertia**, in which users stick with an old standard because no one wants to move first, and its mirror image, excess momentum.

Arthur (1989) placed these ideas in a broader frame. Under increasing returns to adoption, the process of technology competition is path dependent: small historical events early in the process can determine which technology wins, the outcome is not predictable in advance, and the winner need not be the better technology. P. A. David (1985) argued that the QWERTY keyboard layout was such a case.

17.6.2 Core logic and propositions

A standard formulation gives consumer i utility from product k of

$$u_{ik} \;=\; r_{ik} + v\big(n_k^e\big) - p_k,$$

where r_{ik} is stand-alone value, n_k^e is the expected number of users, and $v(\cdot)$ is increasing. Because utility depends on expectations about others' choices, there can be multiple **fulfilled-expectations equilibria**: if everyone expects product A to win, everyone adopts A, and the expectation is confirmed. The theory implies:

- **P1 (Tipping).** With strong network effects and costly multi-homing, markets tend toward a single dominant product.
- **P2 (Expectations).** Firms that shape expectations about future network size (through announcements, alliances, or commitments) can win even with a smaller installed base (M. L. Katz and Shapiro 1985).
- **P3 (Compatibility strategy).** Firms with large networks or strong reputations tend to oppose compatibility, while smaller firms favour it (M. L. Katz and Shapiro 1985).
- **P4 (Path dependence and lock-in).** Early random events can decide the outcome, and an inferior standard can persist because switching requires coordination (Arthur 1989, Farrell and Saloner 1985).
- **P5 (Limits).** Differentiation, local network effects, and cheap multi-homing reduce winner-take-all outcomes.

17.6.3 Mechanism

Figure 17.6 simulates Arthur's argument. Two technologies start with identical support, and each new adopter chooses technology A with a probability that rises more than proportionally with A's current share. Chance early choices decide the winner, and most runs drift toward dominance by one technology.

17.6.4 Evidence and extensions

M. L. Katz and Shapiro (1994) review evidence and theory on systems competition. Empirically, estimating network effects is difficult because users who are connected tend to be similar and are exposed to common shocks. Tucker (2008) used variation in the adoption of a video-messaging technology within an investment bank to separate network effects from correlated behaviour, and found that how much a colleague's adoption mattered depended on that colleague's position in the formal hierarchy and the informal communication network.

The lock-in claim has been contested. Liebowitz and Margolis (1990) re-examined the QWERTY case and argued that the evidence for the superiority of the rival Dvorak layout was weak and that

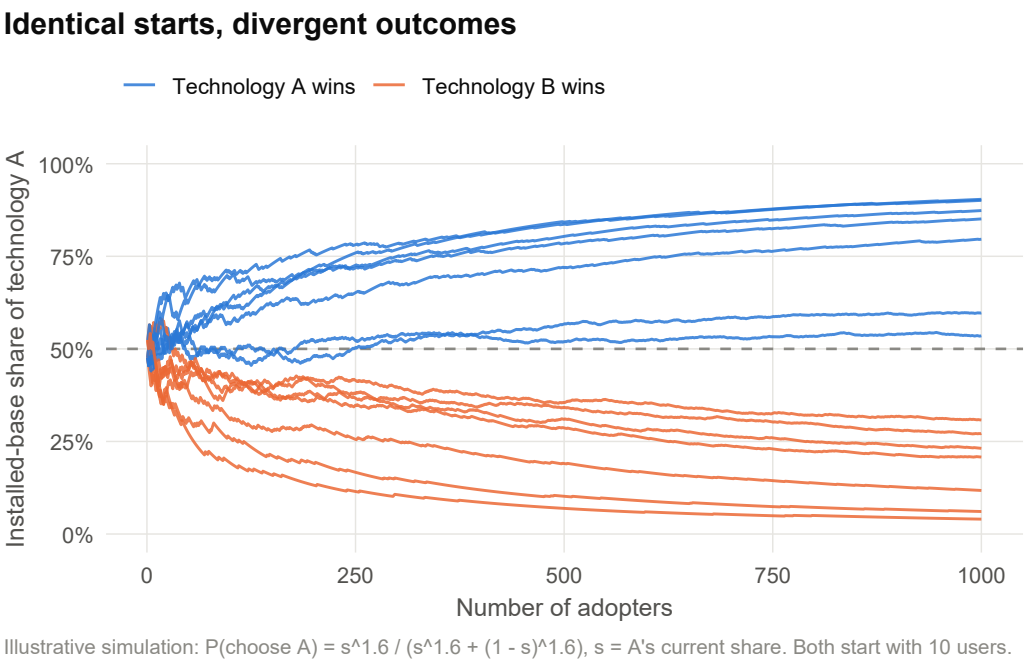


Figure 17.6. Illustrative simulation of increasing returns to adoption. Each line is one market with identical starting conditions. Because each adopter favours the technology with the larger installed base, chance early choices tip markets toward one technology or the other.

markets had not demonstrably locked in an inferior standard. Their critique shifted the burden of proof: documenting path dependence is not the same as documenting inefficient lock-in.

Table 17.7. Competitive strategies in markets with network effects.

Strategy	Network-effects logic	Example	Risk
Penetration pricing	Subsidize early adopters to build the installed base	Free tiers for consumer software	Customers anchor on a low price
Seeding and targeting hubs	Win central users whose adoption influences others	Launching in a dense community first	Hubs may multi-home
Compatibility and adapters	Borrow a rival's network	Backward-compatible consoles, file-format import	Strengthens the rival standard
Open standard or licensing	Grow the whole network quickly	Broad licensing of a format to manufacturers	Lose control of rents
Expectation management	Make adoption look inevitable	Preannouncements, alliance announcements	Credibility loss if late
Lock-in through switching costs	Make exit costly once users join	Proprietary data formats, loyalty programmes	Antitrust and customer resentment

DEBATE AND EVIDENCE

How strong are network effects in practice? Strategy writing often treats network effects as an impregnable moat. History suggests caution: dominant social networks have been displaced, and many networked markets support several competitors. Researchers increasingly distinguish the *strength* of network effects, their *scope* (global versus local), and the *cost of*

multi-homing, each of which moderates the tipping prediction.

17.6.5 Applications in business

BUSINESS CASE

VHS versus Betamax. Sony introduced the Betamax videocassette format in 1975 and JVC introduced VHS in 1976. The formats were incompatible, and the value of a recorder depended on the availability of pre-recorded tapes, an indirect network effect. JVC licensed VHS widely to other manufacturers, and early VHS machines offered longer recording times. As the VHS installed base grew, video rental stores stocked more VHS titles, which further increased VHS's advantage. By the late 1980s VHS dominated, and Sony began producing VHS recorders in 1988. The case illustrates tipping and the strategic value of licensing to build an installed base (Table 17.7). Whether Betamax was technically superior matters less to the theory than how an early lead in complements became decisive.

17.6.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The effect of installed base on a product's market share will be larger in categories with higher multi-homing costs.
- *H2.* Product preannouncements will slow adoption of a rival standard more in markets with strong indirect network effects.

Designs. Demand estimation with installed-base terms and instruments, peer-effects designs exploiting exogenous variation in network exposure, historical case comparisons of standards battles, agent-based simulation.

COMMON PITFALL

Confusing popularity with network effects. Network effects require that each user's value depends on others' use. Popularity driven by quality, advertising, or social learning produces similar adoption curves but different strategic implications. Identify the mechanism before invoking tipping.

17.7 Two-Sided Markets

THEORY CARD

Core claim. A platform that serves two distinct groups linked by **cross-side network effects** chooses not only a price level but a **price structure**. It typically subsidizes the side whose participation creates more value for the other side and whose demand is more price-sensitive, and earns its margin on the other side.

Origins. Industrial organization. Rochet and Tirole (2003), Caillaud and Jullien (2003), G. G. Parker and Van Alstyne (2005), Armstrong (2006), reviewed in Rochet and Tirole (2006) and

Rysman (2009).
Level of analysis. Platform and market.
Key constructs. Cross-side externality, price level versus price structure, subsidy side and money side, multi-homing, competitive bottleneck, chicken-and-egg problem.
Boundary conditions. A market is two-sided when the volume of transactions depends on how the total price is split between the sides, not only on its level. If the sides can pass charges through to each other, the market is effectively one-sided.
Related theories. Network effects, Transaction cost economics (Section 6.3), Competitive dynamics (Section 11.3), Resource dependence theory (Section 7.1).

17.7.1 Origins and intellectual history

Many networked businesses connect two groups that value each other: cardholders and merchants, players and game developers, riders and drivers. Rochet and Tirole (2003) showed that in such markets, the allocation of charges across sides matters, and that the profit-maximizing structure can involve prices below marginal cost on one side. G. G. Parker and Van Alstyne (2005) reached a similar conclusion from information-product design: Adobe could give away its Acrobat Reader to consumers because doing so raised the value of its paid authoring software to document creators. Caillaud and Jullien (2003) analysed the chicken-and-egg problem of attracting both sides at once and the "divide-and-conquer" strategy of subsidizing one side to attract the other.

17.7.2 Core logic and propositions

Rochet and Tirole (2006) offer a precise definition. Let a platform charge a^B per transaction to buyers and a^S to sellers. The market is two-sided if the volume of transactions depends on a^B and a^S separately and not only on their sum $a^B + a^S$. In Armstrong (2006)'s model of a monopoly platform, the utility of a group-1 member is $u_1 = \alpha_1 n_2 - p_1$, where α_1 is the benefit from each member of group 2, and participation is $n_1 = \phi_1(u_1)$. The profit-maximizing price for group 1 is

$$p_1 = f_1 - \alpha_2 n_2 + \frac{\phi_1(u_1)}{\phi_1'(u_1)},$$

where f_1 is the cost of serving a group-1 member. The middle term is the key insight: the price to group 1 is lowered by the external benefit an additional member confers on group 2, and can be zero or negative. The theory implies:

- **P1 (Price structure).** The side that generates larger cross-side benefits and has more elastic demand receives lower prices, possibly below cost.
- **P2 (Skewed pricing is not predation).** A below-cost price on one side is consistent with profit maximization and does not by itself indicate anticompetitive intent (Rochet and Tirole 2006).
- **P3 (Competitive bottlenecks).** When one side multi-homes and the other single-homes, platforms compete for the single-homing side and charge the multi-homing side for access to it (Armstrong 2006).
- **P4 (Chicken and egg).** Platforms must solve a coordination problem at launch, often by subsidizing or seeding one side (Caillaud and Jullien 2003).

17.7.3 Mechanism

Figure 17.7 shows the feedback loop that the price structure manages.

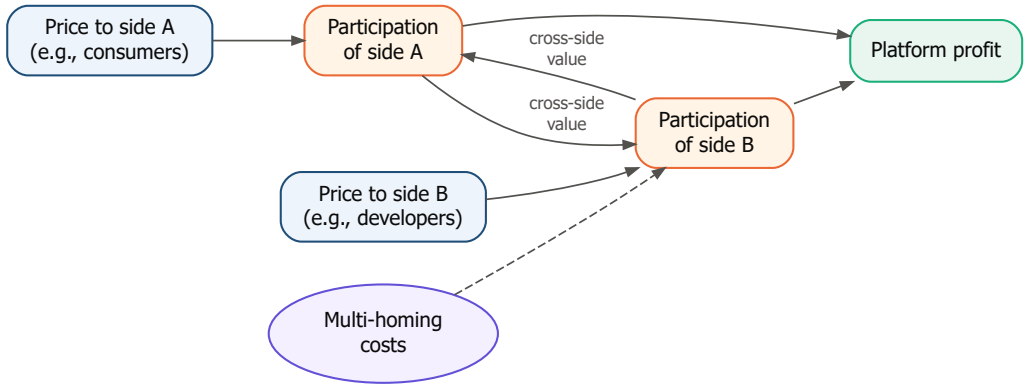


Figure 17.7. Mechanism of a two-sided platform: the price charged to each side determines its participation, and each side's participation raises the value of the platform to the other side. The platform chooses which side to subsidize.

Table 17.8 gives examples of price structures.

Table 17.8. Price structures on two-sided platforms.

Platform	Side A	Side B	Typical price structure
Payment cards	Cardholders	Merchants	Cardholders often pay little or receive rewards. Merchants pay fees per transaction
App stores	Device users	App developers	Many apps free to download. Developers pay commissions on paid apps and in-app purchases
Ride-hailing	Riders	Drivers	Both sides priced, with dynamic prices balancing supply and demand
Video game consoles	Players	Game publishers	Hardware has often been sold near cost. Publishers pay royalties
Search and social media	Users	Advertisers	Users pay nothing. Advertisers pay for attention
Document formats	Readers	Authors	Free reader software. Paid authoring software (Parker & Van Alstyne, 2005)

17.7.4 Evidence and extensions

Payment cards are the most studied case. Interchange fees, paid by the merchant's bank to the cardholder's bank, shift costs between the sides, and regulators have capped them: the European Union's 2015 Interchange Fee Regulation capped consumer debit and credit interchange at 0.2 and 0.3 percent of transaction value, and in the United States the 2010 Dodd-Frank Act (through the Durbin Amendment) capped debit interchange for large issuers. Two-sidedness also reached the United States Supreme Court: in *Ohio v. American Express* (2018), the Court held that the credit-card market must be analysed as two-sided, weighing effects on both merchants and cardholders. Rysman (2009) reviews the broader empirical literature.

DEBATE AND EVIDENCE

Efficient structure or market power? Two-sided theory explains why a platform might charge one side heavily without implying abuse. Critics worry that the theory can be used to excuse high charges to a side that has no alternative, for instance merchants who cannot refuse a popular card or developers who cannot reach users except through a dominant app store. The analytical challenge is to separate the *structure* of prices, which may be legitimately skewed, from their *level*, which can reflect market power. A second debate concerns definitions: almost every firm intermediates between some groups, so two-sided analysis requires showing that cross-side effects are material.

17.7.5 Applications in business

BUSINESS CASE

App store commissions. Apple's App Store has long charged developers a standard commission of 30 percent on paid apps and in-app purchases of digital goods, while letting users download free apps at no charge. Two-sided theory explains the structure: users single-home on a phone platform and are the side developers need to reach, so the platform subsidizes users and charges the side that multi-homes (a competitive bottleneck). The structure has been contested. In 2021 Apple introduced a Small Business Program that cut the commission to 15 percent for developers with up to one million dollars in annual proceeds, and litigation such as *Epic Games v. Apple* and the European Union's Digital Markets Act have targeted its payment and distribution rules. A price structure that maximizes short-run profit can provoke the regulatory and competitive responses that reshape it. Ride-hailing shows a different case: because drivers multi-home across apps and riders compare fares, both sides are price-sensitive, and dynamic pricing must balance them continuously, subject to the fairness limits discussed under Social preferences.

17.7.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Platforms will charge lower prices to the side whose participation generates stronger cross-side effects.
- *H2.* Regulatory caps on fees to one side will be partly offset by higher fees or lower subsidies on the other side (a "waterbed" effect).

Designs. Structural demand models with cross-side effects, difference-in-differences around fee regulation, natural experiments in platform entry, field experiments that vary fees to one side.

COMMON PITFALL

Applying one-sided price-cost tests. Concluding that a zero price to users is predatory, or that a high fee to merchants is exploitative, by looking at one side alone ignores the cross-side externality. Analyse total price and total cost across both sides before drawing conclusions about efficiency or competition.

17.8 Auction Theory

THEORY CARD

Core claim. The rules of a sale (who bids, in what format, with what information and reserve price) determine bidding strategies, revenue, and whether the good goes to the bidder who values it most. Under benchmark assumptions many formats yield the same expected revenue, but departures from those assumptions, especially common values, asymmetric bidders, collusion, and entry costs, make design decisive.
Origins. Economics and game theory. Vickrey (1961), Myerson (1981), Riley and Samuelson (1981), and P. R. Milgrom and Weber (1982). Vickrey shared the 1996 Nobel Memorial Prize, and Milgrom and Robert Wilson received the 2020 prize.
Level of analysis. Market mechanism, bidder, and seller.
Key constructs. Private versus common values, bid shading, dominant strategy, revenue equivalence, reserve price, winner's curse, affiliation, linkage principle, collusion, entry.
Boundary conditions. The benchmark results assume risk-neutral bidders with independent private values from a common distribution, no collusion, and a fixed number of bidders.
Related theories. Game theory (Section 11.4), Information asymmetry (Section 8.1), Search theory, Social preferences.

17.8.1 Origins and intellectual history

Vickrey (1961) founded the field by analysing auctions as games of incomplete information. He showed that when the highest bidder pays the *second*-highest bid, bidding one's true value is a dominant strategy. He also anticipated the **revenue equivalence theorem**, generalized by Myerson (1981) and Riley and Samuelson (1981): with risk-neutral bidders and independent private values, any auction that gives the good to the highest bidder and zero surplus to a bidder with the lowest possible value yields the same expected revenue.

P. R. Milgrom and Weber (1982) extended the analysis to settings in which bidders' values are **affiliated**, as when the value of an oil tract depends partly on the unknown amount of oil, which all bidders estimate with noise. In such settings the English auction yields expected revenue at least as high as the second-price auction, which in turn yields at least as much as the first-price auction, and sellers gain by committing to reveal information they hold (the **linkage principle**). Petroleum engineers had already noticed a practical counterpart: Capen, Clapp, and Campbell (1971) argued that oil companies bidding for Gulf of Mexico leases earned disappointing returns because the winner of a common-value auction is the bidder who most overestimated the tract's value, the **winner's curse**.

17.8.2 Core logic and propositions

With n bidders whose private values are independent and uniformly distributed on $[0, 1]$, the equilibrium strategies are:

$$b^{\text{2nd}}(v) = v, \qquad b^{\text{1st}}(v) = \frac{n-1}{n}\, v.$$

In the second-price auction, bidders bid their values. In the first-price auction, they shade their bids, but the winner's expected payment in both cases equals the expected second-highest value,

$(n-1)/(n+1)$. In a common-value auction where the item is worth V to everyone and bidder i observes a noisy signal s_i, winning is bad news:

$$\mathbb{E}\left[V \mid s_i,\ s_i = \max_j s_j\right] \ <\ \mathbb{E}\left[V \mid s_i\right],$$

so a rational bidder must shade more as the number of competitors grows. Table 17.9 summarizes common formats. The propositions are:

- **P1 (Truthful bidding).** In second-price and English auctions with private values, bidding one's value is a (weakly) dominant strategy (Vickrey 1961).
- **P2 (Revenue equivalence).** Under the benchmark assumptions, standard formats yield equal expected revenue (Myerson 1981, Riley and Samuelson 1981).
- **P3 (Linkage principle).** With affiliated values, formats that link the price to more information, and sellers who disclose information, raise expected revenue (P. R. Milgrom and Weber 1982).
- **P4 (Winner's curse).** In common-value auctions, naive bidders who bid their estimates lose money on average, and the loss grows with the number of bidders (Capen, Clapp, and Campbell 1971, R. H. Thaler 1988).
- **P5 (Design for entry and against collusion).** In practice, attracting bidders and preventing collusion matter more than the fine points of the benchmark ranking (Klemperer 2002).

Table 17.9. Common auction formats.

Format	Rules	Equilibrium bidding (private values)	Typical uses
English (ascending)	Price rises until one bidder remains. Winner pays the final price	Stay in until price reaches own value	Art, real estate, livestock, online consumer auctions
Dutch (descending)	Price falls until a bidder accepts	Strategically equivalent to first-price sealed bid	Flowers, fish, perishable goods
First-price sealed bid	Highest bid wins and pays own bid	Shade bid below value	Public procurement, construction tenders
Second-price sealed bid (Vickrey)	Highest bid wins and pays second-highest bid	Bid true value (dominant strategy)	Benchmark in theory, some online ad systems
Generalized second price	Several ranked slots. Each winner pays the next bid	Not truthful in general	Sponsored search advertising
Simultaneous ascending or clock auction	Many items auctioned together over rounds	Bid on items or packages as prices rise	Spectrum licences, electricity, emission allowances

17.8.3 Mechanism

Figure 17.8 simulates the winner's curse. Each bidder naively bids an unbiased but noisy estimate of a common value. The winning bid is the highest estimate, which is biased upward, increasingly so as bidders are added.

17.8.4 Evidence and extensions

Kagel and Levin (1986) found that inexperienced laboratory bidders in common-value auctions frequently lost money, and that the problem worsened in larger groups. R. H. Thaler (1988) reviews

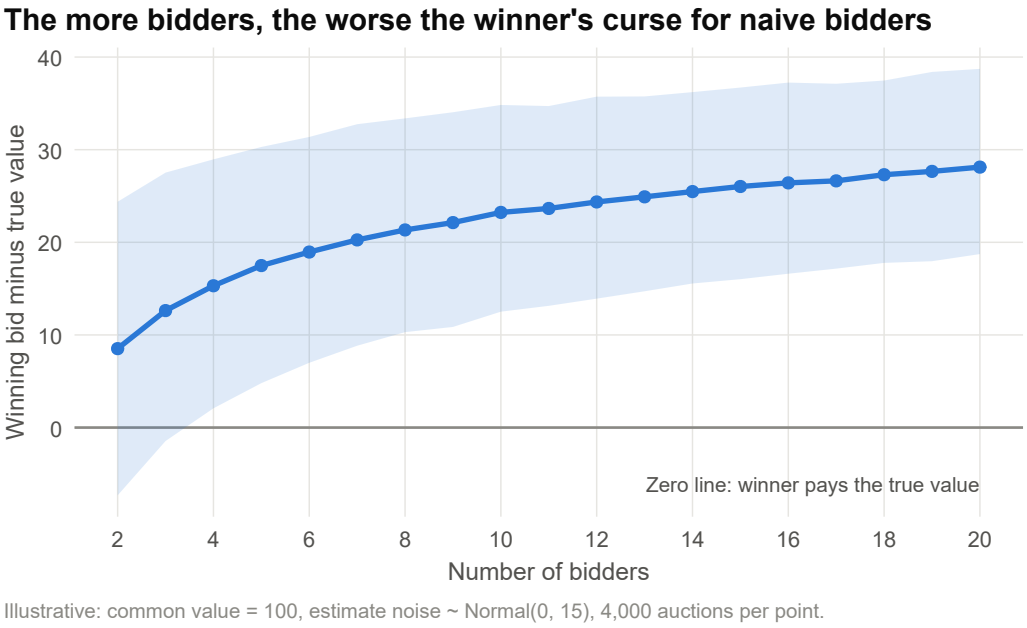


Figure 17.8. Illustrative simulation of the winner's curse. A tract is worth 100 to every bidder. Each bidder receives an unbiased estimate with noise (standard deviation 15) and bids it. The winner's expected overpayment rises with the number of bidders. The band shows the 10th to 90th percentile across simulated auctions.

field evidence from settings such as oil leases, corporate takeovers, and professional sports free agency.

The most consequential applications are in market design. Decades after Coase (1959)'s proposal, the United States Federal Communications Commission began auctioning spectrum licences in 1994, using a simultaneous multiple-round ascending format developed with input from auction theorists (McMillan 1994, P. Milgrom 2004). The United Kingdom's 2000 auction of third-generation mobile licences raised about 22.5 billion pounds, while auctions of similar licences in some other European countries raised far less per capita, a contrast that Klemperer (2002) attributes to design choices that affected entry and the scope for tacit collusion. In online advertising, search engines adopted the **generalized second-price** (GSP) auction to sell ranked ad slots. Edelman, Ostrovsky, and Schwarz (2007) showed that, unlike the Vickrey auction, GSP does not make truthful bidding a dominant strategy, but that it has an equilibrium whose outcome mirrors the Vickrey-Clarke-Groves mechanism.

DEBATE AND EVIDENCE

Theory versus practice. Klemperer (2002) argues that "what really matters" in practical auction design is not the revenue-ranking of formats under textbook assumptions but robustness to collusion, entry deterrence, and political pressure. Ascending auctions are vulnerable to bidders signalling and dividing markets. Sealed-bid auctions are more robust to collusion and encourage entry by weaker bidders. A second debate concerns behavioural deviations: if bidders systematically fall prey to the winner's curse or overbid, the seller's optimal design changes.

17.8.5 Applications in business

BUSINESS CASE

Auctions in procurement and advertising. Firms use **reverse auctions**, in which suppliers bid prices downward, to buy standardized inputs. Auction theory offers three lessons. Qualifying more suppliers usually matters more than the choice of format. When quality is not fully contractible, a pure price auction invites a supplier-side winner's curse followed by quality shading, so buyers often use scoring rules that weigh price against quality. Reserve prices protect the buyer when competition is thin. In online advertising, the choice between second-price and first-price formats is a live design question: in 2019 Google moved its Ad Manager exchange to a first-price auction, which shifted the burden of bid shading onto advertisers and their bidding tools.

17.8.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Winning bids in common-value auctions will exceed ex post values more often when the number of bidders is large and bidders are inexperienced.
- *H2.* Procurement auctions with more qualified bidders will yield lower prices, with diminishing marginal effects.

Designs. Laboratory auctions, structural estimation of bidder values from bid data, natural experiments in auction rules (reserve prices, format changes), comparison of auction and negotiated sales.

COMMON PITFALL

Assuming the winner got a bargain. A high winning bid in a common-value setting is evidence of optimism, not of superior value. Researchers who treat winning bids as measures of value, and managers who treat winning as success, ignore selection on the most favourable estimate. Always ask what the winner's estimate implies conditional on having won.

17.9 New Institutional Economics

THEORY CARD

Core claim. Institutions, the formal rules and informal constraints that structure human interaction, determine transaction costs and incentives, and therefore explain long-run differences in economic performance across societies. Institutions change slowly and path-dependently, shaped by the interests of those with power to change them.

Origins. Economics and economic history. North (1990), North (1991), and Williamson (2000), building on Coase (1937) and Coase (1960). Nobel Memorial Prizes went to North in 1993, Williamson in 2009, and Acemoglu, Johnson, and Robinson in 2024.

Level of analysis. Institution and society, with implications for firms and transactions.

Key constructs. Formal rules, informal constraints, enforcement, organizations as players,

credible commitment, security of property rights, path dependence, levels of institutional analysis.
Boundary conditions. Most informative for comparisons across societies or over long periods, and for explaining why identical contracts or technologies perform differently in different settings.
Related theories. Transaction cost economics (Section 6.3), Institutional theory (Section 7.2), Institutional voids (Section 13.5), Legitimacy (Section 7.3), Coase theorem.

17.9.1 Origins and intellectual history

Neoclassical growth theory explained output with capital, labour, and technology, but could not explain why societies with access to the same technologies diverged so dramatically. North argued that the answer lies in institutions, which he defined as "the humanly devised constraints that structure political, economic and social interaction" (North 1991, 97), both formal (constitutions, laws, property rights) and informal (sanctions, taboos, customs, codes of conduct). Institutions are the **rules of the game**. Organizations such as firms, unions, and parties are the **players**, who exploit the opportunities institutions create and lobby to change the rules (North 1990). Williamson (2000) placed these ideas in a four-level scheme, from slow-changing social structures to the everyday allocation of resources (Figure 17.9). Transaction cost economics occupies the governance level and North's work the institutional environment.

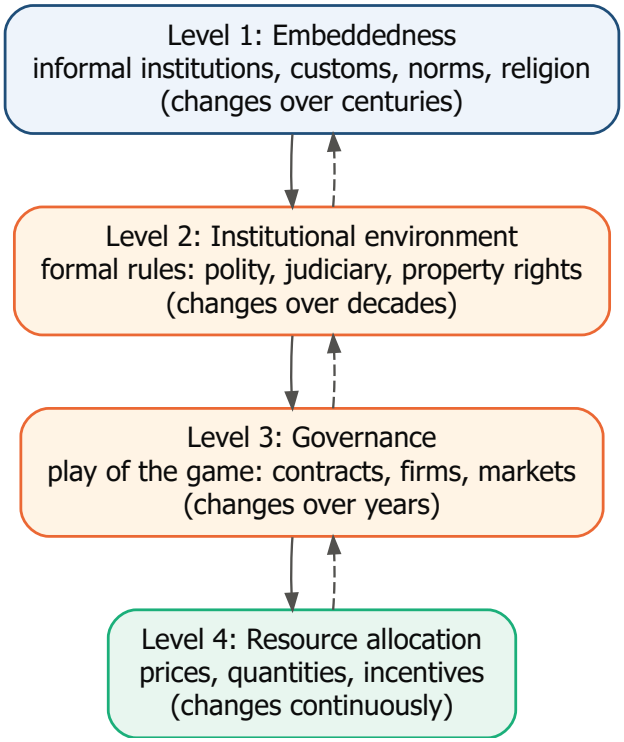


Figure 17.9. Four levels of social analysis in new institutional economics, after Williamson (2000). Higher levels constrain lower levels (solid arrows). Feedback from lower to higher levels (dashed arrows) is slower and weaker.

17.9.2 Core logic and propositions

The argument runs from rules to incentives to performance. Secure property rights and impartial enforcement lower the cost of impersonal exchange and encourage investment. Rules that permit expropriation push exchange into small, personal networks where reputation substitutes for courts. The propositions are:

- **P1 (Institutions and transaction costs).** Institutions that protect property and enforce contracts reduce transaction costs and expand the range of feasible exchange (North 1990).
- **P2 (Credible commitment).** Economic growth requires that the state commit not to expropriate. Political institutions that constrain rulers make such commitments credible (North and Weingast 1989).
- **P3 (Path dependence).** Because institutions create organizations with a stake in preserving them, inefficient institutions can persist for long periods (North 1990).
- **P4 (Hierarchy of levels).** Informal institutions constrain formal rules, which constrain governance choices, which constrain resource allocation (Williamson 2000).
- **P5 (Private ordering).** Where public institutions are weak, private institutions such as reputation-based coalitions can support exchange (Greif 1993).

17.9.3 Evidence and extensions

The most influential empirical study is Acemoglu, Johnson, and Robinson (2001). They argued that European colonizers established extractive institutions where disease made settlement dangerous and institutions protecting property where Europeans settled in large numbers, and that those institutions persisted. Using historical settler mortality as an instrument for current institutions (measured by protection against expropriation risk), they estimated large effects of institutions on income per capita. Rodrik, Subramanian, and Trebbi (2004) similarly concluded that institutional quality dominates geography and trade integration as a determinant of income levels.

Greif (1993) illustrated private ordering: eleventh-century Maghribi traders in the Mediterranean formed a coalition in which an overseas agent who cheated one merchant would not be hired by the others, a reputation mechanism that supported long-distance agency relationships without courts. North and Weingast (1989) argued that the settlement after England's Glorious Revolution of 1688 made government commitments to repay debt credible.

Table 17.10 distinguishes new institutional economics from the sociological Institutional theory (Section 7.2) covered earlier in the book, a frequent source of confusion.

Table 17.10. New institutional economics and sociological institutional theory compared.

Dimension	New institutional economics	Sociological institutional theory
Discipline	Economics, economic history, law	Sociology, organization theory
Institutions are	Rules of the game that shape incentives	Taken-for-granted beliefs, norms, and scripts
Actors are	Boundedly rational, interest seeking	Embedded, seeking legitimacy
Core mechanism	Transaction costs and incentives	Isomorphism and legitimacy
Typical outcome	Efficiency, investment, growth	Similarity of organizational forms and practices
Representative sources	North (1990), Williamson (2000)	Meyer & Rowan (1977), DiMaggio & Powell (1983)

DEBATE AND EVIDENCE

Do institutions cause growth? Glaeser et al. (2004) argued that the institutional measures used in cross-country studies capture policy outcomes rather than durable constraints, and that human capital is a more fundamental cause of both growth and institutional improvement. Albouy (2012) questioned the settler mortality data underlying Acemoglu, Johnson, and Robinson (2001), arguing that many observations were imputed from other regions and that the results weaken with alternative codings. The debate has pushed the field toward within-country natural experiments with sharper identification. A conceptual critique also persists: if "institutions" include everything from constitutions to culture, the theory risks becoming unfalsifiable.

17.9.4 Applications in business

BUSINESS CASE

Contracting where courts are slow. Firms operating where contract enforcement is weak rely more on relational contracting, vertically integrate activities that would be outsourced elsewhere, form diversified business groups (see Institutional voids (Section 13.5)), and write arbitration clauses that move disputes to more reliable forums. The 1958 New York Convention on the Recognition and Enforcement of Foreign Arbitral Awards is itself an institution that lowers the transaction costs of cross-border contracting: by committing signatory states to enforce arbitral awards, it lets firms import credible enforcement that local courts may not provide.

17.9.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Firms in jurisdictions with slower contract enforcement will be more vertically integrated, controlling for industry.
- *H2.* The positive effect of foreign direct investment on local supplier productivity will be larger where property rights are secure.

Designs. Cross-country panels with fixed effects, instrumental variables based on historical shocks, regression discontinuity at historical or administrative borders, comparative historical case studies.

COMMON PITFALL

Treating "institutions" as a residual. When a model fails to explain cross-country differences, it is tempting to attribute the remainder to "institutions". The theory is useful only if the specific rule, its enforcement mechanism, and the incentive it changes are identified and measured. Specify which institution, at which of Williamson's levels, alters which transaction cost.

17.10 Social Preferences

THEORY CARD

Core claim. Many people care not only about their own material payoff but also about the payoffs of others and the fairness of how outcomes come about. They sacrifice money to punish unfair treatment and to reward kindness, and these preferences constrain the prices, wages, and contracts that firms can sustain.

Origins. Experimental and behavioural economics. Güth, Schmittberger, and Schwarze (1982) introduced the ultimatum game. Kahneman, Knetsch, and Thaler (1986) documented community standards of fairness in pricing. Formal models include Rabin (1993), Fehr and Schmidt (1999), and Bolton and Ockenfels (2000). Fehr and Gächter (2000b) review the evidence on reciprocity.

Level of analysis. Individual and dyad, with market-level consequences.

Key constructs. Inequity aversion (disadvantageous and advantageous), reciprocity, intentions, reference transaction, dual entitlement, altruistic punishment, conditional cooperation.

Boundary conditions. Effects are strongest in bilateral relationships, repeated interactions, and settings where fairness norms are salient. In anonymous competitive markets, inequity-averse agents may behave much like selfish ones (Fehr and Schmidt 1999).

Related theories. Equity theory (Section 19.3), Organizational justice (Section 15.9), Reference price (Section 27.11), Efficiency wages, Collective action, Prospect theory (Section 24.2).

17.10.1 Origins and intellectual history

In the **ultimatum game** studied by Güth, Schmittberger, and Schwarze (1982), a proposer divides a sum of money and a responder either accepts the split or rejects it, in which case both get nothing. Self-interest predicts that the proposer offers the smallest positive amount and the responder accepts. Instead, proposers typically offer a substantial share, and responders often reject low offers, paying to punish what they see as unfair. The pattern has been replicated many times (C. Camerer and Thaler 1995).

At about the same time, Kahneman, Knetsch, and Thaler (1986) surveyed telephone respondents about the fairness of everyday pricing and wage decisions. A hardware store that raises the price of snow shovels from 15 to 20 dollars the morning after a large snowstorm was judged unfair by 82 percent of respondents. Firms were judged entitled to raise prices to protect profits when costs rose, but not to exploit shortages. This is the principle of **dual entitlement**: customers are entitled to the terms of a reference transaction, and firms to a reference profit.

17.10.2 Core logic and propositions

Fehr and Schmidt (1999) proposed a tractable model of **inequity aversion**. With n players and material payoffs x, player i's utility is

$$U_i(x) \;=\; x_i \;-\; \frac{\alpha_i}{n-1}\sum_{j\neq i}\max(x_j - x_i,\,0) \;-\; \frac{\beta_i}{n-1}\sum_{j\neq i}\max(x_i - x_j,\,0),$$

with $\beta_i \le \alpha_i$ and $0 \le \beta_i < 1$. The parameter α_i captures aversion to being behind, β_i aversion to being ahead. In a two-player ultimatum game with a pie of one, a responder offered share $s < 1/2$

accepts only if $s - \alpha(1 - 2s) \geq 0$, that is, if

$$s \geq \frac{\alpha}{1 + 2\alpha}.$$

A responder with $\alpha = 1$ rejects any offer below one third. Reciprocity models such as Rabin (1993) add that people respond to perceived *intentions*, not just outcomes: the same low offer is resented more when the proposer could have chosen a fair one. The theory's propositions are:

- **P1 (Costly punishment).** People reject unfair offers and punish unfair behaviour even at a cost to themselves (Güth, Schmittberger, and Schwarze 1982, Fehr and Gächter 2000b).
- **P2 (Reciprocity).** Kind actions elicit kind responses, which supports cooperation and gift exchange beyond what contracts enforce (Fehr and Gächter 2000b).
- **P3 (Dual entitlement).** Price and wage changes justified by cost increases are accepted. Changes that exploit demand shocks or market power are resented (Kahneman, Knetsch, and Thaler 1986).
- **P4 (Market context).** In competitive markets with many traders, fairness concerns have limited effect on prices. In bilateral relationships and bargaining, they can be decisive (Fehr and Schmidt 1999).

17.10.3 Mechanism

Figure 17.10 shows how responders who differ in inequity aversion, each rejecting offers below $\alpha/(1 + 2\alpha)$, generate a smoothly declining rejection curve. Table 17.11 compares the main models.

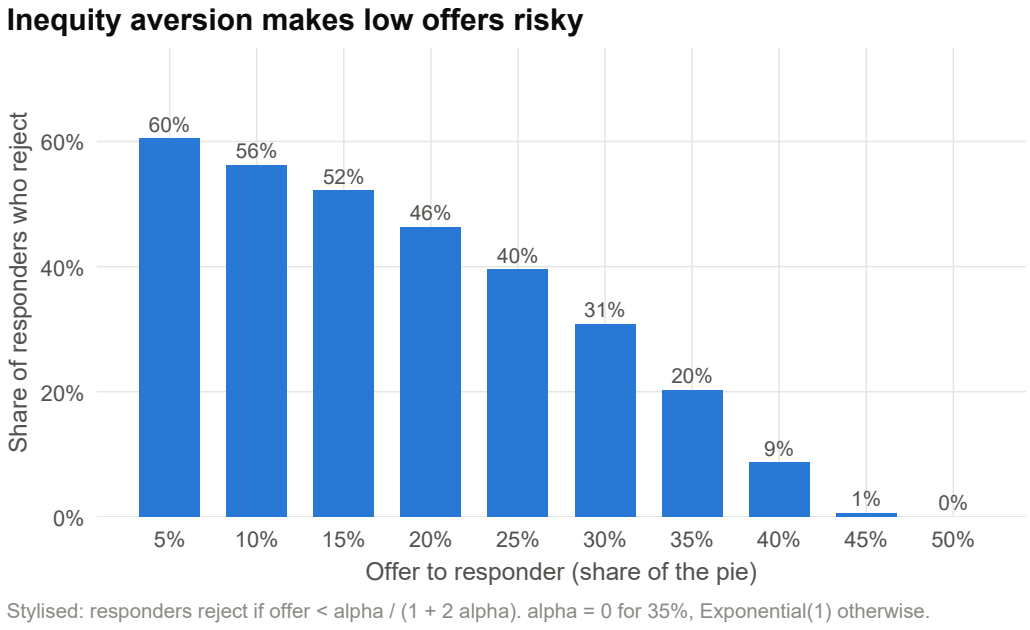


Figure 17.10. Stylised ultimatum-game rejection rates generated by a simulated population of responders: 35 percent purely self-interested and 65 percent inequity-averse with varying sensitivity. Low offers are frequently rejected and near-equal splits almost never are. Illustrative, not experimental data.

Table 17.11. Models of social preferences and fairness.

Model	What people care about	Signature prediction	Key source
Pure altruism	Others' payoffs, positively	Giving in dictator games	Classic welfare economics
Inequity aversion	Differences between own and others' payoffs	Rejection of low ultimatum offers, punishment of free riders	Fehr & Schmidt (1999)
Equity, reciprocity, competition (ERC)	Own share relative to the average	Similar to inequity aversion in two-player games	Bolton & Ockenfels (2000)
Intention-based reciprocity	Whether others meant to be kind or unkind	Same outcome judged differently by intention	Rabin (1993)
Dual entitlement	Reference transaction and reference profit	Cost-based price rises fair, demand-based ones unfair	Kahneman, Knetsch & Thaler (1986)

17.10.4 Evidence and extensions

The *direction* of ultimatum game results is widespread but their *magnitude* varies. Henrich et al. (2001) ran economic games, chiefly the ultimatum game, in 15 small-scale societies and found that offers and rejections varied considerably and were related to how much a society's economic life involved cooperation and market exchange. Oosterbeek, Sloof, and Kuilen (2004) meta-analysed cross-country variation in ultimatum game results, and university subject pools may not be representative of humanity (Henrich, Heine, and Norenzayan 2010). S. D. Levitt and List (2007) caution that laboratory settings may inflate social preferences because participants know they are observed, stakes are small, and the choice set is artificial. The broad consensus is that social preferences are real and heterogeneous: some people behave selfishly, many are conditional cooperators, and institutions determine which type sets the tone (Fehr and Gächter 2000b, Fischbacher, Gächter, and Fehr 2001).

DEBATE AND EVIDENCE

Preferences or norms? One interpretation is that people have stable preferences over the distribution of payoffs. Another is that behaviour in these games reflects social norms and the desire to be seen as fair, which depend on context and observability. Preference models imply that fair behaviour persists under anonymity, while norm models imply that it weakens when no one is watching or when the situation is framed as a market. Evidence supports elements of both.

17.10.5 Applications in business

BUSINESS CASE

Price gouging backlash. The fairness constraint documented by Kahneman, Knetsch, and Thaler (1986) recurs whenever firms raise prices in response to demand shocks. In 1999 Coca-Cola's chief executive drew widespread criticism after remarking to a magazine that vending machines could raise prices in hot weather, and the company stated that it had no plans to introduce such machines. Ride-hailing platforms faced similar criticism when surge pricing coincided with emergencies, and in 2014 Uber reached an agreement with the New York State Attorney General to limit surge pricing during declared emergencies and abnormal market disruptions, and many US states prohibit excessive price increases during declared emergencies.

The guidance is clear: increases framed as passing on costs are tolerated, increases that exploit scarcity are resented, and capping prices in emergencies can protect long-run goodwill.

17.10.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Customer complaints and churn following a price increase will be lower when the increase is attributed to cost rises than to demand.
- *H2.* Workers will reduce effort more after a wage cut attributed to labour-market slack than after an equal cut attributed to the firm's losses.

Designs. Laboratory games (ultimatum, dictator, trust, public goods), vignette surveys in the style of Kahneman, Knetsch, and Thaler (1986), field experiments on price communication, analysis of social media reactions to pricing events.

COMMON PITFALL

Assuming fairness norms are universal and fixed. Fairness judgements depend on the reference transaction, on what is customary in a market, and on culture. A surge price that is accepted for concert tickets may be condemned for bottled water after a hurricane. Measure the reference point before predicting a backlash, and expect norms to shift as practices become familiar.

17.11 Chapter Summary

KEY TAKEAWAYS

- Externalities can in principle be solved by bargaining, but positive transaction costs make rights assignment and institutional design decisive. Tradable permits are a Coasean market in rights (Coase 1960, Pigou 1920).
- Groups underprovide shared goods because benefits are non-excludable. Selective incentives, privileged members, and peer sanctions overcome free riding (Olson 1965, Fehr and Gächter 2000a).
- With incomplete contracts, ownership allocates residual control rights and shapes investment incentives. Integration helps one party and hurts another (Grossman and Hart 1986, Hart and Moore 1990).
- Costly search sustains price dispersion and frictional unemployment as equilibrium outcomes (Stigler 1961, Mortensen and Pissarides 1994).
- Firms may pay above market wages to elicit effort and reduce turnover, which can leave unemployment in equilibrium (Akerlof 1982, Shapiro and Stiglitz 1984).
- Network effects create tipping and path dependence. Two-sided platforms manage cross-side effects through price structure, often subsidizing one side (M. L. Katz and Shapiro 1985, Rochet and Tirole 2003).
- Auction rules shape bids and revenue. The winner's curse, entry, and collusion matter more in practice than benchmark revenue rankings (P. R. Milgrom and Weber 1982,

Klemperer 2002).
- Institutions, the rules of the game, shape transaction costs and long-run performance, and fairness preferences constrain what markets can sustain (North 1990, Fehr and Schmidt 1999, Kahneman, Knetsch, and Thaler 1986).

Table 17.12 presents the comparison of the chapter's theories.

Table 17.12. Comparison of the chapter's theories.

Theory	Market imperfection addressed	Key mechanism	Signature prediction	Typical business use
Coase theorem	Externalities	Bargaining over rights, transaction costs	Efficiency independent of rights when bargaining is cheap	Emissions trading, dispute resolution
Collective action	Public goods	Free riding, selective incentives	Large groups underprovide shared goods	Consortia, trade associations, open source
Property rights theory	Incomplete contracts	Residual control rights	Owner is the party whose investment matters most	Make-or-buy, joint venture design
Search theory	Costly information	Optimal stopping	Price dispersion, frictional unemployment	Pricing transparency, hiring speed
Efficiency wages	Unobservable effort	Gift exchange, shirking deterrence	Above-market wages, involuntary unemployment	Wage policy, retention
Network effects	Consumption externalities	Positive feedback, expectations	Tipping and lock-in	Standards battles, launch strategy
Two-sided markets	Cross-group externalities	Price structure across sides	Subsidized side and money side	Platform pricing
Auction theory	Private information about value	Bidding strategy under rules	Revenue equivalence, winner's curse	Procurement, advertising, spectrum
New institutional economics	Weak rules and enforcement	Transaction costs, credible commitment	Institutions drive long-run performance	Market entry, governance choice
Social preferences	Self-interest assumption	Inequity aversion, reciprocity	Costly punishment of unfairness	Price communication, wage setting

17.12 Discussion Questions

1. Coase argued that harm is reciprocal. Choose a current externality (for example, noise from a delivery hub or data-centre water use) and analyse it from both parties' perspectives. Which institution (private bargaining, liability, tax, or tradable permits) would minimize total costs, and why?
2. Olson predicted that large groups struggle to act collectively, yet open-source projects with thousands of contributors thrive. Which selective incentives and institutional features explain this, and how would you test your explanation?
3. Design a study that distinguishes property rights theory from transaction cost economics. What setting would make the two theories predict different ownership structures?
4. A comparison platform enters a market for home insurance. Using search theory, predict

how average prices, price dispersion, and product design will change. What would count as evidence of obfuscation?

5. A platform must decide which side to subsidize at launch. What information would you need, and how would the answer change if one side could cheaply multi-home?
6. A retailer is considering dynamic pricing that raises prices when demand spikes. Using dual entitlement, advise management on design and communication, and propose a field experiment to evaluate the policy.

17.13 Further Reading

- Coase (1960): the original argument that harm is reciprocal and that transaction costs, not externalities as such, determine the right institutional response.
- Olson (1965): the classic statement of the free-rider problem in groups, still the starting point for research on associations and lobbying.
- Rochet and Tirole (2006): a concise progress report on two-sided markets that defines the concept precisely and summarizes its pricing implications.
- Klemperer (2002): a practical guide to auction design that explains why entry and collusion dominate textbook concerns.
- North (1991): a short and readable overview of how institutions shape economic performance over history.
- Kahneman, Knetsch, and Thaler (1986): the foundational study of fairness as a constraint on pricing and wage setting.

Part IV

Individuals and Groups

Chapter 18

Culture and Cross-Cultural Theory

A television advertisement that sells life insurance in Chicago by promising "freedom to live your dream" may leave viewers in Seoul unmoved, while an advertisement that shows three generations protected by the same policy may move them. A junior engineer in Stockholm interrupts her vice-president to point out an error in a slide. An equally competent engineer in a high-hierarchy subsidiary waits until after the meeting, or says nothing at all. A manufacturer that has grown successfully through acquisitions at home chooses a joint venture, not an acquisition, when it enters a market whose business customs it does not understand. In each case the economics of the situation are similar. What differs is the set of shared assumptions that people bring to it.

This chapter introduces the theories that social scientists use to describe and explain such differences. We begin with what "culture" means as a scientific construct and with the methodological traps that accompany it. We then develop **Hofstede's cultural dimensions theory**, the most cited framework in international business, and its key derivative, **cultural distance**. We examine **power distance** in depth: both Mulder's power distance reduction theory and the practical problem of voice in hierarchical workplaces. We turn to the psychological level with **individualism-collectivism** and **independent versus interdependent self-construal**, then to two rival country-level frameworks, the **GLOBE** project and **Schwartz's theory of basic values**, and finally to **cultural tightness-looseness**. The chapter closes with the principal critiques of the dimensional approach and with practical guidance on negotiation and multicultural teams (Table 18.1).

Table 18.1. Chapter at a glance: theories of culture and cross-cultural difference.

Theory	Core question	Level of analysis	Key source
Hofstede's cultural dimensions	On which value dimensions do national cultures differ?	Nation	Hofstede (1980, 2001)
Cultural distance	How different is a host country from a home country, and does it matter?	Country dyad / firm	Kogut & Singh (1988)
Power distance theory	How do people respond to unequal power, and why do some accept it?	Individual / group / nation	Mulder (1977), Hofstede (1980)
Self-construal and individualism-collectivism	Is the self defined as separate from or embedded in others?	Individual / nation	Markus & Kitayama (1991), Triandis (1995)
GLOBE	How do cultural practices and values shape leadership?	Nation / organization	House et al. (2004)
Schwartz's basic values	What universal structure organizes human values?	Individual / nation	Schwartz (1992)

Table 18.1. Chapter at a glance: theories of culture and cross-cultural difference. *(continued)*

Theory	Core question	Level of analysis	Key source
Tightness-looseness	How strong are social norms and how tolerated is deviance?	Nation / state	Gelfand et al. (2011)

18.1 What Do We Mean by Culture?

DEFINITION

Definition 18.1. Culture is a system of shared meanings, values, beliefs, norms, and practices, that is learned, transmitted across generations, and that distinguishes the members of one group from those of another. Hofstede, Hofstede, and Minkov (2010) describe it as the *collective programming of the mind*: patterns of thinking, feeling, and acting that are acquired early and are largely taken for granted.

Three distinctions organize the cross-cultural literature and recur throughout this chapter.

First, **values versus practices**. Values are broad preferences for some states of affairs over others ("what *should* be"). Practices are the observable ways things are done ("what *is*"). Hofstede's work concentrates on values. The GLOBE project measures both and finds, provocatively, that they do not always move together (House et al. 2004).

Second, **etic versus emic** approaches. An *etic* approach compares cultures on dimensions assumed to be universal, which permits quantification and ranking. An *emic* approach studies a culture on its own terms, from the inside, using concepts that may not translate. Dimensional theories are etic by design. Their power comes from comparability and their limitations from the same source.

Third, and most important for research design, the **level of analysis**. Most dimensional frameworks are defined at the *national* level: they describe how societies differ on average, not how any given person thinks. Inferring individual attitudes from country scores commits the **ecological fallacy**: the error of assuming that relationships observed among aggregates hold among the individuals who make them up (W. S. Robinson 1950). Hofstede himself repeatedly warned against this use of his scores (Hofstede 2001, 2011). Psychological constructs such as self-construal are defined at the individual level, and the two levels must be linked explicitly, typically with multilevel models.

18.2 Hofstede's Cultural Dimensions Theory

THEORY CARD

Core claim. National cultures differ systematically on a small number of value dimensions, power distance, individualism, masculinity, uncertainty avoidance, long-term orientation, and indulgence, and these differences shape institutions, organizations, and behaviour.
Origins. Cross-cultural psychology and organization studies, Hofstede (1980), based on employee attitude surveys at IBM subsidiaries, expanded in Hofstede (2001) and Hofstede, Hofstede, and Minkov (2010), summarised in Hofstede (2011).
Level of analysis. Nation (country-level averages of individual responses).

Key constructs. Power distance, individualism-collectivism, masculinity-femininity, uncertainty avoidance, long- versus short-term orientation, indulgence-restraint, dimension scores (indices).
Boundary conditions. Scores describe central tendencies of national populations, not individuals. They assume that nations are meaningful cultural units and that matched samples isolate national culture from occupation and organization.
Related theories. Power distance theory, GLOBE, Schwartz values, Self-construal, Institutional theory (Section 7.2).

18.2.1 Origins and intellectual history

Geert Hofstede, a Dutch social psychologist working in IBM's personnel research function, gained access to an unusually large data set: employee attitude questionnaires administered in two rounds, around 1968 and 1972, across IBM's subsidiaries worldwide, amounting to more than 100,000 questionnaires (Hofstede 1980, 2011). Because respondents worked for the same company, in similar jobs, and were surveyed with the same instrument, Hofstede argued that the remaining systematic differences between subsidiaries could be attributed to *national* culture. Factor-analysing country means of the work-goal and attitude items, he identified four dimensions, which he related to basic problems every society must solve: how to handle inequality, how tightly to bind the individual to the group, how to divide emotional roles between the genders, and how to deal with an uncertain future.

The model grew in two steps. A fifth dimension, initially called *Confucian dynamism* and later **long-term orientation**, emerged from the Chinese Value Survey : an instrument deliberately designed by Chinese scholars to avoid a Western bias in item selection (Hofstede and Bond 1988). A sixth dimension, **indulgence versus restraint**, was added in the third edition of *Cultures and Organizations* on the basis of Michael Minkov's analyses of World Values Survey data, which also supplied re-estimated long-term orientation scores for many more countries (Hofstede, Hofstede, and Minkov 2010).

18.2.2 Core logic and propositions

Table 18.2 summarises the six dimensions.

Table 18.2. Hofstede's six dimensions of national culture (summarised from Hofstede, Hofstede, & Minkov, 2010, and Hofstede, 2011).

Dimension	Societal problem addressed	Low pole	High pole	Workplace illustration
Power distance (PDI)	How to handle inequality	Inequality minimised, superiors accessible	Hierarchy accepted as natural and proper	Consultative vs. directive supervision
Individualism (IDV)	Bond between individual and group	Collectivism: loyalty to in-groups for protection	Individualism: loose ties, 'I' identity	Group vs. individual incentives and feedback
Masculinity (MAS)	Distribution of emotional roles	Femininity: caring, quality of life, modesty	Masculinity: achievement, assertiveness, competition	Resolving conflict by compromise vs. by contest
Uncertainty avoidance (UAI)	Tolerance of ambiguity	Ambiguity tolerated, few rules	Ambiguity threatening, rules and structure valued	Detailed contracts and procedures

Table 18.2. Hofstede's six dimensions of national culture (summarised from Hofstede, Hofstede, & Minkov, 2010, and Hofstede, 2011). *(continued)*

Dimension	Societal problem addressed	Low pole	High pole	Workplace illustration
Long-term orientation (LTO)	Focus on future vs. present and past	Respect for tradition, quick results	Perseverance, thrift, adaptation	Patient market building vs. quarterly targets
Indulgence (IVR)	Gratification of desires	Restraint: gratification regulated by norms	Indulgence: enjoying life and having fun	Attitudes toward leisure and positive emotion at work

The dimensions are scored as indices, typically on a scale from roughly 0 to 100 (a few countries exceed 100 on some indices). The scores are meaningful only in *relative* terms: a power distance score of 80 is not "high" in any absolute sense, but it places a country above most others in the sample. Figure 18.1 compares six economically important countries.

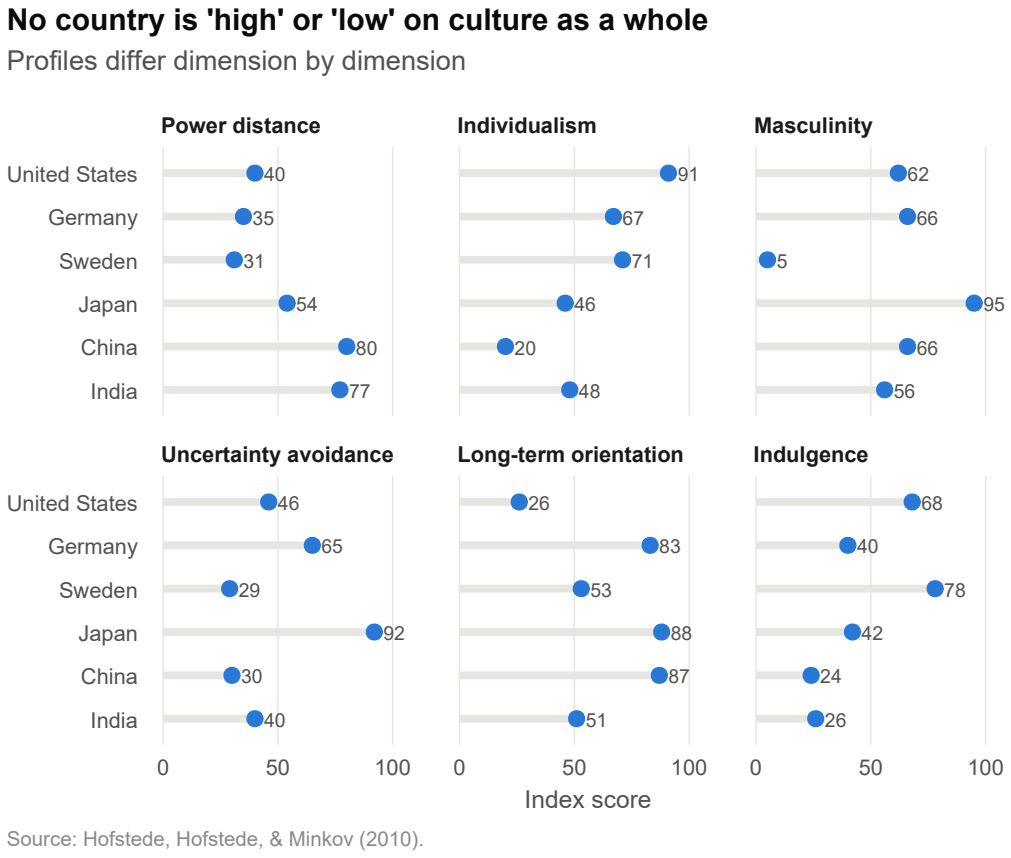


Figure 18.1. Hofstede dimension scores for six countries. Index values as published in Hofstede, Hofstede, and Minkov (2010). Long-term orientation and indulgence scores are the World Values Survey-based estimates introduced in that edition. Scores are relative country positions, not absolute measures.

Several features of the figure are instructive. Japan and Germany, both industrial exporters, differ sharply on individualism. Sweden and Japan, both wealthy, sit at opposite ends of masculinity. China and India share high power distance but differ on long-term orientation. National culture is

a *profile*, not a single position on a "Western-Eastern" continuum.

Hofstede's theoretical claims can be stated as propositions:

- **P1 (Dimensionality).** Differences among national cultures can be described parsimoniously by a small number of independent value dimensions.
- **P2 (Stability).** Countries' *relative* positions on these dimensions change slowly, because values are acquired in childhood and reproduced by institutions. Absolute levels may drift as societies modernise (Hofstede 2001).
- **P3 (Institutional correlates).** Dimension scores covary with national institutions: for example, power distance with centralisation of authority, individualism with national wealth, uncertainty avoidance with the density of rules and laws (Hofstede 2001, Hofstede, Hofstede, and Minkov 2010).
- **P4 (Fit).** Management practices are more effective when they fit the cultural profile of the setting in which they are used. Imported practices that clash with local values are resisted or reinterpreted.

18.2.3 Evidence and extensions

Hofstede's framework has been used in thousands of studies across management, marketing, accounting, and psychology. Two systematic reviews summarise what this literature has learned. Kirkman, Lowe, and Gibson (2006) reviewed a quarter-century of empirical work that used the framework. They classified studies into those treating culture as a main effect (culture predicts an outcome) and those treating it as a moderator (culture changes the strength of another relationship), and they documented recurring problems: heavy concentration on individualism-collectivism, frequent use of country as a proxy for culture without measuring values, and inattention to within-country variation.

Taras, Kirkman, and Steel (2010) meta-analysed the relationships between Hofstede's dimensions, measured at the individual and group levels, and a wide range of organizational outcomes. Their analysis found that cultural values do predict outcomes, but that effects were, overall, modest, and were stronger for attitudes and emotion-laden outcomes (such as organizational commitment (Section 15.10) and team-related attitudes) than for job performance, absenteeism, or turnover. The predictive power of culture also varied across samples, which reinforces the point that culture is one of several influences on behaviour, not a master variable.

DEBATE AND EVIDENCE

A triumph of faith? The sharpest critique is McSweeney (2002), who argues that Hofstede's identification of *national* culture rests on questionable assumptions: that IBM employees were representative of their nations. That the occupational and organizational cultures of IBM were uniform across countries, so that national differences could be "isolated" by subtraction. That a survey of attitudes can reveal deep and enduring values, and that nations are the right unit of cultural analysis. Hofstede replied that the IBM matched-sample design was a strength, not a weakness, representativeness matters less than comparability, and that the dimensions have been validated by many later replications with other populations (Hofstede 2001, 2011). The debate is not resolved, but it has changed practice: careful researchers now measure values directly, test for within-country heterogeneity, and treat country scores as one imperfect proxy among several (Kirkman, Lowe, and Gibson 2006, Tung and Verbeke

2010).

18.2.4 Cultural distance

The most influential derivative of Hofstede's work in strategy is the construct of **cultural distance**: how different the culture of a foreign market is from that of a firm's home country. Kogut and Singh (1988) proposed an index that combines the four original dimensions, standardising each by its variance across countries:

$$CD_j = \frac{1}{4}\sum_{i=1}^{4}\frac{(I_{ij} - I_{iu})^2}{V_i},$$

where I_{ij} is country j's score on dimension i, I_{iu} is the home country's score (the United States in the original study), and V_i is the variance of dimension i. Kogut and Singh showed that, for entries into the United States, greater cultural distance, and higher uncertainty avoidance in the investor's home country, made firms more likely to choose joint ventures or greenfield investments over acquisitions. The intuition is that acquiring an existing firm means managing its culture, which is costly when that culture is unfamiliar. The idea complements the Uppsala internationalization model, in which firms enter "psychically close" markets first and deepen commitment as they learn (Johanson and Vahlne 1977).

Table 18.3 applies the formula to the six countries in Figure 18.1, taking the United States as the home country.

Table 18.3. Illustrative Kogut-Singh cultural distance from the United States, computed from the four original Hofstede indices. Variances are taken across these six countries only. Published studies use the variance across the full country sample, so the index values here are for illustration.

Country	PDI gap	IDV gap	MAS gap	UAI gap	Kogut-Singh index
Germany	-5	-24	4	19	0.41
Sweden	-9	-20	-57	-17	1.27
India	37	-43	-6	-6	1.54
Japan	14	-45	33	46	2.16
China	40	-71	4	-16	3.07

DEBATE AND EVIDENCE

Is distance the right metaphor? Shenkar (2001) identified several "illusions" in the cultural-distance construct: *symmetry* (the distance from A to B is assumed equal to that from B to A, although a Swedish firm entering China may face different frictions than a Chinese firm entering Sweden). *Stability* (distance is treated as fixed, although contact and convergence change it). *Linearity* (each additional unit of distance is assumed to add the same difficulty). *Causality* (culture is assumed to be the only source of difference that matters), and *discordance* (all differences are assumed to create friction, whereas some differences can be complementary). Shenkar called for measures of *cultural friction* that reflect the actual interaction between the two parties.

18.2.5 Applications in business

BUSINESS CASE

Entry mode and the Walmart Germany experience. Walmart entered Germany in the late 1990s by acquiring two domestic hypermarket chains, Wertkauf and Interspar, and withdrew in 2006, selling its stores to Metro. Commentators frequently cite cultural mismatch, for example, the transfer of American service rituals such as store greeters and scripted friendliness, as a contributing factor. A careful reading of the public record suggests a multi-causal story: Walmart also faced a mature, price-competitive discount sector, regulatory constraints on opening hours and pricing, a fragmented store base inherited through acquisition, and difficulties in labour relations. The case is therefore best used to illustrate Kogut and Singh's logic, acquisition requires managing an inherited organizational and national culture, rather than as proof that culture alone determines success.

BUSINESS CASE

Illustrative scenario: choosing an entry mode. A mid-sized German industrial equipment maker wants to enter both the United States and Japan. On the four original dimensions, Japan is more distant from Germany than is the United States on masculinity and uncertainty avoidance, and both differ from Germany on individualism. Applying the cultural-distance logic, the firm might acquire a distributor in the United States, where it can integrate an existing organization with modest cultural friction, but enter Japan through a joint venture with a local partner that already holds relationships with customers and regulators. The theory supplies a hypothesis, not a decision rule: the choice also depends on the availability of acquisition targets, intellectual-property risk, and the firm's prior international experience.

18.2.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The positive effect of individual performance-based pay on employee effort will be weaker in subsidiaries located in more collectivistic countries.
- *H2.* Cultural distance between acquirer and target countries will be negatively associated with post-acquisition integration speed, and this effect will be weaker when the acquirer has prior experience in the target country.
- *H3.* Consumers in high uncertainty-avoidance countries will adopt new product categories more slowly, controlling for income and infrastructure.

Measures. Country scores from Hofstede, Hofstede, and Minkov (2010) (cite the edition and note which countries are estimated). Hofstede's Values Survey Module administered to one's own sample. Kogut-Singh or Euclidean distance indices. Individual-level value scales when the hypothesis is about individuals.

Designs. Multilevel models with individuals nested in countries. Matched samples across subsidiaries of one multinational. Panel data on entry decisions. Experiments with participants from several countries (with measurement-invariance tests).

Data sources. Firm-level foreign direct investment databases, multinational employee surveys. The World Values Survey and European Social Survey. GLOBE and Schwartz country

scores as robustness checks.

COMMON PITFALL

The ecological fallacy. Country scores describe nations, not people. It is a misuse of the theory to assume that a particular Japanese manager is high in uncertainty avoidance because Japan's index is 92, or to correlate country scores with individual behaviour without modelling the two levels separately (W. S. Robinson 1950, Hofstede 2011). Within-country variation on most values is large relative to between-country differences.

18.3 Power Distance Theory

THEORY CARD

Core claim. People respond actively to unequal power: they try to reduce the distance between themselves and more powerful others and to preserve or increase the distance from less powerful others. Societies differ in how much inequality their members expect and accept.
Origins. Social psychology of power, Mulder (1977) formulated *power distance reduction theory*, Hofstede (1980) adopted "power distance" as a dimension of national culture.
Level of analysis. Individual and group (Mulder), nation (Hofstede).
Key constructs. Power distance, power distance reduction tendency, acceptance of inequality, centralisation, voice, dependence on superiors.
Boundary conditions. The reduction tendency is strongest when the distance to the more powerful is small and when the less powerful have relevant expertise. Large gaps in expertise can make participation *increase* inequality.
Related theories. Hofstede, GLOBE, Social exchange theory (Section 9.1), Resource dependence theory (Section 7.1).

18.3.1 Origins: Mulder's power distance reduction theory

The Dutch social psychologist Mauk Mulder studied power in laboratory groups and organizations, and summarised his program in *The Daily Power Game* (Mulder 1977). Mulder defined power distance as the degree of inequality in power between a less powerful individual and a more powerful other within the same social system. His theory makes three core claims. First, individuals strive to *reduce* the power distance between themselves and more powerful others. Second, this tendency is stronger the *smaller* the distance: people close to the top strive harder to close the gap than people far below, for whom the gap seems unbridgeable. Third, the powerful tend to *maintain or increase* the distance from those below them.

Mulder's research carried a counter-intuitive implication for participative management. If the powerful hold much more relevant expertise, giving the less powerful a formal role in decision-making can expose their weaker position and reinforce, rather than reduce, the existing hierarchy. Participation equalises power only when participants bring comparable knowledge (Mulder 1977).

Hofstede borrowed Mulder's term and moved it to the societal level. In his usage, power distance captures the extent to which the less powerful members of a society's institutions and organiza-

tions expect and accept that power is distributed unequally (Hofstede 2001, Hofstede, Hofstede, and Minkov 2010). The key word is *accept*: inequality exists everywhere, but cultures differ in whether it is seen as legitimate and natural or as something to be minimised and justified.

18.3.2 Power distance at work

Table 18.4 contrasts typical organizational patterns in low and high power-distance settings.

Table 18.4. Power distance and the workplace (adapted from the descriptive contrasts in Hofstede, Hofstede, & Minkov, 2010).

Aspect	Lower power distance	Higher power distance
Meaning of hierarchy	Inequality of roles, established for convenience	Reflects existential inequality between levels
Structure	Decentralised, flatter pyramids	Centralised, tall pyramids
Subordinates expect	To be consulted	To be told what to do
Ideal superior	Resourceful democrat	Benevolent autocrat, 'good father'
Pay differentials	Narrow range between top and bottom	Wide range
Status symbols and privileges	Frowned upon	Normal and popular
Disagreeing with the boss	Accepted as part of the job	Seen as disrespectful, raised indirectly if at all

Power distance has measurable consequences for organizational processes. In cross-national studies, Brockner et al. (2001) found that the beneficial effect of giving employees **voice** in decisions, a classic procedural-justice mechanism, was weaker in higher power-distance cultures. Where employees do not expect to be consulted, the opportunity to speak up matters less for their reactions to decisions. This does not mean that voice is unimportant in such settings. It means that the psychological meaning of voice, and thus its motivational effect, depends on cultural expectations.

18.3.3 Mechanism: why hierarchy silences

Voice is discretionary, upward communication intended to improve the organization (Morrison 2011). Two beliefs suppress it: that speaking up is *risky* (it may damage one's image or relationships) and that it is *futile* (nothing will change). Power distance raises both. Leaders, in turn, can lower them by being visibly open to input (Detert and Burris 2007) and by building **psychological safety**: a shared belief that the team is safe for interpersonal risk-taking (A. Edmondson 1999). Figure 18.2 summarises the mechanism.

18.3.4 Applications in business

BUSINESS CASE

Crew resource management in aviation. From the late 1970s onward, accident investigations repeatedly identified failures of communication and coordination on the flight deck: junior crew members who noticed a problem but did not challenge the captain effectively. The industry's response, **crew resource management** (CRM), trains all crew members in assertive, structured communication, shared situational awareness, and decision-making, and it has evolved through several generations (R. L. Helmreich, Merritt, and Wilhelm 1999). Helm-

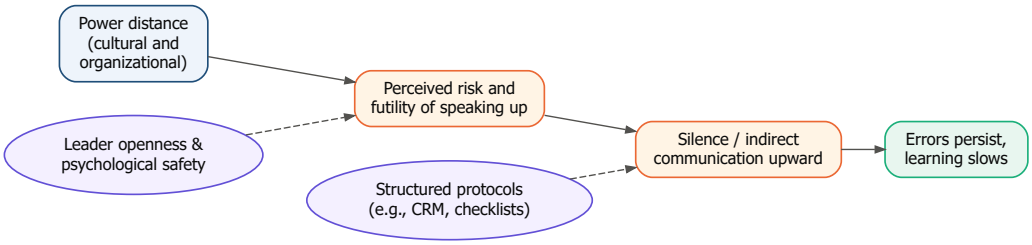


Figure 18.2. Mechanism linking power distance to employee silence and error persistence. Leader openness, psychological safety, and structured protocols such as crew resource management moderate the path from hierarchy to perceived risk and futility of speaking up.

reich and colleagues documented systematic national differences in pilots' attitudes toward command authority and argued that CRM training must be adapted to national, organizational, and professional cultures rather than exported unchanged (R. L. Helmreich and Merritt 1998). Practical tools include standardised phrasing for escalating concerns and explicit invitations by captains for crew to speak up: design solutions that lower the risk of voice regardless of cultural background.

DEBATE AND EVIDENCE

A contested anecdote. Popular accounts, notably Gladwell (2008), link the accident record of a particular national airline in the 1990s primarily to cultural deference in the cockpit. Aviation-safety specialists have criticised such single-cause narratives: official accident investigations typically identify multiple contributing factors (training, procedures, equipment, fatigue, oversight), and attributing outcomes to national culture alone risks both oversimplification and stereotyping. The defensible lesson is narrower, that hierarchy can suppress voice and that structured communication protocols help, and it applies in any high-hierarchy setting, including hospitals, engineering teams, and financial trading desks.

BUSINESS CASE

Illustrative scenario: a "speak-up" program in a high power-distance subsidiary. A multinational's quality audits reveal that defects in its plant in a high power-distance country are detected late, although line operators often notice them early. Exhortations to "speak up" have little effect. Following the mechanism in Figure 18.2, the plant introduces anonymous defect-reporting kiosks, a daily stand-up in which supervisors must ask each operator a specific question, and recognition for the most valuable early warning. Each tool reduces the perceived risk or futility of voice without asking employees to violate deeply held norms about deference.

18.3.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Individual power-distance orientation will weaken the positive effect of participative leadership on employee job satisfaction.
- *H2.* Leader openness will have a stronger positive effect on employee voice in high power-distance teams than in low power-distance teams (because baseline risk perceptions are higher).
- *H3.* Consistent with Mulder, employees one level below a decision-maker will engage in more power-reduction attempts (e.g., coalition building) than employees several levels below.

Measures. Individual power-distance orientation scales, voice and silence scales (Morrison 2011), psychological safety (Section 14.8) (A. Edmondson 1999), hierarchical steepness (number of levels, pay ratios).

Designs. Multilevel field surveys, vignette experiments manipulating leader openness, archival data on incident reporting in hospitals or airlines.

COMMON PITFALL

Equating high power distance with bad management. Low power distance is not universally superior. Directive leadership can be efficient and legitimate where it is expected, and imposing participative practices without regard to local expectations can create confusion rather than empowerment. The managerial question is how to secure the *functions* of voice, error detection and learning, in ways that fit the local context.

18.4 Individualism, Collectivism, and Self-Construal

THEORY CARD

Core claim. Cultures differ in how they define the self. In individualist cultures the self is construed as an autonomous entity defined by internal attributes (independent self-construal). In collectivist cultures it is construed as embedded in relationships and defined by social roles (interdependent self-construal). These construals shape cognition, emotion, and motivation.

Origins. Cultural psychology, H. R. Markus and Kitayama (1991), cross-cultural psychology, Triandis (1995), Hofstede's individualism dimension (Hofstede 1980).

Level of analysis. Individual (self-construal) and nation (individualism-collectivism).

Key constructs. Independent self-construal, interdependent self-construal, in-group, horizontal versus vertical individualism and collectivism.

Boundary conditions. Both construals exist within every person and can be primed situationally. National averages conceal large individual differences.

Related theories. Hofstede, Regulatory focus (Section 20.1), Social identity theory (Section 22.2), Optimal distinctiveness (Section 22.3).

18.4.1 Core logic

H. R. Markus and Kitayama (1991) argued that much of Western psychology assumed an *independent* self, bounded, unique, and stable across contexts, and that this assumption does not describe how people in many cultures experience themselves. In an **interdependent** self, relationships and roles are not external influences on a pre-existing self. They are constitutive of it. The authors traced the consequences across three domains. *Cognition*: interdependent selves attend more to context and to others' perspectives. *Emotion*: they experience and value more "other-focused" emotions, such as sympathy and shame, relative to "ego-focused" ones such as pride and anger. *Motivation*: they are motivated to fit in, maintain harmony, and meet obligations, rather than primarily to stand out and express inner attributes.

Triandis (1995) developed the cultural counterpart, **individualism-collectivism**, and added a second axis. *Horizontal* patterns emphasise equality, *vertical* patterns accept hierarchy. The result is four types: horizontal individualism (self-reliance without status-seeking, often associated with Scandinavia), vertical individualism (competition and achievement), horizontal collectivism (cooperation among equals), and vertical collectivism (duty and deference within a hierarchical in-group) (Triandis and Gelfand 1998) (Table 18.5).

Table 18.5. Independent versus interdependent self-construal (adapted from Markus & Kitayama, 1991. Last row from Lee, Aaker, & Gardner, 2000).

Feature	Independent self-construal	Interdependent self-construal
Definition of self	Separate from social context	Connected with social context
Structure	Bounded, unitary, stable	Flexible, variable across roles
Important features	Internal, private (abilities, thoughts, feelings)	External, public (status, roles, relationships)
Tasks	Be unique, express self, promote own goals	Belong, fit in, promote others' goals
Role of others	Self-evaluation through social comparison	Self-definition through relationships
Basis of self-esteem	Ability to express self and validate attributes	Ability to adjust, restrain self, maintain harmony
Typical regulatory emphasis	Promotion (gains, aspirations)	Prevention (duties, avoiding losses)

18.4.2 Evidence and extensions

Self-construal has been measured with self-report scales (Singelis 1994) and manipulated experimentally by priming, which allows causal tests that country comparisons cannot provide. One important link concerns motivation. A. Y. Lee, Aaker, and Gardner (2000) showed that people with a predominantly independent self-construal place more weight on promotion-focused information (gains and aspirations), whereas those with a predominantly interdependent self-construal place more weight on prevention-focused information (losses and obligations). The pattern appeared both in comparisons of North American and East Asian participants and when self-construal was primed within a culture. This connects cultural psychology to regulatory focus theory (Section 20.1), and it explains why the persuasiveness of promotion- versus prevention-framed messages varies across markets (J. L. Aaker and Lee 2001).

DEBATE AND EVIDENCE

How different are individualists and collectivists? Oyserman, Coon, and Kemmelmeier (2002) meta-analysed cross-national comparisons of individualism and collectivism and found that differences were real but smaller and less consistent than often assumed: Americans were not uniformly more individualistic than all other groups, and differences depended heavily on how the constructs were measured. Heine et al. (2002) identified a further methodological trap, the **reference-group effect**: respondents answer Likert items by comparing themselves with others in their own culture, so a person surrounded by collectivists may rate themselves as relatively individualistic. Finally, Henrich, Heine, and Norenzayan (2010) showed that much behavioural science rests on samples from Western, educated, industrialised, rich, and democratic (WEIRD) societies, which are often outliers rather than typical of humanity. Together these critiques argue for multiple methods, scenario studies, behavioural measures, and priming experiments, rather than cross-national self-report comparisons alone.

18.4.3 Applications in business

BUSINESS CASE

Global advertising: standardise or adapt? T. Levitt (1983) famously argued that technology and travel were homogenising consumer preferences, so that global firms should sell standardised products with standardised messages. Research on culture and persuasion suggests a more nuanced conclusion. S.-P. Han and Shavitt (1994) found that advertisements in the United States more often used individualistic appeals (personal benefit, self-improvement) and those in Korea more often used collectivistic appeals (family, group harmony), and that experimentally, culturally matched appeals tended to be more persuasive, particularly for products typically used with others. Combined with the regulatory-focus evidence (A. Y. Lee, Aaker, and Gardner 2000, J. L. Aaker and Lee 2001), the practical rule is *standardise the brand's core promise and visual identity. Adapt the appeal and framing*. A financial-services brand might emphasise "achieve your ambitions" (promotion, independent) in one market and "protect those who depend on you" (prevention, interdependent) in another, while keeping the same brand platform.

TRY IT YOURSELF

Collect six print or video advertisements for the same product category, say, cars or insurance, from two countries with different individualism scores. Code each advertisement for (a) individual versus group benefit, (b) promotion versus prevention framing, and (c) direct versus indirect claims. Do the patterns match the predictions of self-construal theory? What alternative explanations (category conventions, regulation, brand positioning) could account for any differences?

18.5 The GLOBE Project

THEORY CARD

Core claim. Societal and organizational cultures can be described by nine dimensions measured separately as *practices* ("as is") and *values* ("should be"). Cultures hold shared implicit theories of leadership that determine which leader behaviours are seen as effective.
Origins. The Global Leadership and Organizational Behavior Effectiveness (GLOBE) research program, led by Robert House. House et al. (2004).
Level of analysis. Society and organization, leadership perceptions.
Key constructs. Cultural practices, cultural values, culturally endorsed implicit leadership theories (CLTs), societal clusters.
Boundary conditions. Data come from middle managers in three industries (financial services, food processing, telecommunications). Generalisation to other populations is an assumption.
Related theories. Hofstede, Power distance theory, Upper echelons theory (Section 6.4).

18.5.1 Core logic

GLOBE was designed partly to address limitations of Hofstede's work. A network of researchers in many countries surveyed middle managers in local organizations in 62 societies (House et al. 2004). Two design choices distinguish the project. First, it **expanded and refined the dimensions** (Table 18.6): for example, it split collectivism into institutional and in-group forms, and split Hofstede's masculinity into gender egalitarianism and assertiveness. Second, it measured **practices and values separately**, asking respondents both how things *are* done in their society and how they *should* be done.

Table 18.6. The nine GLOBE cultural dimensions (summarised from House et al., 2004) and their nearest Hofstede counterparts.

GLOBE dimension	What it captures	Closest Hofstede counterpart
Power distance	Expectation that power is stratified and concentrated at higher levels	Power distance
Uncertainty avoidance	Reliance on norms, rules, and procedures to reduce unpredictability	Uncertainty avoidance
Institutional collectivism	Encouragement of collective distribution of resources and collective action	Individualism (reversed)
In-group collectivism	Pride, loyalty, and cohesiveness in families and organizations	Individualism (reversed)
Gender egalitarianism	Minimisation of gender role differences	Masculinity (partly)
Assertiveness	Assertiveness, confrontation, and aggressiveness in relationships	Masculinity (partly)
Future orientation	Planning, investing in the future, delaying gratification	Long-term orientation (partly)
Performance orientation	Encouragement and reward of performance improvement and excellence	No direct counterpart
Humane orientation	Encouragement and reward of fairness, altruism, generosity, and care	No direct counterpart

On leadership, GLOBE identified six broad leadership dimensions: charismatic/ value-based, team-

oriented, participative, humane-oriented, autonomous, and self-protective. Charismatic/value-based and team-oriented leadership were broadly endorsed as contributing to outstanding leadership across societies, whereas the endorsement of participative, autonomous, and self-protective leadership varied more with culture (House et al. 2004). The central proposition is that leader effectiveness depends on the match between leader behaviour and the *culturally endorsed implicit leadership theory* held by followers.

18.5.2 Evidence and extensions

One of GLOBE's most discussed findings is that, for several dimensions, societies' practice scores and value scores were *negatively* correlated across countries. Societies that reported high uncertainty avoidance or high power distance as current practice tended to say they wanted *less* of it. One interpretation is that values express aspirations formed in reaction to current conditions: a "grass is greener" effect.

DEBATE AND EVIDENCE

What did GLOBE measure? Hofstede (2006) argued that GLOBE's items, by asking respondents to describe their society or to state what it should be, captured researchers' categories rather than respondents' own values, and that the negative practice-value correlations suggest the "should be" items do not measure values in his sense. He reanalysed GLOBE's data and claimed that they could be reduced to dimensions resembling his own. Javidan et al. (2006) replied that GLOBE's theory-driven, multi-dimensional approach was more transparent and that practices and values are conceptually distinct constructs whose divergence is a finding, not an artefact. For researchers, the exchange is a reminder that the choice of framework is a measurement decision with substantive consequences. For robustness, many studies now report results using both.

18.5.3 Applications in business

BUSINESS CASE

Illustrative scenario: a global leadership competency model. A professional services firm uses a single 360-degree feedback instrument worldwide, with "empowers others and invites dissent" as a core competency. GLOBE's findings suggest three adjustments. First, retain the competencies that are broadly endorsed, vision, integrity, and team-building, as global standards. Second, treat participative behaviour as a *contingent* competency whose behavioural indicators are localised (for example, consulting privately before a meeting rather than inviting open dissent during it). Third, interpret ratings relative to local norms, because the same behaviour can be rated differently against different implicit leadership theories.

18.6 Schwartz's Theory of Basic Values

THEORY CARD

Core claim. Human values form a universal circular structure: ten basic values, each defined by its motivational goal, are arranged so that adjacent values are compatible and opposing values conflict. Individuals and cultures differ in their *priorities*, not in the *structure*.
Origins. Social psychology. S. H. Schwartz (1992), based on surveys in 20 countries. Extended to culture-level value orientations in S. H. Schwartz (1999).
Level of analysis. Individual (basic values) and nation (cultural value orientations).
Key constructs. Ten basic values, four higher-order values (openness to change, conservation, self-enhancement, self-transcendence), compatibility and conflict among values.
Boundary conditions. The structure is a *continuum*, value boundaries are conventions. Value priorities predict behaviour most strongly when values are activated and relevant to the choice.
Related theories. Hofstede, GLOBE, Self-construal.

18.6.1 Core logic

S. H. Schwartz (1992) defined values as desirable, trans-situational goals that serve as guiding principles in people's lives. What distinguishes values from one another is the **motivational goal** they express. Ten such goals were identified (Table 18.7). The theory's distinctive claim is structural: pursuing any value has psychological, practical, and social consequences that may conflict or be compatible with pursuing others. Achievement and benevolence, for example, can conflict, since pursuing personal success may obstruct actions that enhance the welfare of others. Conformity and tradition are compatible, since both subordinate the self to social expectations.

Table 18.7. Schwartz's ten basic values (paraphrased from Schwartz, 1992).

Value	Defining motivational goal	Higher-order value
Self-direction	Independent thought and action: choosing, creating, exploring	Openness to change
Stimulation	Excitement, novelty, and challenge in life	Openness to change
Hedonism	Pleasure and sensuous gratification for oneself	Openness to change / self-enhancement
Achievement	Personal success through demonstrating competence by social standards	Self-enhancement
Power	Social status and prestige, control over people and resources	Self-enhancement
Security	Safety, harmony, and stability of society, relationships, and self	Conservation
Conformity	Restraint of actions likely to upset others or violate social norms	Conservation
Tradition	Respect for and acceptance of cultural or religious customs	Conservation
Benevolence	Preserving and enhancing the welfare of close others	Self-transcendence
Universalism	Understanding, tolerance, and protection of all people and nature	Self-transcendence

Figure 18.3 draws the circular structure. Two bipolar axes organise the circle: **openness to change**

versus conservation (independence and novelty versus order and resistance to change) and **self-enhancement versus self-transcendence** (pursuit of one's own interests versus concern for others).

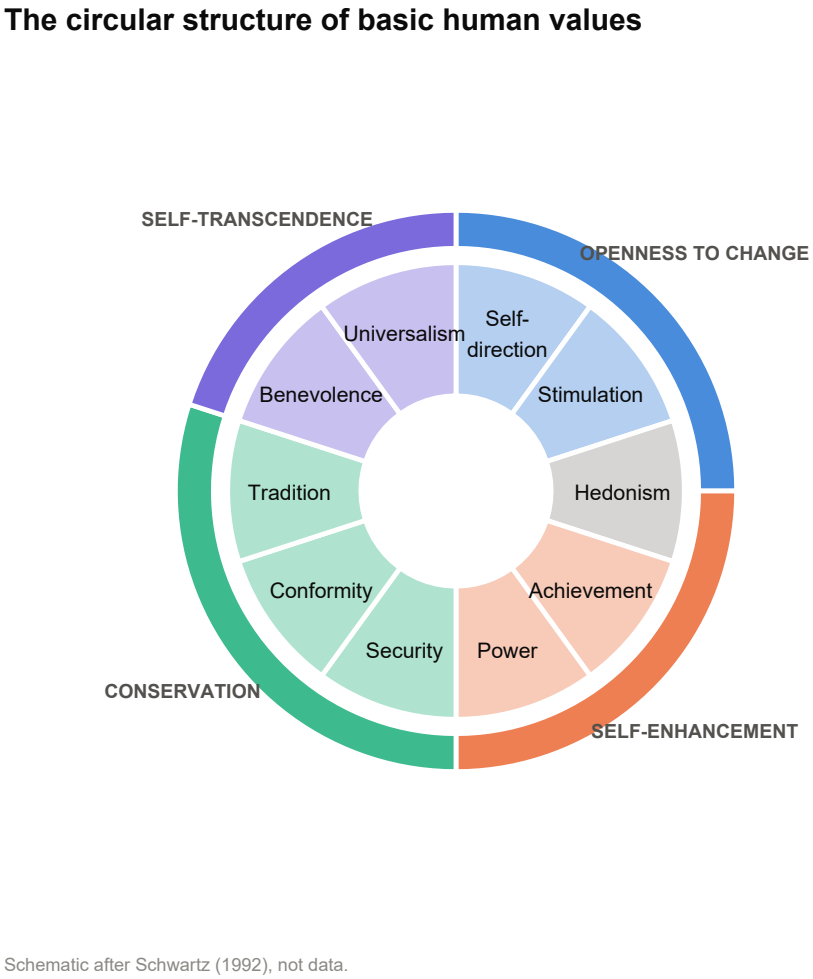


Figure 18.3. Schematic of Schwartz's (1992) circular value structure. Adjacent values are motivationally compatible, values on opposite sides conflict. The outer ring shows the four higher-order values. Hedonism straddles openness to change and self-enhancement. In Schwartz's original figure, tradition and conformity share one wedge (tradition on the outside). They are drawn side by side here for legibility.

At the cultural level, S. H. Schwartz (1999) derived seven value orientations organised along three bipolar dimensions that describe how societies resolve basic problems: *embeddedness versus autonomy* (intellectual and affective), *hierarchy versus egalitarianism*, and *mastery versus harmony*. The cultural-level and individual-level structures are related but not identical: a further reminder that levels of analysis (Section 2.6) must be kept distinct.

18.6.2 Evidence and applications

The circular structure has been replicated across many countries and samples, and Schwartz's instruments are now included in large comparative surveys. Because the theory specifies which values conflict, it yields sharper predictions than lists of independent dimensions: any outcome (for example, willingness to adopt an innovation) should correlate positively with values on one side of

the circle and negatively with values on the opposite side, with a sinusoidal pattern of correlations around the circle.

BUSINESS CASE

Illustrative scenario: positioning and employer branding with the value circle. A sustainable outdoor-apparel brand targets consumers high in universalism and self-direction. The value circle warns that messages invoking power and status, "be the envy of the trail", sit on the opposite side and may undermine credibility with the core segment, whereas adjacent values such as benevolence ("gear that lasts, so less ends up in landfill") and stimulation ("explore further") reinforce one another. The same logic applies to employer branding: a recruitment message that combines conservation values (security, stability) with openness values (autonomy, novelty) sends mixed motivational signals unless the trade-off is made explicit.

RESEARCH TOOLKIT

Using Schwartz's theory in research. Measure values with the Schwartz Value Survey or the Portrait Values Questionnaire. Centre scores within person to capture *priorities*. Test circumplex predictions with the full set of ten values rather than cherry-picking one. At the country level, use Schwartz's cultural orientation scores as an alternative to Hofstede for robustness checks.

18.7 Tight and Loose Cultures

THEORY CARD

Core claim. Cultures differ in the strength of their social norms (Section 22.5) and their tolerance of deviance. **Tight** cultures have strong norms and sanction deviance. **Loose** cultures have weaker norms and greater latitude. Tightness is an adaptive response to ecological and historical threats that require coordination.
Origins. Cultural psychology, Gelfand et al. (2011), a 33-nation study building on earlier anthropological ideas, extended to U.S. states in Harrington and Gelfand (2014), popularised in Gelfand (2018).
Level of analysis. Nation, state, organization, and the everyday situation.
Key constructs. Strength of norms, tolerance of deviance, ecological and historical threat, situational constraint, self-regulation.
Boundary conditions. Tightness is distinct from collectivism and power distance. Tightness can be domain-specific (e.g., tight in workplace norms, loose in leisure).
Related theories. Hofstede, Social norms (Section 22.5), Conformity (Section 22.4), Institutional theory (Section 7.2).

18.7.1 Core logic and mechanism

Gelfand et al. (2011) asked respondents in 33 nations how clear and pervasive social norms were in their society and how much deviance was tolerated, and combined the answers into a tightness-looseness score. Tightness was associated with a history of ecological and human-made threats: higher population density, resource scarcity, territorial threats, natural disasters, and pathogen

prevalence. The theoretical argument is functional: when threats are severe, groups that coordinate effectively survive, and strong norms are a coordination technology. Tightness then becomes embedded in institutions (more autocratic governance, more media restrictions, stricter punishment), in everyday situations (which permit a narrower range of acceptable behaviour), and in individuals' psychology (stronger self-monitoring and impulse control) (Figure 18.4).

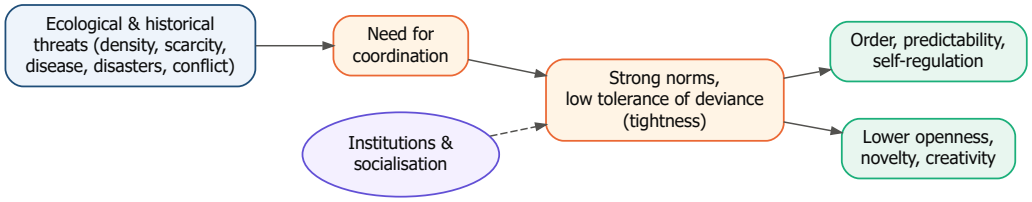


Figure 18.4. Mechanism of cultural tightness-looseness (after Gelfand et al., 2011). Threats increase the need for coordination, which favours strong norms and low tolerance of deviance. Tightness produces order and predictability but constrains openness and creativity.

The trade-off is central. Tight cultures enjoy more order, synchrony, and self-control. Loose cultures enjoy more openness, tolerance, and creativity, at the cost of more disorganisation. Neither pole is superior in general. Gelfand (2018) argues that extremes on either side carry costs. Within the United States, Harrington and Gelfand (2014) found substantial variation among states, with tighter states concentrated in the South and looser states more often on the coasts, and with tightness again associated with historical threat.

18.7.2 Applications in business

BUSINESS CASE

Creativity in global crowdsourcing. Chua, Roth, and Lemoine (2015) studied creative contests on an international crowdsourcing platform, on which clients from many countries post innovation challenges and individuals worldwide submit ideas. They found that cultural tightness shaped success in this global creative work: tight-culture settings were less receptive to creative contributions from culturally distant outsiders. For firms using open innovation (Section 12.6), the implication is that where contributors come from, and where the evaluating client sits, can shape which ideas succeed: independent of idea quality.

BUSINESS CASE

Illustrative scenario: post-merger integration. A tight-culture manufacturer, whose plants run on detailed standard operating procedures and strict punctuality, acquires a loose-culture software start-up whose engineers value flexible hours and improvisation. Integration plans that impose the acquirer's rules wholesale are likely to provoke turnover among the engineers. Plans that leave everything loose may fail to deliver the reliability the acquirer's customers expect. Tightness-looseness theory suggests a *differentiated* design: tight where coordination failures are costly (security, safety, release processes) and loose where variation creates value (exploration, prototyping).

RESEARCH TOOLKIT

Using tightness-looseness in research. Measure tightness with Gelfand et al.'s six-item scale at the individual level and aggregate after testing agreement. Use country or state scores as moderators of norm-related outcomes (compliance, imitation, innovation). Test whether tightness explains variance beyond collectivism and power distance, with which it is correlated but not identical.

18.8 Culture in Practice: Negotiation and Multicultural Teams

Cultural theories earn their place in business education by improving everyday decisions in negotiation, teamwork, and leadership. Three findings stand out.

Negotiation style. Hall (1976) distinguished **low-context** cultures, in which meaning is carried mainly by explicit words, from **high-context** cultures, in which much meaning is carried by relationships, setting, and what is left unsaid. Building on this idea, Adair and Brett (2005) analysed the sequences of behaviour in intra- and inter-cultural negotiations and found that negotiators from low-context cultures tended to exchange information directly, through questions and answers about priorities, whereas negotiators from high-context cultures tended to exchange information indirectly, by making and inferring from offers. Both routes can reach integrative agreements. Problems arise when the two styles meet and each side misreads the other. Brett (2007) synthesises such findings into practical advice for negotiating across cultures.

Multicultural teams. Stahl et al. (2010) meta-analysed research on culturally diverse teams and found offsetting effects: cultural diversity was associated with more creativity but also with more conflict and weaker social integration, so its net effect on performance depended on the task and the team context. Diversity is a resource whose value depends on how it is managed.

Voice and hierarchy. As discussed under power distance theory, members from higher power-distance backgrounds may be less likely to challenge leaders openly. Leaders of global teams can design for this by soliciting input asynchronously, rotating the role of devil's advocate, and explicitly rewarding dissent (Table 18.8).

Table 18.8. From cultural theory to managerial design.

Managerial task	Relevant theory	Key risk	Design response
Choosing an entry mode	Cultural distance	Inherited culture of an acquired firm is hard to manage	Joint venture or greenfield where distance is high. Staged commitment
Global advertising	Self-construal, regulatory focus	Standardised appeal misfires	Standardise brand core, adapt appeal and framing
Cross-cultural negotiation	High/low context	Misreading direct vs. indirect information exchange	Prepare both Q&A and offer-based strategies
Leading global teams	GLOBE, power distance	Silence from members who defer to hierarchy	Asynchronous input, structured dissent
Post-merger integration	Tightness-looseness	Clash between rule-bound and improvisational units	Tight where coordination is critical, loose elsewhere
Global HR practices	Hofstede, Schwartz	Incentives conflict with local values	Localise incentive mix and feedback style

18.9 Critiques and the Research Frontier

Cross-cultural theory has matured through criticism. Five critiques deserve attention from anyone who uses these frameworks.

1. **Nation as proxy for culture.** Countries contain multiple ethnic, regional, religious, and class cultures. Some cultural regions span several countries (McSweeney 2002, Tung and Verbeke 2010).
2. **Within-country variation.** Individual differences within countries are often larger than differences between country means, which limits what a country score can say about any person or firm (Taras, Kirkman, and Steel 2010).
3. **Measurement equivalence.** Survey items may not mean the same thing across languages, and response styles and reference-group effects can distort comparisons (Heine et al. 2002).
4. **Change over time.** National cultures evolve with economic development, migration, and technology. Relative positions may be more stable than absolute scores, but stability is an empirical question (Hofstede 2001).
5. **Sample bias.** Much evidence comes from WEIRD samples, and frameworks built from managerial or student samples may not generalise (Henrich, Heine, and Norenzayan 2010).

Current research responds to these critiques by combining multiple frameworks (Hofstede, GLOBE, Schwartz, tightness-looseness) as robustness checks, by using multilevel designs, by studying subnational culture (for example, U.S. states or Chinese provinces), and by turning from static "distance" to dynamic processes of cultural friction, learning, and adaptation (Shenkar 2001, Tung and Verbeke 2010).

COMMON PITFALL

Culture as residual. A common misuse is to attribute any unexplained cross-country difference to "culture" without measuring cultural variables. Countries also differ in income, institutions, legal systems, and industry structure. A credible cultural explanation measures culture, controls for institutional alternatives, and ideally shows the hypothesised mechanism at the individual level.

18.10 Chapter Summary

KEY TAKEAWAYS

- Culture is a system of shared values, norms, and practices. Dimensional theories describe *national* central tendencies, and applying them to individuals risks the **ecological fallacy** (W. S. Robinson 1950, Hofstede 2011).
- **Hofstede's** six dimensions, power distance, individualism, masculinity, uncertainty avoidance, long-term orientation, and indulgence, remain the most widely used framework, with modest but real predictive validity (Hofstede, Hofstede, and Minkov 2010, Taras, Kirkman, and Steel 2010).
- **Cultural distance** predicts entry-mode choices (Kogut and Singh 1988), but the construct rests on simplifying assumptions of symmetry, stability, and linearity (Shenkar 2001).
- **Power distance** shapes whether people speak up. Mulder's reduction theory and re-

search on voice and psychological safety explain why hierarchy silences and how design can counteract it (Mulder 1977, A. Edmondson 1999, Detert and Burris 2007).
- **Independent and interdependent self-construals** shape cognition, emotion, and motivation, including promotion versus prevention focus (H. R. Markus and Kitayama 1991, A. Y. Lee, Aaker, and Gardner 2000).
- **GLOBE** separates practices from values and links culture to implicit leadership theories (House et al. 2004). **Schwartz** maps values onto a universal circle of compatibility and conflict (S. H. Schwartz 1992).
- **Tight cultures** trade openness for order. Tightness emerges where threats demand coordination (Gelfand et al. 2011).

Table 18.9 presents the comparison of the chapter's cultural frameworks.

Table 18.9. Comparison of the chapter's cultural frameworks.

Framework	Unit of analysis	Core constructs	Signature prediction	Typical business use
Hofstede	Nation	Six value dimensions	Practices that fit national values work better	HR localisation, market segmentation
Cultural distance	Country pair	Standardised dimension differences	Distance favours JVs and greenfield over acquisition	Entry mode, M&A integration
Power distance theory	Individual / nation	Acceptance of inequality, reduction tendency	Voice is weaker and less valued in high-PD settings	Speak-up programs, safety-critical teams
Self-construal	Individual	Independent vs. interdependent self	Matched appeals persuade more	Advertising appeals and framing
GLOBE	Society / organization	Nine dimensions, practices vs. values, CLTs	Leader effectiveness depends on endorsed leadership prototypes	Global leadership development
Schwartz	Individual / nation	Ten values on a circle	Opposing values have opposite correlations	Positioning, employer branding
Tightness-looseness	Nation / state	Norm strength, tolerance of deviance	Tight cultures: order, loose: creativity	Integration design, innovation sourcing

18.11 Discussion Questions

1. Hofstede's dimensions were derived from employees of a single multinational. Is this a strength (matched samples) or a weakness (unrepresentative samples)? Design a study that could adjudicate between McSweeney's and Hofstede's positions.
2. A study finds that country-level individualism correlates with per-capita patent counts. What alternative explanations should be ruled out before concluding that individualism causes innovation? What individual-level evidence would strengthen the causal claim?
3. Using the Kogut-Singh formula, compute the cultural distance between two countries of your choice. Then apply Shenkar's critique: in what ways might the *experienced* friction differ from the index?

4. A U.S. hospital chain is opening a facility staffed largely by clinicians trained in high power-distance settings. Drawing on Mulder, Edmondson, and the CRM literature, propose three practices to protect patient safety.
5. GLOBE finds that practices and values are often negatively correlated across societies. Offer two theoretical explanations and describe the data you would need to distinguish them.
6. When should a global brand standardise its advertising, and when should it adapt? Formulate a testable hypothesis that integrates self-construal and product category.

18.12 Further Reading

- Hofstede (2011): a concise, freely available overview of the six-dimension model by its author, including its history and intended uses.
- Hofstede, Hofstede, and Minkov (2010): the most complete statement of the model, with country scores and extensive discussion of implications for organizations.
- McSweeney (2002): the most forceful critique of Hofstede's methodology. Essential for understanding the framework's limits.
- Taras, Kirkman, and Steel (2010): a meta-analysis that quantifies how well cultural values predict organizational outcomes.
- H. R. Markus and Kitayama (1991): the classic statement of independent and interdependent self-construal and its psychological consequences.
- House et al. (2004): the GLOBE study. Read alongside Hofstede (2006) and Javidan et al. (2006) for the debate it sparked.
- S. H. Schwartz (1992): the foundational paper on the circular structure of values.
- Gelfand et al. (2011): a short, influential article introducing tightness-looseness at the national level.

Chapter 19

Motivation

A coffee shop hands out loyalty cards: buy ten coffees, get one free. Customers who are one stamp away from the reward return noticeably sooner than customers who have just started a new card. A sales organization raises its quarterly quota and sees a burst of effort, and, in the final week of the quarter, a burst of deals that are pulled forward, discounted, or quietly booked in ways that violate policy. A volunteer organization begins paying a small stipend and finds that volunteers work *less* than before. A basketball team that trails by a single point at half-time goes on to win more often than one that leads by a point. Each of these puzzles is about **motivation**: the forces that determine the direction, intensity, and persistence of effort.

This chapter develops the major theories that explain such patterns. After a brief contrast between *content* and *process* theories, we examine **expectancy theory**, which treats motivation as a product of beliefs about effort, performance, and rewards. **Equity theory**, which explains how social comparison of rewards produces fairness judgements and behaviour. **Goal-setting theory**, the most practically influential theory of work motivation, together with the debate about its side effects. The **goal-gradient hypothesis**, which explains why effort accelerates as a goal approaches. **Self-determination theory**, which distinguishes autonomous from controlled motivation and explains when rewards backfire, and the striking finding that **losing can lead to winning**. The theories apply as readily to customers in a loyalty program as to employees under a sales plan, and the chapter draws its examples from both (Table 19.1).

Table 19.1. Chapter at a glance: theories of motivation.

Theory	Core question	Level of analysis	Key source
Need theories (Maslow, Herzberg)	What needs or job factors energise behaviour?	Individual	Maslow (1943), Herzberg et al. (1959)
Expectancy theory	How do beliefs about effort, performance, and rewards combine into motivation?	Individual	Vroom (1964)
Equity theory	How do people judge and respond to the fairness of their rewards?	Individual / dyad	Adams (1963, 1965)
Goal-setting theory	Why do specific, difficult goals raise performance?	Individual / team	Locke & Latham (1990, 2002)
Goal-gradient hypothesis	Why does effort accelerate as a goal approaches?	Individual	Hull (1932), Kivetz et al. (2006)
Self-determination theory	When is motivation autonomous, and when do rewards undermine it?	Individual	Deci & Ryan (1985), Ryan & Deci (2000)

Losing leads to winning	Why can being slightly behind increase success?	Individual / team	Berger & Pope (2011)

19.1 Content and Process Theories

DEFINITION

Definition 19.1. Motivation is the set of psychological processes that direct, energise, and sustain goal-directed behaviour. Researchers typically distinguish its **direction** (which activity a person chooses), **intensity** (how much effort is exerted), and **persistence** (how long effort is sustained) (Latham and Pinder 2005, Kanfer, Frese, and Johnson 2017).

Early theories of work motivation asked *what* motivates people. These **content theories** identify needs or job characteristics that energise behaviour. Maslow (1943) proposed a hierarchy of needs, physiological, safety, love and belonging, esteem, and self-actualisation, in which higher needs become motivating only once lower needs are reasonably satisfied. The hierarchy is intuitively appealing and remains popular in management education, but empirical tests have found little support for a strict ordering of needs or for the claim that satisfying one level activates the next (Wahba and Bridwell 1976).

Herzberg, Mausner, and Snyderman (1959) interviewed engineers and accountants about episodes in which they had felt exceptionally good or bad about their jobs. The two kinds of episode involved different factors. **Hygiene factors**, pay, supervision, working conditions, company policy, were associated with dissatisfaction when they were poor. **Motivators**, achievement, recognition, responsibility, the work itself, advancement, were associated with satisfaction when present. The resulting two-factor theory has been criticised on methodological grounds, notably that people tend to attribute good experiences to themselves and bad experiences to the environment, but its practical message endured: removing sources of dissatisfaction does not by itself create motivation.

Later **process theories** asked *how* motivation arises and changes: how people form beliefs about effort and reward, compare their outcomes with others', set and pursue goals, and internalise the reasons for acting. Most of this chapter concerns process theories, which yield sharper, testable predictions (Table 19.2).

Table 19.2. Content versus process theories of motivation.

Feature	Content theories	Process theories
Central question	What motivates people?	How is motivated behaviour initiated, directed, and sustained?
Core constructs	Needs, motives, job factors	Expectancies, valences, comparisons, goals, regulation
Exemplars	Maslow's hierarchy, Herzberg's two-factor theory	Expectancy, equity, goal-setting, self-determination theories
Typical managerial implication	Identify and satisfy important needs	Design beliefs, comparisons, goals, and incentives
Empirical status	Mixed, strict hierarchies poorly supported	Generally stronger, with well-specified boundary conditions

19.2 Expectancy Theory

THEORY CARD

Core claim. People are motivated to exert effort when they believe that effort will lead to performance (expectancy), that performance will lead to outcomes (instrumentality), and that those outcomes are valued (valence). The three beliefs combine multiplicatively.
Origins. Industrial-organizational psychology, Vroom (1964), extended by L. W. Porter and Lawler (1968).
Level of analysis. Individual (choices among levels of effort or courses of action).
Key constructs. Expectancy (E), instrumentality (I), valence (V), motivational force, effort, performance.
Boundary conditions. Assumes deliberate, calculative choice. Predicts better for choices among alternatives within a person than for differences between people. Less suited to habitual or impulsive behaviour.
Related theories. Goal-setting theory, Self-determination theory, Agency theory (Section 6.1), Prospect theory (Chapter 24).

19.2.1 Origins and core logic

Vroom (1964) brought the logic of subjective expected utility into work psychology. A worker considering how hard to try asks, in effect, three questions. *If I try harder, will I perform better?* (**expectancy**, a subjective probability between 0 and 1). *If I perform better, will I be rewarded?* (**instrumentality**, the perceived link between performance and outcomes). *Do I care about the rewards?* (**valence**, the anticipated satisfaction from outcomes, which can be positive or negative). In its simplest form:

$$\text{Motivational force} \;=\; E \times \sum_{k} \left(I_k \times V_k\right),$$

where the sum runs over the outcomes k (pay, recognition, promotion, effort costs) that performance may produce. The multiplicative form carries the theory's most important prediction: if *any* component is close to zero, motivation is close to zero, no matter how large the others are. A generous bonus (high valence) does not motivate a salesperson who believes the quota is unattainable (low expectancy) or who suspects that bonuses are awarded by favouritism (low instrumentality).

L. W. Porter and Lawler (1968) extended the model by adding that effort translates into performance only when ability and role clarity are present, and that the satisfaction produced by rewards depends on whether they are perceived as equitable: a bridge to equity theory.

19.2.2 Mechanism

Figure 19.1 visualises the multiplicative logic. Motivation is high only in the corner of the space where expectancy and valence are both high, and the whole surface shrinks when instrumentality is weak.

Table 19.3 translates each component into managerial levers and research measures.

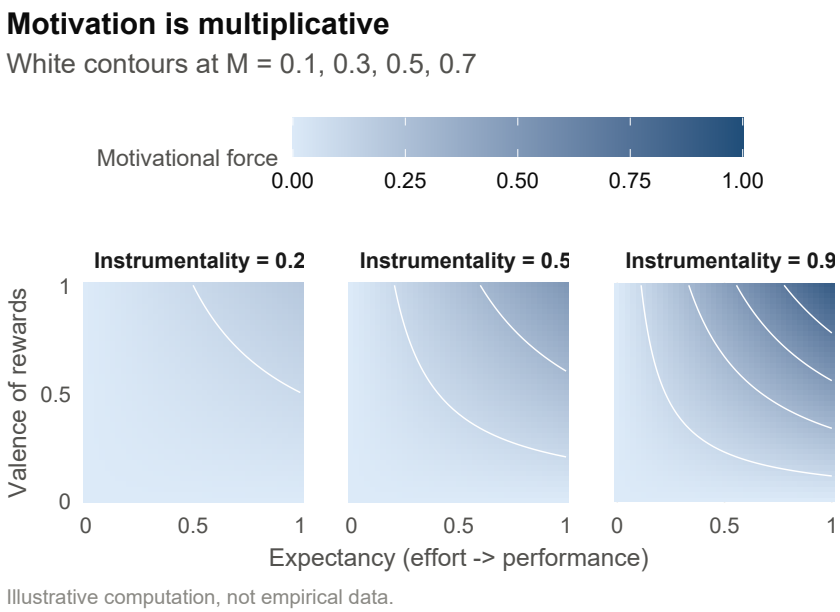


Figure 19.1. Illustrative motivational force under the multiplicative expectancy model (M = E x I x V). Each panel fixes instrumentality at a different level. Motivation is high only when all three components are high. A near-zero value on any component collapses motivation.

Table 19.3. Components of expectancy theory and how to influence them.

Component	Belief in question	What lowers it	Managerial levers	Example measure
Expectancy (E)	Effort -> performance	Unrealistic targets, lack of skill, tools, or authority	Training, resources, attainable goals, self-efficacy building	Subjective probability of reaching a target with high vs. normal effort
Instrumentality (I)	Performance -> outcomes	Opaque or discretionary rewards, caps, favouritism	Transparent pay-for-performance rules, credible commitments	Perceived link between performance and pay or promotion
Valence (V)	Value of outcomes	Rewards that do not match preferences	Cafeteria-style rewards, recognition, meaningful work	Rated attractiveness of each outcome

19.2.3 Evidence and extensions

Van Eerde and Thierry (1996) meta-analysed tests of Vroom's models. Each component, valence, instrumentality, and expectancy, was related to work criteria such as performance, effort, intention, preference, and choice, but the multiplicative combination did not generally predict better than the components entered separately or additively. Effects were also larger in *within-subject* designs, which compare a person's choices across alternatives, than in *between-subject* designs, which compare people. The second finding fits Vroom's original intent: the theory models how an individual chooses among options, not why one person is more motivated than another.

The expectancy component has a close cousin in Bandura's concept of **self-efficacy (Section 21.5)**:

the belief that one can execute the behaviours required to produce an outcome (Bandura 1977). Self-efficacy has become one of the most robust predictors of effort and persistence, and it is a central moderator in goal-setting theory. Steel and König (2006) integrated expectancy and valence with time discounting in *temporal motivation theory*, which explains procrastination as the result of rewards that are valued but distant.

19.2.4 Applications in business

BUSINESS CASE

Sales quotas and the shape of incentive plans. Sales compensation is the purest field laboratory for expectancy theory, because the links between effort, performance (bookings), and pay are explicit. Research on nonlinear plans, quotas with lump-sum bonuses, shows that salespeople respond strategically to where they stand relative to the threshold. Oyer (1998) documented that sales across many industries are higher at the end of fiscal years, consistent with salespeople timing effort and deals to meet annual targets. Larkin (2014) found that enterprise-software salespeople facing accelerating commission schedules timed and priced deals to maximise their own pay, at a cost to the firm. Chung, Steenburgh, and Sudhir (2014) estimated a dynamic model of salespeople's responses to a bonus-based plan and concluded that bonuses enhanced sales productivity, with periodic bonuses helping to keep weaker performers on pace and overachievement commissions sustaining the effort of stronger performers who had already reached their quotas. In expectancy terms, a quota that is far out of reach destroys expectancy. A quota already met destroys instrumentality for the rest of the period. Both lower effort.

RESEARCH TOOLKIT

Using expectancy theory in research.

- *H1.* Raising a sales quota will increase effort among salespeople whose pre-period performance is close to the new quota and decrease effort among those far below it (an inverted-U in expectancy).
- *H2.* The effect of bonus size on effort will be stronger when the performance-pay link is transparent (high instrumentality).

Designs. Within-person choice experiments (conjoint designs over effort-reward options). Field data on nonlinear incentive plans exploiting distance to threshold. Structural models of dynamic effort. **Measures.** Subjective probabilities and ratings for each component. Self-efficacy scales.

COMMON PITFALL

Testing a within-person theory between persons. Correlating one person's average expectancy score with another person's performance tests a different question from the one Vroom posed. Where possible, measure E, I, and V for several alternatives per person and model choices within individuals (Van Eerde and Thierry 1996).

19.3 Equity Theory

THEORY CARD

Core claim. People judge the fairness of their outcomes by comparing their ratio of outcomes to inputs with that of a referent other. Unequal ratios create psychological tension, which people reduce by changing inputs, outcomes, perceptions, comparison others, or the relationship itself.
Origins. Social psychology of exchange, J. S. Adams (1963), J. S. Adams (1965), rooted in social exchange theory (Section 9.1) and social comparison theory (Section 22.1).
Level of analysis. Individual and dyad (employee-employer, employee-coworker).
Key constructs. Inputs, outcomes, referent other, under-reward and over-reward inequity, distributive justice, equity sensitivity.
Boundary conditions. Requires a salient comparison other and comparable inputs and outcomes. Reactions to over-reward are weaker than to under-reward. Individuals differ in equity sensitivity.
Related theories. Social exchange theory (Section 9.1), Social comparison (Section 22.1), Expectancy theory.

19.3.1 Core logic

J. S. Adams (1963) proposed that people evaluate an exchange relationship by comparing two ratios:

$$\frac{O_p}{I_p} \quad \text{versus} \quad \frac{O_o}{I_o},$$

where O denotes perceived outcomes (pay, recognition, status, interesting assignments), I denotes perceived inputs (effort, skill, education, seniority, time), p is the person, and o is the comparison other. **Equity** exists when the ratios are equal, not when outcomes are equal. A senior engineer who earns more than a junior colleague perceives no inequity if the difference is proportional to differences in inputs.

When the ratios differ, the person experiences **inequity**: *under-reward* inequity (one's own ratio is lower) produces anger and resentment, whereas *over-reward* inequity (one's own ratio is higher) produces guilt, although the threshold for noticing over-reward is typically higher. Because the tension is unpleasant, people act to restore equity. J. S. Adams (1965) catalogued the available responses (Table 19.4).

Table 19.4. Responses to inequity (after Adams, 1965).

Response	Under-reward example	Over-reward example	Organizational consequence
Alter own inputs	Reduce effort, withdraw extra-role behaviour	Work harder or longer	Productivity and citizenship behaviour change
Alter own outcomes	Request a raise, take resources (including theft)	Decline a perk (rare)	Pay negotiations, deviance
Cognitively distort own inputs or outcomes	'My job is less demanding than I thought'	'I really do contribute more'	Attitudes adjust without behaviour change
Act on or distort the other	Pressure the other to work harder, belittle their inputs	Encourage the other to do less	Team conflict, norm enforcement

Table 19.4. Responses to inequity (after Adams, 1965). *(continued)*

Response	Under-reward example	Over-reward example	Organizational consequence
Change the comparison other	Compare with a worse-off group instead	Compare with a better-paid group	Shifts in reference groups
Leave the field	Absenteeism, quitting, transfer	Rare	Turnover

19.3.2 Mechanism

Figure 19.2 shows the causal chain. Two moderators are especially important. First, **equity sensitivity**: some people ("benevolents") tolerate under-reward, others ("entitleds") prefer over-reward, and others ("equity sensitives") follow Adams' predictions closely (Huseman, Hatfield, and Miles 1987). Second, **interactional justice**: honest, respectful explanations of how outcomes were decided reduce the intensity of reactions to the same distributive outcome (Greenberg 1990, Colquitt et al. 2001).

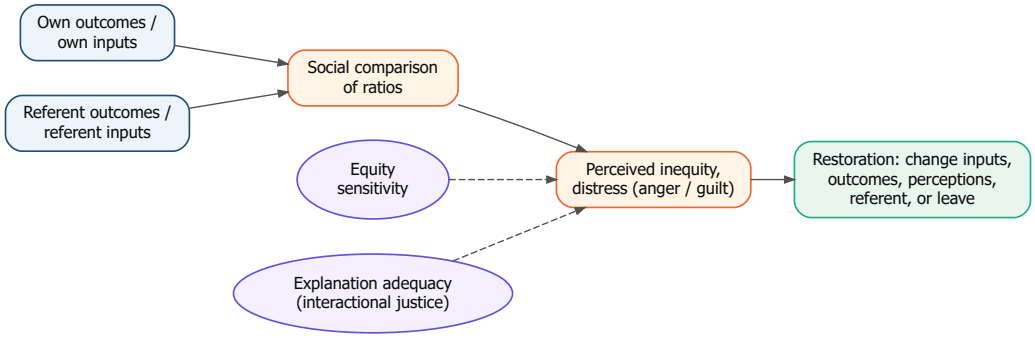


Figure 19.2. Mechanism of equity theory. Comparing one's own outcome/input ratio with a referent's creates perceived inequity, which produces distress (anger or guilt) and motivates restoration. Equity sensitivity and the adequacy of explanations moderate the process.

19.3.3 Evidence and extensions

The most cited field demonstration is Greenberg (1990). In a manufacturing company, two plants experienced a temporary pay cut, while a third plant did not. Employee theft rose in the plants with the pay cut and fell back when normal pay resumed. Crucially, theft rose much less in the plant where management explained the pay cut thoroughly and with expressions of regret than in the plant that received a brief, impersonal explanation. Theft, in equity terms, was a way of restoring outcomes. The explanation reduced the perceived unfairness of the same pay cut.

Equity theory evolved into the broader field of **organizational justice (Section 15.9)**. In addition to **distributive justice** (fairness of outcomes, the domain of equity theory), researchers identified **procedural justice**, fairness of the process, including the opportunity for voice (J. Thibaut and Walker 1975), and **interactional justice**, which Colquitt et al. (2001) showed is best split into interpersonal and informational components. Their meta-analysis supported four distinct but correlated forms of justice with partly different consequences (Table 19.5).

Table 19.5. Four dimensions of organizational justice (following Colquitt et al., 2001).

Justice dimension	Fairness of ...	Typical criteria	Managerial lever
Distributive	Outcomes	Equity (proportional to inputs), equality, need	Pay structure, bonus allocation
Procedural	Decision processes	Voice, consistency, bias suppression, accuracy, correctability	Transparent criteria, appeal mechanisms
Interpersonal	Treatment by decision-makers	Dignity, respect, propriety	Manager training, respectful delivery
Informational	Explanations	Truthfulness, adequacy, timeliness of explanations	Explaining pay decisions and cuts

19.3.4 Applications in business

BUSINESS CASE

Pay transparency and fairness. Pay transparency laws and policies make referent outcomes visible. Card et al. (2012) exploited a natural experiment: after a newspaper launched a searchable website of University of California employees' salaries, researchers informed a random subset of employees about the site. Employees whose pay was below the median for their unit and occupation reported lower job satisfaction and greater intention to look for a new job. Those above the median showed no corresponding increase in satisfaction. The asymmetry is exactly what equity theory predicts: under-reward hurts more than over-reward helps. For firms adopting pay transparency, the lesson is that the *inputs* side of the ratio must also be visible and credible, published pay bands, clear criteria for progression, and explanations for differences, or transparency will mainly reveal perceived inequities.

RESEARCH TOOLKIT

Using equity theory in research.

- *H1.* Disclosure of peer salaries will reduce satisfaction among below-median employees more than it increases satisfaction among above-median employees.
- *H2.* The negative effect of a pay cut on discretionary effort will be weaker when the cut is accompanied by an adequate explanation.

Measures. Colquitt's justice scales, equity sensitivity instrument (Huseman, Hatfield, and Miles 1987), behavioural outcomes such as turnover, absenteeism, and shrinkage. **Designs.** Natural experiments in disclosure, field experiments on explanations, laboratory games with manipulated referent outcomes.

COMMON PITFALL

Assuming the referent. Researchers and managers often assume that employees compare themselves with the obvious peer group. Equity theory makes the choice of referent a variable, and people may compare with former colleagues, friends in other firms, or their own past outcomes. Measure the referent or manipulate it, do not assume it.

19.4 Goal-Setting Theory

THEORY CARD

Core claim. Specific, difficult goals lead to higher performance than easy goals, vague goals, or exhortations to "do your best", provided that people are committed to the goal, have the ability to reach it, and receive feedback.
Origins. Industrial-organizational psychology, Locke and Latham (1990), summarised in Locke and Latham (2002).
Level of analysis. Individual, team, and organization.
Key constructs. Goal specificity, goal difficulty, goal commitment, feedback, task complexity, self-efficacy, learning versus performance goals.
Boundary conditions. Effects weaken for complex tasks where people lack strategies, when commitment is low, and when goals conflict. Poorly designed goals produce side effects.
Related theories. Expectancy theory, Goal gradient, Prospect theory (Chapter 24), Self-regulation (Chapter 20).

19.4.1 Core logic and propositions

Goal-setting theory was built inductively from hundreds of laboratory and field studies. Locke and Latham (2002) summarise its core propositions:

- **P1 (Difficulty).** Performance increases with goal difficulty, up to the limits of ability, provided the person remains committed.
- **P2 (Specificity).** Specific, difficult goals produce higher performance than vague "do your best" goals, because vague goals allow a wide range of acceptable performance.
- **P3 (Mechanisms).** Goals affect performance through four mechanisms: they *direct* attention to goal-relevant activities, *energise* effort, prolong *persistence*, and prompt the discovery and use of task *strategies*.
- **P4 (Moderators).** Goal effects are stronger when commitment is high (goals are important and self-efficacy is high), when feedback shows progress, and when tasks are simple. For complex tasks, learning goals can outperform performance goals.

Figure 19.3 sketches the basic relationship.

Goals also act as **reference points**. Heath, Larrick, and Wu (1999) showed that goals divide outcomes into gains and losses: falling short of a goal feels like a loss, and, consistent with prospect theory (Chapter 24) (Kahneman and Tversky 1979), the marginal value of effort is greatest near the goal. This reference-point view explains why goals motivate even without external rewards and foreshadows both the goal-gradient and the losing-leads-to-winning phenomena discussed below. Field data confirm reference-point effects in performance: professional golfers putt more accurately for par than for an otherwise comparable birdie (Pope and Schweitzer 2011), and marathon finishing times bunch just ahead of round-number targets such as four hours (E. J. Allen et al. 2017) (Table 19.6).

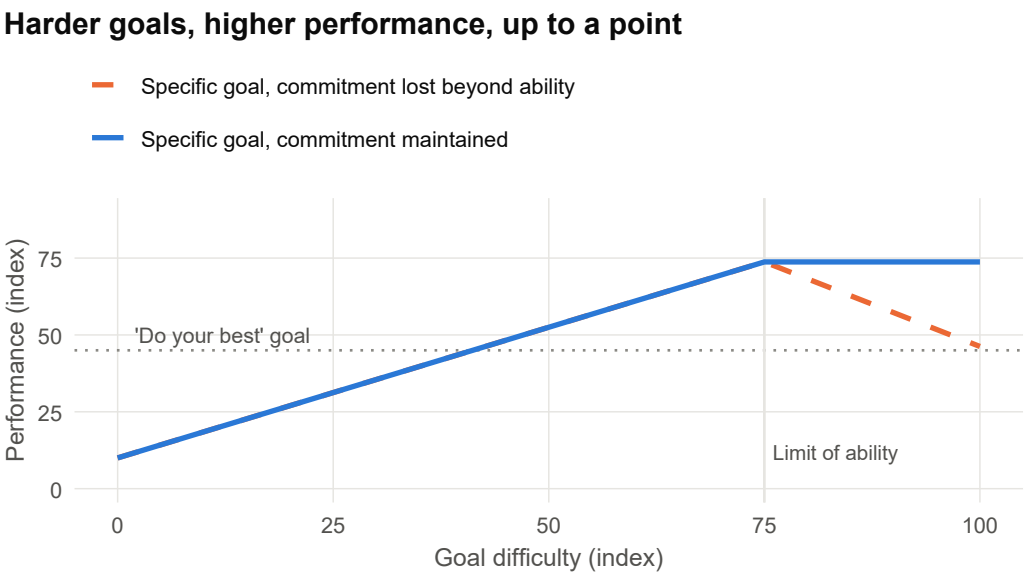


Figure 19.3. Stylised goal-difficulty function. Performance rises with goal difficulty up to the limit of ability. Beyond that point, performance plateaus when commitment holds and falls when people abandon an impossible goal. A vague 'do your best' goal produces moderate performance.

Table 19.6. Moderators of goal-setting effects (summarised from Locke & Latham, 2002).

Moderator	How it changes the goal effect	Design implication
Goal commitment	Effects require commitment, increased by importance and self-efficacy	Explain why the goal matters, build confidence, consider participation
Feedback	Allows people to track progress and adjust effort or strategy	Provide frequent, specific progress information
Task complexity	Effects smaller on complex tasks requiring new strategies	Use learning goals (e.g., 'identify five new approaches') on novel tasks
Ability	Performance plateaus at the limit of ability	Calibrate difficulty to capability, train
Goal conflict	Multiple goals compete for attention	Prioritise, avoid incompatible targets

19.4.2 Evidence and the "goals gone wild" debate

The evidence base for goal-setting theory is among the largest in organizational psychology, and its effects have been replicated across tasks, countries, and levels from individuals to organizations (Locke and Latham 1990, 2002). The theory's very success, however, provoked a debate about its side effects.

DEBATE AND EVIDENCE

Goals gone wild? Ordóñez et al. (2009) argued that goal setting had been "overprescribed" and that specific, challenging goals have systematic side effects: they narrow attention so that unrewarded dimensions of performance (quality, safety, learning) are neglected. They encourage riskier behaviour, they can crowd out intrinsic motivation, and, most seriously, they motivate unethical behaviour. They drew on organizational cases, including Sears' auto-

repair centres and the Ford Pinto, and on experimental evidence. In one such experiment, Schweitzer, Ordóñez, and Douma (2004) found that people with unmet specific goals were more likely to overstate their performance than people told to do their best, and that the tendency was strongest when they fell just short of the goal. Locke and Latham (2009) replied that the critics relied heavily on anecdotes, that the side effects are well known and are addressed within goal-setting theory, and that the remedy is better goal setting, appropriate difficulty, multiple goals that include quality and ethics, and monitoring, not abandoning goals. The productive synthesis is that goals are powerful and therefore need the same governance as any powerful incentive.

19.4.3 Applications in business

BUSINESS CASE

Wells Fargo's cross-selling targets. Wells Fargo's retail bank long promoted an aggressive cross-selling strategy, summarised internally by the slogan "Eight is great", a goal of eight products per household. In September 2016, the Consumer Financial Protection Bureau, the Office of the Comptroller of the Currency, and the Los Angeles City Attorney announced settlements totalling $185 million over the opening of accounts that customers had not authorised. The regulators' analysis found that employees may have opened roughly 1.5 million deposit accounts and submitted roughly 565,000 credit-card applications without customers' consent. The bank had dismissed about 5,300 employees over several years for related misconduct. Subsequent reviews identified a far larger number of potentially unauthorised accounts, and in 2020 the bank agreed to pay $3 billion to resolve criminal and civil investigations by the U.S. Department of Justice and the Securities and Exchange Commission. A report commissioned by the bank's independent directors in 2017 attributed the misconduct to a sales culture and performance-management system built around high, specific product targets. In the vocabulary of this chapter, the case combines the conditions that Ordóñez et al. (2009) and Schweitzer, Ordóñez, and Douma (2004) identified: difficult, specific, and highly monitored goals. Narrow measurement (accounts opened rather than accounts used). Pressure that intensified as employees approached their targets, and weak controls against gaming.

RESEARCH TOOLKIT

Using goal-setting theory in research.

- *H1.* Specific, difficult goals will increase output quantity but reduce quality when quality is not included in the goal.
- *H2.* The probability of misreporting will be highest for employees who end a period just below their goal.
- *H3.* Learning goals will outperform performance goals on novel, complex tasks.

Designs. Randomised goal assignments in the lab or field. Regression discontinuity around goal thresholds. Bunching estimators for outcomes near round-number goals (E. J. Allen et al. 2017). **Measures.** Goal difficulty and specificity, commitment scales, objective performance, audit data on misreporting.

COMMON PITFALL

Measuring what is easy rather than what matters. Goals direct attention to whatever is measured. A goal defined on a proxy (accounts opened, calls handled, tickets closed) will be pursued at the expense of the underlying objective (customer value, problem resolution). Pair quantity goals with quality and integrity metrics, and audit the behaviours that the goal makes tempting.

19.5 The Goal-Gradient Hypothesis

THEORY CARD

Core claim. The tendency to approach a goal increases with proximity to the goal: effort accelerates as the reward draws near.
Origins. Behaviourist learning theory. Hull (1932) and Hull (1934) on rats running toward food. J. S. Brown (1948) on gradients of approach and avoidance. Revived in consumer research by Kivetz, Urminsky, and Zheng (2006).
Level of analysis. Individual (animals, consumers, employees).
Key constructs. Goal distance, perceived progress, effort acceleration, illusionary (endowed) progress, post-reward resetting.
Boundary conditions. Effects depend on *perceived* rather than actual progress. Motivation can sag in the middle of a goal. In competition, being far behind can trigger quitting.
Related theories. Goal-setting theory, Losing leads to winning, Prospect theory (Chapter 24).

19.5.1 Origins and core logic

Hull (1932) formulated the goal-gradient hypothesis to explain maze learning: responses closer to the reward are reinforced more strongly, so animals move faster as they approach the goal. Hull (1934) measured rats' running speed along a straight alley and found that they ran faster as they neared the food. J. S. Brown (1948) showed that gradients of approach (and of avoidance) were steeper when the animal's motivation was stronger.

Seven decades later, Kivetz, Urminsky, and Zheng (2006) asked whether the same principle governs human consumers in reward programs. In a field study at a university café that used a "buy ten, get one free" coffee card, customers purchased coffee more frequently as they approached the free coffee. In an online music-rating program that rewarded participants after a given number of ratings, participants visited more often and rated more songs per visit as they neared the reward. After earning a reward, purchase rates fell back: a *post-reward resetting* of motivation. The researchers also documented **illusionary goal progress**: a 12-stamp card pre-stamped with two stamps was completed faster than a 10-stamp card with no pre-stamps, although both required ten purchases.

Figure 19.4 shows the pattern in stylised form.

19.5.2 Mechanism and extensions

Why does proximity motivate? Three explanations complement one another. In Hull's behaviourist account, actions closer to the reward are more strongly reinforced. In a cognitive account, each action produces a larger proportional reduction in the remaining distance as the goal approaches, the

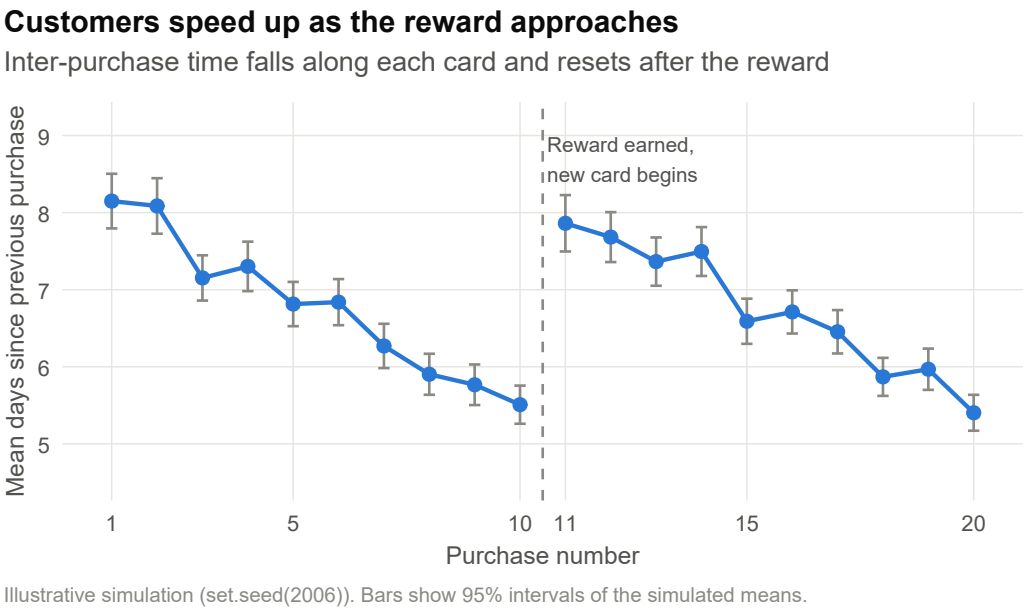


Figure 19.4. Illustrative goal-gradient pattern in a 'buy ten, get one free' program: the interval between purchases shortens as customers approach the reward and resets after the reward is earned. Simulated data (500 customers) constructed to mirror the qualitative pattern reported by Kivetz, Urminsky, and Zheng (2006).

last stamp eliminates 100% of the remaining distance, the first only 10%, so the *perceived marginal impact* of effort rises. In a reference-point account, falling just short of a goal is experienced as a loss (Heath, Larrick, and Wu 1999).

The perceived-progress mechanism explains the **endowed progress effect**. Nunes and Drèze (2006) gave car-wash customers loyalty cards that required eight purchases for a free wash. For half the customers the card had eight spaces. For the other half it had ten spaces with two already stamped. Customers given the "endowed" cards completed them substantially more often and faster, although both groups needed the same eight purchases. Progress that is framed as already under way appears to increase commitment to finishing.

Motivation is not always monotonic in progress. Bonezzi, Brendl, and De Angelis (2011) showed that motivation can be "stuck in the middle": people focus on progress made so far (*to-date*) near the start and on what remains (*to-go*) near the end, and both frames produce their largest marginal effects at the extremes, so motivation sags mid-way. Koo and Fishbach (2008) found that whether emphasising accomplished or remaining actions increases motivation depends on commitment: emphasising what has been accomplished helps when commitment is uncertain, and emphasising what remains helps when commitment is already high.

Competition modifies the gradient further. Fershtman and Gneezy (2011) studied tournaments in which the stakes were raised and found that higher stakes increased effort among those still in contention but also led more participants to quit. And Dai et al. (2018) showed that high expectations can make *favoured* competitors more likely to quit when they fall behind: quitting provides a face-saving excuse that protects the inference of high ability.

19.5.3 Applications in business

BUSINESS CASE

Loyalty programs and gamification. The goal gradient is built into the architecture of modern loyalty and engagement systems: punch cards and stamp cards, points thresholds for status tiers, progress bars on profile completion, and streaks in learning apps. The research suggests five design principles. (1) *Make progress visible*: the gradient depends on perceived distance. (2) *Endow initial progress*: start customers part-way along the card, as in Nunes and Drèze (2006). (3) *Shorten the middle*: break long goals into sub-goals to avoid the mid-point slump (Bonezzi, Brendl, and De Angelis 2011). (4) *Plan for the reset*: after a reward, offer an immediate new goal so that resetting does not become attrition (Kivetz, Urminsky, and Zheng 2006). (5) *Beware of competitive leaderboards*: those far from the top may disengage (Fershtman and Gneezy 2011).

TRY IT YOURSELF

Design two versions of a loyalty card for a local bakery: a standard ten-stamp card and an endowed-progress card. Specify how you would randomise cards across customers, what outcome you would measure (completion rate, time to completion, post-reward return), and how you would distinguish a goal-gradient effect from seasonal variation in purchases.

19.6 Self-Determination Theory

THEORY CARD

Core claim. Motivation varies not only in amount but in *type*. Autonomous motivation, acting from interest or personally endorsed values, produces better persistence, performance on complex tasks, and well-being than controlled motivation driven by rewards, punishments, or guilt. Satisfying the needs for autonomy, competence, and relatedness fosters autonomous motivation.
Origins. Social psychology, Deci and Ryan (1985), R. M. Ryan and Deci (2000), applied to work by Gagné and Deci (2005).
Level of analysis. Individual (with social-contextual antecedents).
Key constructs. Intrinsic motivation, extrinsic motivation, internalisation, external, introjected, identified, and integrated regulation, basic psychological needs, undermining (crowding-out) effect.
Boundary conditions. Undermining effects apply to tasks that are initially interesting. Rewards perceived as informational rather than controlling can support motivation.
Related theories. Expectancy theory, Goal-setting theory, Agency theory (Section 6.1).

19.6.1 Core logic

Deci and Ryan (1985) distinguished **intrinsic motivation**, doing an activity for its inherent satisfaction, from **extrinsic motivation**: doing it for a separable outcome. Self-determination theory refines the second category. Extrinsic motivation can be more or less internalised, ranging from **external regulation** (acting to obtain rewards or avoid punishment), through **introjected regu-**

lation (acting to avoid guilt or to maintain self-esteem), to **identified regulation** (acting because one personally values the outcome) and **integrated regulation** (the value is fully assimilated into one's sense of self) (R. M. Ryan and Deci 2000). The theory groups external and introjected regulation as *controlled* motivation, and identified, integrated, and intrinsic regulation as *autonomous* motivation (Figure 19.5).

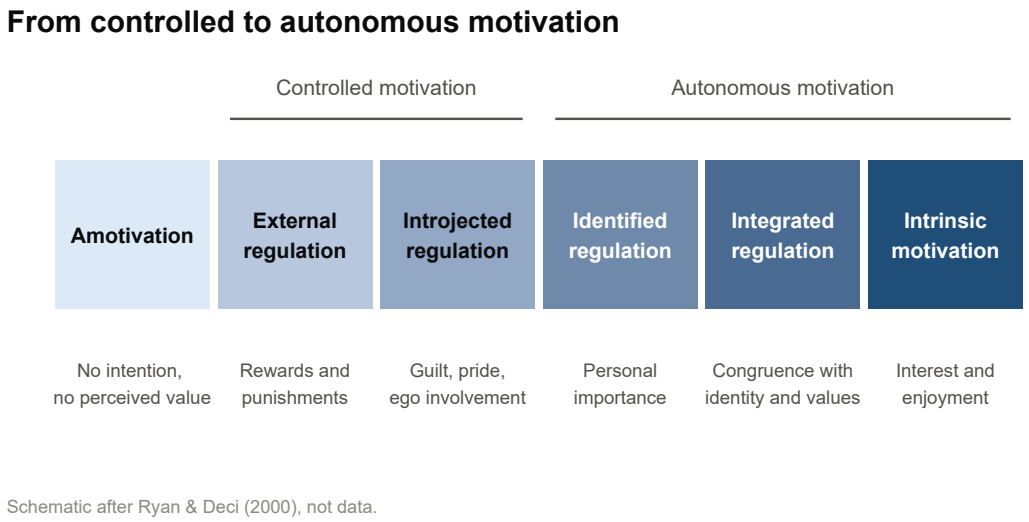


Figure 19.5. The self-determination continuum (after Ryan & Deci, 2000). Forms of motivation are ordered by the degree to which regulation is experienced as autonomous. External and introjected regulation are 'controlled', identified, integrated, and intrinsic regulation are 'autonomous'.

Social contexts move people along this continuum by supporting or thwarting three **basic psychological needs**: *autonomy* (experiencing one's behaviour as self-endorsed), *competence* (feeling effective), and *relatedness* (feeling connected to others) (R. M. Ryan and Deci 2000). At work, autonomy-supportive supervision, acknowledging employees' perspectives, offering choice, and explaining the rationale for requests, is associated with more autonomous motivation, better performance on complex tasks, and higher well-being (Gagné and Deci 2005).

19.6.2 Mechanism: when rewards backfire

The most provocative prediction of self-determination theory is the **undermining effect**: tangible rewards can reduce intrinsic motivation for an activity that was initially interesting. In a classic study, Lepper, Greene, and Nisbett (1973) found that preschool children who drew with markers in order to receive an expected "good player" award later spent less free time drawing than children who had not been offered the award. Self-determination theory explains the effect through the *functional significance* of the reward: a reward experienced as **controlling** shifts the perceived locus of causality from "I do this because I enjoy it" to "I do this for the reward" and thwarts autonomy. A reward experienced as **informational** conveys competence and need not undermine motivation (Figure 19.6).

19.6.3 Evidence and extensions

Deci, Koestner, and Ryan (1999) meta-analysed experiments on rewards and intrinsic motivation, measured by free-choice persistence and self-reported interest. Tangible rewards that were expected and contingent on doing, completing, or performing a task undermined intrinsic motivation

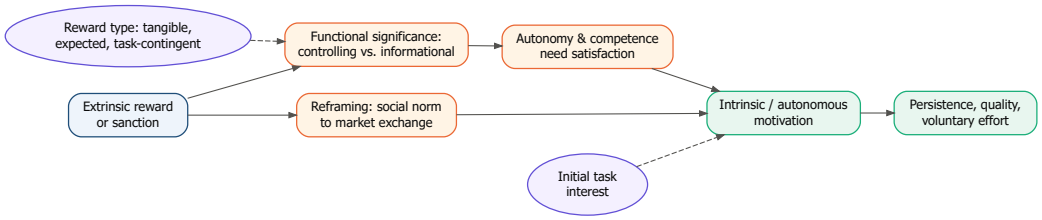


Figure 19.6. Mechanism of motivation crowding. Whether an extrinsic reward undermines or supports intrinsic motivation depends on whether it is experienced as controlling or informational, which in turn depends on reward type and contingency, and on whether the reward reframes the situation from a social to a market exchange.

for interesting tasks. Unexpected tangible rewards and verbal rewards (positive feedback) did not, and verbal rewards tended to enhance it.

Economists documented parallel phenomena under the label **motivation crowding** (Frey and Jegen 2001). Gneezy and Rustichini (2000b) found that paying participants a small piece rate for answering IQ-test questions produced *lower* performance than paying nothing, whereas a larger piece rate raised performance above the small-payment level. Similarly, students collecting charitable donations collected less when offered a small commission than when offered nothing. Hence the title: "pay enough or don't pay at all". In a field experiment at Israeli day-care centres, Gneezy and Rustichini (2000a) found that introducing a modest fine for parents who collected their children late *increased* late arrivals, and the higher level persisted after the fine was removed. The fine appears to have converted a moral obligation into a price that parents were willing to pay.

DEBATE AND EVIDENCE

How general is crowding out? Cameron and Pierce (1994) meta-analysed many of the same experiments and concluded that rewards had little overall detrimental effect on intrinsic motivation, prompting a long-running exchange about which studies to include and how to classify reward conditions. Deci, Koestner, and Ryan (1999) was in part a response. A later synthesis by Cerasoli, Nicklin, and Ford (2014) suggests a reconciliation: intrinsic motivation predicted performance, especially its *quality*, while extrinsic incentives better predicted performance *quantity*. The two jointly predicted performance, and the role of intrinsic motivation was weaker when incentives were directly tied to performance. The practical conclusion is not that incentives are harmful, but that their *design*, contingency, framing, and the type of task, determines whether they complement or crowd out intrinsic motivation.

19.6.4 Applications in business

BUSINESS CASE

Intrinsic motivation in knowledge work. Amabile (1996) argued that intrinsic motivation is conducive to creativity, whereas controlling extrinsic constraints , close monitoring, evaluation emphasised for its own sake, rewards contingent on narrow outputs, tend to undermine it, and that informational feedback and recognition can support it. For managers of engineers, designers, analysts, and researchers, self-determination theory implies several practices: give teams meaningful choice over *how* they meet goals. Frame metrics as feedback rather than

surveillance. Use recognition and development opportunities that signal competence, and avoid piece-rate incentives for open-ended creative work, where contingent tangible rewards are most likely to crowd out interest (Gagné and Deci 2005, Deci, Koestner, and Ryan 1999). Pay still matters, as a fair, predictable base that removes concern about money, consistent with equity theory, but it is a poor tool for stimulating curiosity.

RESEARCH TOOLKIT

Using self-determination theory in research.

- *H1.* Performance-contingent bonuses will increase output quantity but reduce the novelty of solutions in open-ended design tasks.
- *H2.* The negative effect of contingent rewards on persistence will be weaker when rewards are framed as feedback about competence.
- *H3.* Introducing small payments for prosocial behaviour will reduce participation relative to no payment (crowding out), but larger payments will increase it.

Measures. Free-choice persistence, Work Extrinsic and Intrinsic Motivation or Motivation at Work scales, basic need satisfaction scales. **Designs.** Randomised reward framing, field experiments with payment levels (Gneezy and Rustichini 2000b), panel data following changes in incentive regimes.

COMMON PITFALL

Treating "intrinsic versus extrinsic" as a zero-sum dichotomy. Most work behaviour is extrinsically motivated in some sense, yet it can still be highly autonomous (identified or integrated). The relevant question is usually not whether to use incentives, but how to design them so that employees experience them as fair and informational rather than controlling.

19.7 Losing Leads to Winning

THEORY CARD

Core claim. Being slightly behind a competitor or a reference point can increase effort enough to raise the probability of eventual success.

Origins. Behavioural economics and psychology. Berger and Pope (2011), building on reference-dependent preferences (Section 24.5) (Kahneman and Tversky 1979) and goals as reference points (Heath, Larrick, and Wu 1999).

Level of analysis. Individual and team in competitive settings.

Key constructs. Reference point, distance from the reference point, motivational effort, competitive feedback.

Boundary conditions. Applies only when the deficit is *small*. Large deficits reduce perceived attainability and can trigger quitting. Requires that effort can still affect the outcome.

Related theories. Goal gradient, Goal-setting theory, Prospect theory (Chapter 24), Social comparison (Section 22.1).

19.7.1 Core logic and evidence

Berger and Pope (2011) analysed a large archive of professional and college basketball games. Being ahead at half-time naturally predicts winning, but the relationship was not smooth near zero: teams that were behind by a single point at half-time won *more* often than teams that were ahead by a single point. To isolate the mechanism, the authors ran laboratory experiments in which participants competed on an effortful task and received manipulated feedback at a break. Participants told that they were slightly behind their opponent exerted more effort than those told that they were tied or slightly ahead.

The explanation combines two ideas from this chapter. First, the tie acts as a salient **reference point**. Trailing by a little places a team in the domain of losses, where the value function is steepest and the motivation to close the gap is strongest (Kahneman and Tversky 1979, Heath, Larrick, and Wu 1999). Second, a small deficit preserves **expectancy**: catching up is clearly feasible. A large deficit preserves the loss frame but destroys expectancy, which is why the effect should, and in the data does, appear only for small margins. Figure 19.7 shows the pattern in stylised form.

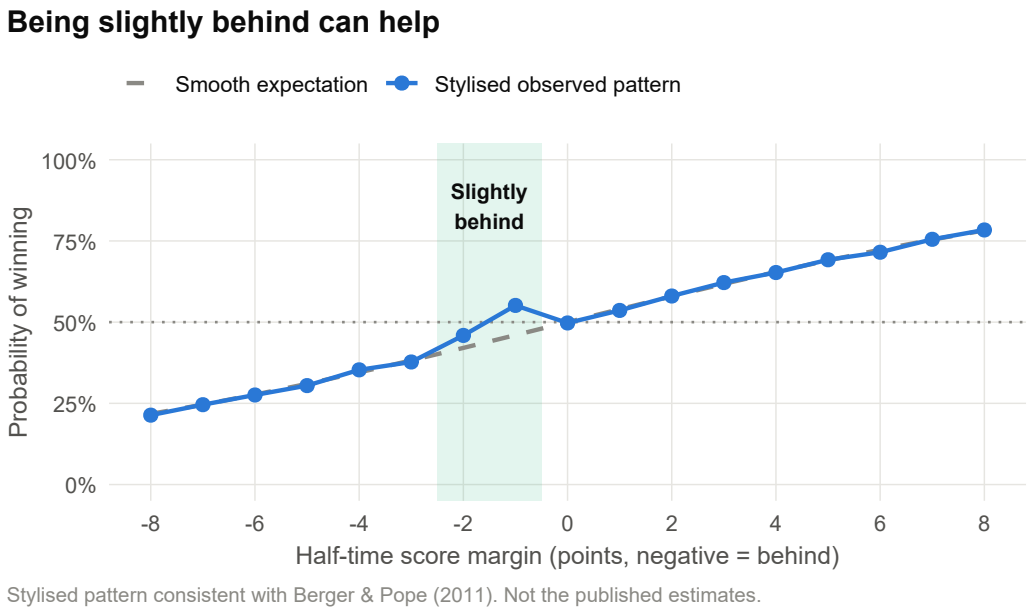


Figure 19.7. Stylised illustration of the 'losing leads to winning' pattern reported by Berger and Pope (2011): win probability generally rises with the half-time margin, but teams slightly behind win more often than a smooth trend predicts. Values are illustrative, not the published estimates.

DEBATE AND EVIDENCE

How robust is the effect? The basketball pattern is a discontinuity at a single point in a noisy outcome, and such patterns invite scrutiny about specification and generalisability. The laboratory evidence strengthens the causal story by manipulating feedback directly, and the broader literature on reference points (Pope and Schweitzer 2011, E. J. Allen et al. 2017) supports the mechanism. At the same time, competitive settings differ: in tournaments with high stakes, falling behind can induce quitting rather than effort (Fershtman and Gneezy 2011), and favourites who fall behind may withdraw effort to protect their image (Dai et al. 2018). The theory therefore predicts a narrow band in which deficits motivate, bounded on one side

by complacency and on the other by discouragement.

19.7.2 Applications in business

BUSINESS CASE

Illustrative scenario: designing sales contests and progress feedback. A regional sales director runs a monthly contest between branches. A conventional leaderboard shows every branch its rank. The losing-leads-to-winning logic suggests a different design: give each branch feedback relative to a *near* competitor ("you are 3% behind the Riverside branch") rather than to the leader, reset contests frequently so that most branches are within reach at the mid-point, and avoid displaying large deficits that invite quitting. The same logic applies to fundraising thermometers, fitness apps that compare users with similar peers, and internal hackathons that publish interim standings.

RESEARCH TOOLKIT

Using the theory in research.

- *H1.* Teams or individuals slightly behind at an interim checkpoint will exert more subsequent effort than those slightly ahead. The effect will reverse for large deficits.
- *H2.* The effect will be stronger when interim standings are salient and when effort can still affect outcomes.

Designs. Regression discontinuity at interim ties in sports, sales contests, or crowdfunding, laboratory experiments manipulating interim feedback, field experiments randomising peer-comparison feedback. **Outcomes.** Effort measures (activity counts, time on task) and final success.

19.8 Integrating the Theories

The theories in this chapter are complementary rather than rival. Expectancy theory explains the *calculus* of effort. Goal-setting theory explains how specific targets focus that calculus. The goal gradient and losing-leads-to-winning phenomena show how the calculus changes with distance from a reference point. Equity theory adds that rewards are judged *relative to others*, and self-determination theory warns that the *type* of motivation matters as much as its quantity. A well-designed incentive system, for customers or employees, attends to all five: attainable goals (expectancy), credible links between performance and reward (instrumentality), fair comparisons (equity), progress visibility (gradient), and autonomy-supportive framing (self-determination).

19.9 Chapter Summary

KEY TAKEAWAYS

- **Content theories** (Maslow, Herzberg) ask what motivates. **Process theories** ask how motivation works and generally have stronger empirical support (Wahba and Bridwell 1976, Latham and Pinder 2005).
- **Expectancy theory** treats motivation as E x I x V. A weak link anywhere collapses moti-

vation, although meta-analytic evidence favours the components more than their exact multiplicative combination (Vroom 1964, Van Eerde and Thierry 1996).
- **Equity theory** explains reactions to the *ratio* of outcomes to inputs relative to a referent. Under-reward provokes withdrawal and even theft, mitigated by adequate explanations (J. S. Adams 1965, Greenberg 1990, Colquitt et al. 2001).
- **Specific, difficult goals** raise performance through direction, effort, persistence, and strategy (Locke and Latham 2002), but poorly governed goals can motivate narrow focus and misconduct (Ordóñez et al. 2009, Schweitzer, Ordóñez, and Douma 2004).
- **Effort accelerates near goals**, in rats, coffee buyers, and online raters, and perceived progress can be engineered (Hull 1932, Kivetz, Urminsky, and Zheng 2006, Nunes and Drèze 2006).
- **Self-determination theory** distinguishes autonomous from controlled motivation. Expected, contingent tangible rewards can crowd out intrinsic motivation, and small payments or fines can backfire (R. M. Ryan and Deci 2000, Deci, Koestner, and Ryan 1999, Gneezy and Rustichini 2000b, 2000a).
- **Being slightly behind** can increase effort and success, within a narrow band bounded by complacency and discouragement (Berger and Pope 2011).

Table 19.7 presents the comparison of the chapter's motivation theories.

Table 19.7. Comparison of the chapter's motivation theories.

Theory	Key mechanism	Signature prediction	Main boundary condition	Typical business use
Expectancy theory	Beliefs about effort, performance, and rewards	A near-zero component collapses motivation	Calculative choice, within-person	Quota and bonus design
Equity theory	Social comparison of outcome/input ratios	Under-reward triggers withdrawal or restitution	Salient referent, equity sensitivity	Pay transparency, pay-cut communication
Goal-setting theory	Direction, effort, persistence, strategy	Specific difficult goals beat 'do your best'	Commitment, feedback, task complexity	Performance management, OKRs
Goal gradient	Perceived progress and marginal impact	Effort accelerates near the goal	Perceived not actual distance, competition	Loyalty programs, gamification
Self-determination theory	Need satisfaction, internalisation	Controlling rewards undermine interest	Initially interesting tasks	Knowledge-work incentives, volunteering
Losing leads to winning	Reference point plus attainability	Small deficits raise success	Only for small deficits	Contests, peer-comparison feedback

19.10 Discussion Questions

1. Vroom's model is multiplicative, but meta-analytic evidence suggests that additive combinations predict as well. What does this imply about how people actually combine expectancy, instrumentality, and valence? Design an experiment that distinguishes the two models.
2. A firm is about to publish salary bands internally. Using equity theory and the organizational-justice literature, predict which employees will react most negatively and propose three steps

to mitigate those reactions.

3. Was the Wells Fargo cross-selling scandal a failure of goal setting or a failure of goal governance? Use the exchange between Ordóñez et al. and Locke and Latham to build your argument.
4. Kivetz et al. found that customers accelerate as they approach a reward and slow down after it. How would you redesign a coffee-shop loyalty program to exploit the first finding and counteract the second?
5. When are monetary incentives likely to crowd out intrinsic motivation, and when are they likely to crowd it in? Formulate a hypothesis that specifies the moderating conditions.
6. Under what conditions would you expect "losing leads to winning" to fail in a business setting such as a sales contest or a start-up competition? How would you test your prediction?

19.11 Further Reading

- Latham and Pinder (2005): an authoritative review of work-motivation theory and evidence at the turn of the century.
- Vroom (1964): the founding statement of expectancy theory. Read with Van Eerde and Thierry (1996) for the empirical record.
- J. S. Adams (1965): the definitive presentation of equity theory and the responses to inequity. Pair with Greenberg (1990) for a vivid field test.
- Locke and Latham (2002): a concise synthesis of 35 years of goal-setting research. Read with Ordóñez et al. (2009) and Locke and Latham (2009) for the debate on side effects.
- Kivetz, Urminsky, and Zheng (2006): the modern demonstration that the goal gradient governs consumer behaviour in reward programs.
- R. M. Ryan and Deci (2000): the clearest short introduction to self-determination theory. See Deci, Koestner, and Ryan (1999) for the evidence on rewards and intrinsic motivation.
- Gneezy and Rustichini (2000b) and Gneezy and Rustichini (2000a): two short, influential papers showing when incentives backfire.
- Berger and Pope (2011): an elegant combination of field and laboratory evidence that being slightly behind can increase success.

Chapter 20

Self-Regulation and Construal

Two advertisements for the same insurer run side by side. One promises that a policy will "keep everything you have worked for safe". The other invites customers to "get more out of every journey". The first wins easily among homeowners comparing household cover. The second performs better among people booking an overseas holiday. A retirement plan that asks employees to raise their contribution rate *today* attracts few takers, yet the same plan, offered as a commitment to save more *after the next pay rise*, is embraced. A product team that cheerfully commits to an ambitious launch twelve months away finds, as the date approaches, that every conversation is about logistics, budgets, and what can realistically be cut.

None of these puzzles is about what people want. They are about **how people pursue what they want**: the strategies they use, the level of abstraction at which they represent goals, and their struggles to act on long-term intentions when short-term temptations are close at hand. This chapter introduces four closely related theories of **self-regulation**, the processes by which people guide their thoughts, feelings, and behaviour toward desired end-states. **Regulatory focus theory** distinguishes two motivational systems: one oriented to advancement, one to security. **Regulatory fit** explains why the *manner* of goal pursuit can add value of its own. **Construal level theory** links psychological distance to abstract or concrete thinking. And theories of **self-control and intertemporal choice** explain why preferences reverse as rewards draw near, and what individuals and organizations can do about it. Together they form one of the most practically useful toolkits in consumer behaviour, organizational behaviour, behavioural economics, and public policy (Table 20.1).

Table 20.1. Chapter at a glance: theories of self-regulation and construal.

Theory	Core question	Level of analysis	Key source
Regulatory focus theory	Do people pursue goals by seeking gains or by guarding against losses?	Individual (chronic or situational)	Higgins (1997, 1998)
Regulatory fit	Why does the manner of goal pursuit change the value of what is chosen?	Individual x situation	Higgins (2000), Cesario et al. (2004)
Construal level theory	How does psychological distance shape mental representation and choice?	Individual	Liberman & Trope (1998), Trope & Liberman (2010)
Self-control and intertemporal choice	Why do people act against their own long-run interests, and how can they bind themselves?	Individual / household / organization	Thaler & Shefrin (1981), Laibson (1997)

20.1 Regulatory Focus Theory

THEORY CARD

Core claim. People regulate behaviour through two distinct motivational systems. A **promotion focus** is concerned with advancement, accomplishment, and aspirations and is sensitive to the presence or absence of gains. A **prevention focus** is concerned with safety, duties, and responsibilities and is sensitive to the absence or presence of losses.
Origins. Social psychology. Higgins (1997) and Higgins (1998), growing out of self-discrepancy theory (Higgins 1987).
Level of analysis. Individual. Focus can be a chronic disposition or a state induced by the situation (task instructions, message framing, cultural context).
Key constructs. Promotion and prevention focus, ideals versus oughts, gains and non-gains versus non-losses and losses, eager versus vigilant strategic means, risky versus conservative response bias.
Boundary conditions. Predictions require that the outcome can be framed as an opportunity or a threat, and that people have latitude in *how* they pursue it. Focus effects are weaker when choices are highly constrained or when people are indifferent to the outcome.
Related theories. Regulatory fit, Prospect theory (Chapter 24), Self-determination theory (Section 19.6), Hofstede's cultural dimensions (Section 18.2), Goal-setting theory (Section 19.4).

20.1.1 Origins and intellectual history

For most of the twentieth century, psychology explained motivation with the **hedonic principle**: people approach pleasure and avoid pain. Higgins (1997) argued that this principle, though true, is incomplete because it says nothing about *how* people approach pleasure and avoid pain. Two people may both want a promotion at work, yet one experiences the goal as a hope to be realised and the other as an obligation to be met. The first will feel cheerful on success and dejected on failure. The second will feel relieved and calm on success and anxious on failure.

The theory grew out of Higgins's earlier **self-discrepancy theory**, which distinguished between the *ideal self* (the person one hopes to be) and the *ought self* (the person one believes one should be) and showed that discrepancies from each produce different emotions: dejection-related emotions for ideal discrepancies and agitation-related emotions for ought discrepancies (Higgins 1987). Regulatory focus theory generalised this insight into two self-regulatory systems that serve different survival needs: **nurturance** (growth, the presence of positive outcomes) and **security** (protection, the absence of negative outcomes) (Higgins 1997, 1998).

20.1.2 Core logic and propositions

DEFINITION

Definition 20.1. Promotion focus. A self-regulatory state in which goals are represented as ideals, hopes, wishes, and aspirations, and outcomes are coded as *gains* versus *non-gains*.
Prevention focus. A self-regulatory state in which goals are represented as oughts, duties, obligations, and responsibilities, and outcomes are coded as *non-losses* versus *losses*.

A crucial feature of the theory is that both systems are forms of approaching a desired end-state. Prevention is *not* the same as avoidance: a prevention-focused employee approaches a desired state (a project delivered without errors) by guarding against mistakes. What differs is the **strategic means** used. Table 20.2 summarises the two systems.

Table 20.2. Promotion and prevention focus compared (based on Higgins, 1997, 1998, Crowe & Higgins, 1997).

Dimension	Promotion focus	Prevention focus
Underlying need	Nurturance, growth, advancement	Security, safety, protection
Goal representation	Ideals: hopes and aspirations	Oughts: duties and obligations
Outcomes that matter	Gains vs. non-gains	Non-losses vs. losses
Preferred strategic means	Eagerness: ensure hits, avoid errors of omission	Vigilance: ensure correct rejections, avoid errors of commission
Response bias under uncertainty	Risky ('yes' bias)	Conservative ('no' bias)
Emotion after success	Cheerfulness, elation	Quiescence, relief, calm
Emotion after failure	Dejection, disappointment	Agitation, anxiety
Typical message that resonates	'Get energized', 'gain an edge'	'Stay protected', 'avoid costly mistakes'
Example product categories	Travel, education, entertainment, investment growth	Insurance, security, health screening, compliance

The theory's central propositions can be stated as follows:

- **P1 (Two systems).** Promotion and prevention are distinct self-regulatory systems that can be activated chronically or situationally. They are not the two ends of a single continuum, and an individual can be high or low on each.
- **P2 (Outcome sensitivity).** Promotion-focused people are more sensitive to the presence and absence of positive outcomes. Prevention-focused people are more sensitive to the absence and presence of negative outcomes (Higgins 1997).
- **P3 (Strategic inclination).** Promotion focus induces *eager* strategies, seizing opportunities, generating alternatives, and accepting the risk of false alarms. Prevention focus induces *vigilant* strategies, avoiding mistakes and accepting the risk of missed opportunities (Crowe and Higgins 1997).
- **P4 (Emotional signature).** Success and failure produce cheerfulness and dejection under promotion, and quiescence and agitation under prevention (Higgins 1997, Brockner and Higgins 2001).
- **P5 (Antecedents).** Chronic focus develops from socialization histories (for example, being rewarded for accomplishments versus being protected from danger) and from self-construal (Section @ref(self-construal (Section 18.4))), with independent self-views tilting toward promotion and interdependent self-views toward prevention (Higgins et al. 2001, A. Y. Lee, Aaker, and Gardner 2000).

20.1.3 Mechanism

Figure 20.1 traces the causal chain. Chronic and situational sources set the regulatory focus. Focus determines the strategic means a person prefers, and those means shape decisions and emotions.

The clearest early test of the strategic-means link comes from Crowe and Higgins (1997). They

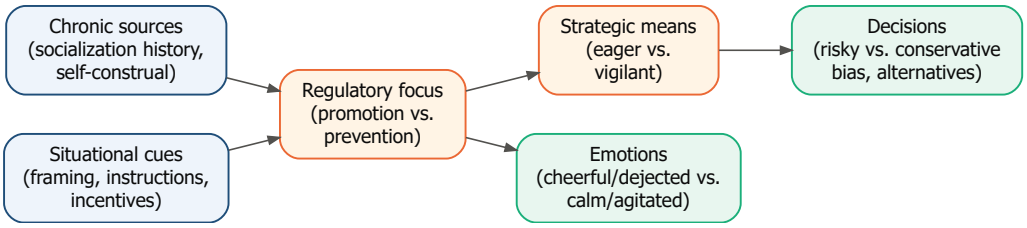


Figure 20.1. Mechanism of regulatory focus: chronic and situational sources activate a promotion or prevention focus, which shapes the preferred strategic means (eager versus vigilant) and, through them, decisions and emotional responses.

induced promotion or prevention focus by framing the same task in terms of accomplishment or of safety, and then asked participants to make recognition judgements under uncertainty. In signal-detection terms, promotion-focused participants showed a **risky bias**, they were more willing to say "yes, I saw it", accepting more false alarms in order to secure hits, whereas prevention-focused participants showed a **conservative bias**, accepting more misses in order to avoid false alarms. The same objective task was approached with different strategies depending on how the goal was represented.

20.1.4 Evidence and extensions

The theory has been extended in several directions that matter for business research.

- **Culture and self-construal.** A. Y. Lee, Aaker, and Gardner (2000) found that people with a more independent self-construal regarded promotion-framed information as more important, whereas those with a more interdependent self-construal regarded prevention-framed information as more important, a pattern that also appeared in comparisons of participants from North American and East Asian cultural backgrounds. The link connects regulatory focus with the cultural dimensions discussed under Hofstede (Section 18.2).
- **Persuasion.** J. L. Aaker and Lee (2001) showed that appeals consistent with the reader's self-view were better remembered and more persuasive: promotion-framed appeals (for example, emphasising energy and enjoyment) for those with an independent self-view, and prevention-framed appeals (emphasising protection against disease) for those with an interdependent self-view. The title of their article , "I" seek pleasures and "we" avoid pains, captures the result.
- **Role models.** Promotion-focused people are inspired by positive role models who show what success looks like, whereas prevention-focused people are more motivated by negative role models who illustrate failures to be avoided (Lockwood, Jordan, and Kunda 2002).
- **Work and organizations.** Brockner and Higgins (2001) brought the theory into organizational behaviour, arguing that leaders, job design, and feedback practices induce promotion or prevention states with predictable emotional consequences. A meta-analysis by Lanaj, Chang, and Johnson (2012) found that the two foci relate to different sets of work outcomes, with promotion focus more closely linked to outcomes involving initiative and innovation and prevention focus more closely linked to safety-related performance.
- **Measurement.** The Regulatory Focus Questionnaire (RFQ) asks people about their history of promotion and prevention success (Higgins et al. 2001). Experiments more often *induce* focus by asking participants to write about hopes and aspirations versus duties and obligations,

or by framing task incentives as gains or losses.

DEBATE AND EVIDENCE

Is regulatory focus distinct from neighbouring constructs? Critics have asked whether promotion and prevention are simply relabelled versions of approach and avoidance motivation, of positive and negative affectivity, or of gain and loss framing in prospect theory (Chapter 24). Defenders answer that the theory makes predictions those frameworks do not: prevention is an *approach* toward safety rather than avoidance. The two systems are characterised by distinct strategies (eager versus vigilant) rather than by valence alone, and the emotional signatures differ in kind, not only in sign (Higgins 1997, 1998). A second debate concerns measurement. Chronic measures and situational inductions do not always produce the same pattern, and different questionnaires capture somewhat different facets of the constructs. Researchers should state whether they theorise about chronic dispositions, momentary states, or both, and choose operationalizations accordingly.

20.1.5 Applications in business

Regulatory focus offers a practical vocabulary for segmenting customers, designing products, and writing messages. Some product categories are inherently coded as prevention goods (insurance, home security systems, antivirus software, retirement annuities), others as promotion goods (travel, education, fitness equipment marketed for performance, growth funds), and many can be positioned either way. A massage chair can be sold as a reward and a pleasure (promotion) or as protection against back injury (prevention).

BUSINESS CASE

Illustrative scenario: one insurer, two journeys. A mid-sized insurer sells both household cover and travel insurance through the same website. For household cover, the buyer is typically thinking about what could be lost, a house, possessions, savings, and the natural framing is protective: "Make sure a burst pipe never becomes a financial disaster." For travel insurance, the purchase occurs in the middle of planning something the customer *wants*. The dominant state is promotional, and copy that speaks of making the most of the trip ("Go further, worry less, enjoy every day of it") fits better. Regulatory focus theory predicts that tailoring headlines, imagery, and calls to action to the focus evoked by each purchase context will raise engagement and conversion relative to a single generic message. Because the prediction is about fit between message and mindset, it is directly testable with an A/B experiment that crosses message frame with purchase context.

Focus can also shift with the economic environment. A plausible, though, to our knowledge, not directly tested, application concerns financial markets: after a severe market downturn such as the collapse of technology stocks in 2000-2002, many investors may move from a promotion state (seeking upside) toward a prevention state (protecting capital), increasing the appeal of products and messages that emphasise security. This is best treated as a hypothesis to be tested, for example by examining whether prevention-framed fund advertising gains relative effectiveness after market crashes.

20.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Consumers primed with a prevention focus will show greater preference for products with established safety certifications than consumers primed with a promotion focus.
- *H2.* Sales employees with a strong chronic promotion focus will generate more new leads but also more unqualified leads than employees with a strong prevention focus (eager versus vigilant strategies).
- *H3.* The effect of a manager's transformational, vision-based communication on employee innovation will be mediated by employees' situational promotion focus.

Designs. Experiments that induce focus (writing tasks about hopes versus duties. Gain versus loss incentive framing) followed by choice or performance tasks. Field experiments that vary ad copy across purchase contexts. Panel surveys that measure chronic focus with the RFQ and link it to behaviour.

Measures. Regulatory Focus Questionnaire (Higgins et al. 2001), experimental manipulation checks (e.g., ratings of how much one thought about aspirations versus responsibilities), signal-detection indices of response bias (Crowe and Higgins 1997), emotion measures distinguishing dejection from agitation.

Contexts. Advertising and message testing, insurance and financial products, health and safety behaviour, innovation and safety climates in organizations, negotiation (focus on maximising gains versus avoiding a bad deal).

COMMON PITFALL

Equating regulatory focus with positive versus negative framing. A loss-framed message is not automatically a prevention message, and a gain-framed message is not automatically a promotion message. Regulatory focus refers to the person's *motivational orientation* and preferred *strategies*. Message framing is one way to *induce* or *fit* that orientation. Studies that manipulate only gain/loss wording and then interpret the results as evidence about promotion and prevention need a manipulation check showing that focus itself changed.

20.2 Regulatory Fit

THEORY CARD

Core claim. When people pursue a goal in a manner that sustains their regulatory orientation, eager means under promotion, vigilant means under prevention, they experience **regulatory fit**. Fit makes people "feel right" about what they are doing, strengthens engagement, and transfers value to the object of choice or to a persuasive message.

Origins. Social psychology, Higgins (2000), with experimental demonstrations in Higgins et al. (2003) and Cesario, Grant, and Higgins (2004), consumer applications in A. Y. Lee and Aaker (2004).

Level of analysis. Individual × situation (the match between a person's orientation and the means or message they encounter).

Key constructs. Regulatory orientation, strategic means, fit versus non-fit, feeling right, engagement strength, processing fluency (Section 23.8), value transfer.
Boundary conditions. Fit effects depend on the feeling being attributed to the target. When people recognise that their positive feeling comes from an irrelevant source, the effect weakens. Fit amplifies existing evaluations rather than creating them from nothing.
Related theories. Regulatory focus, Processing fluency (Section 23.8), Feelings-as-information theory (Section 23.7), Elaboration likelihood model (Section 26.1).

20.2.1 Core logic

Classic theories of decision making locate value in *outcomes*: a choice is good because what it delivers is good. Higgins (2000) proposed an additional source of value: **value from fit**. When the way a decision is made suits the decision maker's orientation, people feel right about the decision and value its outcome more, independently of the outcome's hedonic properties.

The idea was tested directly in studies where participants chose between a coffee mug and a pen. Participants who made their choice using strategies that fit their orientation, thinking about what they would *gain* under promotion, or what they would *lose* under prevention, subsequently assigned more monetary value to the same chosen mug than participants whose strategies did not fit (Higgins et al. 2003). Nothing about the mug changed, only the experience of choosing did.

Cesario, Grant, and Higgins (2004) extended the logic to persuasion. When the framing of a persuasive message fitted recipients' orientation, the message was judged more persuasive. Their account was that fit creates a subjective experience of **feeling right**, which recipients use as information about the message: consistent with feelings-as-information theory (Section 23.7). In line with this account, the fit effect diminished when participants were led to attribute their feelings to a source other than the message.

A. Y. Lee and Aaker (2004) offered a complementary, cognitive account in consumer research. Promotion-focused consumers are more persuaded by **gain frames** (emphasising the presence or absence of positive outcomes), and prevention-focused consumers are more persuaded by **loss frames** (emphasising the absence or presence of negative outcomes). Their studies indicated that fit increases **processing fluency**, the subjective ease with which a message is processed, and that fluency, in turn, helps explain more favourable attitudes. Figure 20.2 sketches the characteristic crossover pattern.

20.2.2 Mechanism

Two mechanisms have been proposed and are not mutually exclusive. The **experiential** route holds that fit produces a feeling of rightness and stronger engagement, which is then misattributed to the target (Cesario, Grant, and Higgins 2004, Higgins 2000). The **fluency** route holds that fit makes information easier to process, and that ease is read as a signal of truth or liking (A. Y. Lee and Aaker 2004). Both routes predict that fit effects should be larger when people rely on feelings (low involvement, time pressure, ambiguous attributes) and smaller when they are motivated and able to scrutinise arguments: an implication that links fit to the elaboration likelihood model (Section 26.1).

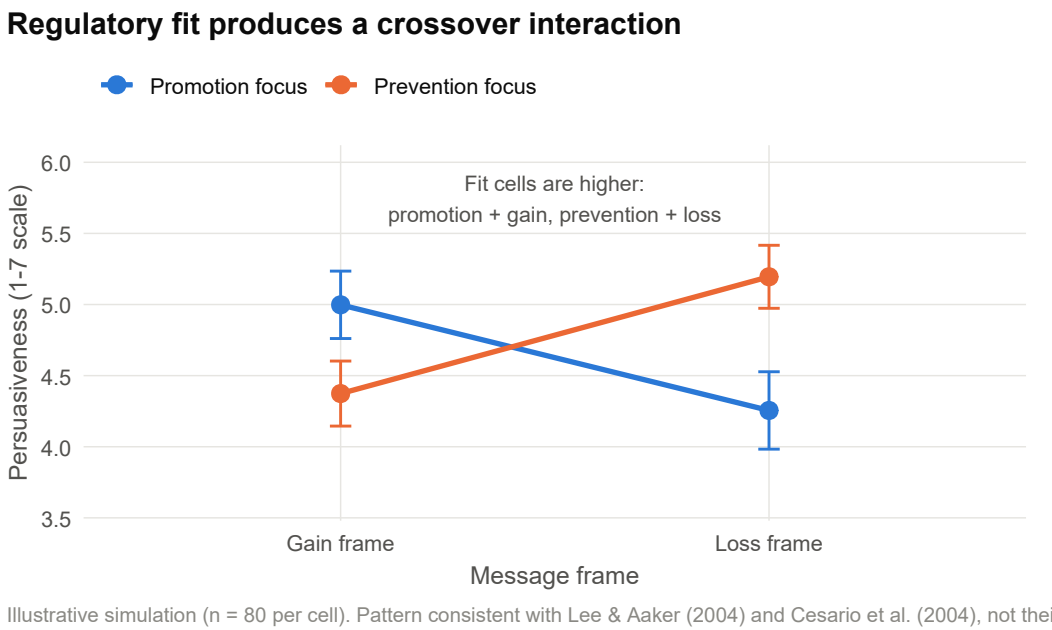


Figure 20.2. Illustrative regulatory fit interaction: persuasion is higher when the message frame matches the audience's regulatory focus (promotion with gain frames, prevention with loss frames). Simulated data. Points are cell means with 95% confidence intervals.

20.2.3 Applications in business

BUSINESS CASE

Message-market fit in digital advertising. Regulatory fit gives marketers a principled reason to match *how* a benefit is described to the mindset in which it is encountered. A sunscreen positioned as a means to "enjoy every hour outdoors" fits a promotion state. The same product described as a way to "avoid sunburn and skin damage" fits prevention. Because many digital channels reveal the context in which an ad is viewed (a travel-planning search versus a medical-information search), firms can test fit by crossing message frame with context and comparing click-through and conversion. The theory also warns against a common practice, mixing gain and loss language in the same headline, which may dilute fit for both segments.

TRY IT YOURSELF

Take a product you know well and write two one-sentence value propositions: one that fits a promotion focus (gains, aspirations, eager means) and one that fits a prevention focus (non-losses, responsibilities, vigilant means). Identify a customer context in which each would be encountered, and specify the outcome metric you would use to test the prediction.

COMMON PITFALL

Testing fit without measuring or manipulating the orientation. A message effect that varies across segments is not, by itself, evidence of regulatory fit. A fit study needs both factors, orientation *and* means or frame, and should show the crossover interaction. Where pos-

sible, measure the proposed mechanism (feeling right, fluency, engagement) and test whether making the source of the feeling salient weakens the effect.

20.3 Construal Level Theory

THEORY CARD

Core claim. People represent psychologically distant objects and events abstractly, in terms of their central, goal-relevant features (**high-level construal**), and psychologically near objects and events concretely, in terms of their contextual, incidental details (**low-level construal**). Construal level in turn shapes prediction, evaluation, and choice.
Origins. Social psychology, temporal construal theory (Liberman and Trope 1998, Trope and Liberman 2003), generalised to all forms of psychological distance in Trope and Liberman (2010), consumer applications in Trope, Liberman, and Wakslak (2007).
Level of analysis. Individual.
Key constructs. Psychological distance (temporal, spatial, social, hypothetical), construal level, desirability versus feasibility, "why" versus "how" representations, primary versus secondary features.
Boundary conditions. Effects require that the target can be represented at more than one level of abstraction. When concrete details are themselves the goal (for example, an engineer specifying tolerances) distance may matter less.
Related theories. Self-control, Regulatory focus, Dual-process theory (Section 23.6), Mental accounting (Section 23.5).

20.3.1 Origins and intellectual history

Construal level theory (CLT) began with a narrow observation about time. People planning a holiday for next year imagine sunshine, rest, and a change of scene. People leaving for the same holiday next week think about packing, airport transfers, and whether the pet sitter has confirmed. Liberman and Trope (1998) formalised this as **temporal construal**: distant-future events are represented in terms of their **desirability** (the value of the end-state, the *why* of an action), whereas near-future events are represented in terms of their **feasibility** (the ease or difficulty of reaching it, the *how*). The distinction echoes **action identification theory**, which showed that any action can be described at a higher level ("improving my health") or a lower level ("walking on the treadmill") (Vallacher and Wegner 1987).

Trope and Liberman (2003) developed the temporal account into a general theory of how time affects representation and preference. Trope and Liberman (2010) then argued that time is only one of several dimensions of **psychological distance**, the subjective experience that something is removed from the self, here, and now, and that all of them operate through the same representational mechanism.

20.3.2 Core logic and propositions

Table 20.3 lists the four distance dimensions identified by Trope and Liberman (2010), and Table 20.4 contrasts the two levels of construal.

Table 20.3. The four dimensions of psychological distance (Trope & Liberman, 2010).

Dimension	Near (low distance)	Far (high distance)	Business example
Temporal	Now, tomorrow, next week	Next year, a decade from now	Product launch next week vs. next year
Spatial	Here, in this office or city	There, in another country or region	Local store vs. overseas market
Social	Self, close friends, ingroup	Strangers, other people, outgroup	Choosing for oneself vs. for a typical customer
Hypothetical	Likely, real, certain	Unlikely, imagined, uncertain	Committed budget vs. contingency scenario

Table 20.4. High-level versus low-level construal (based on Liberman & Trope, 1998, Trope & Liberman, 2003, 2010).

Feature	High-level construal (distant)	Low-level construal (near)
Representation	Abstract, schematic, decontextualised	Concrete, detailed, contextualised
Features emphasised	Primary, central, goal-relevant	Secondary, incidental, goal-irrelevant
Action identification	Why (ends, purposes)	How (means, procedures)
Choice criterion	Desirability of outcomes	Feasibility of execution
Arguments weighted	Pros (reasons for)	Cons (obstacles against)
Information used	Ideals, values, prototypes	Practical constraints, situational details
Example description	'Building a healthier lifestyle'	'Putting trainers in my bag tonight'

The theory's main propositions can be stated as follows:

- **P1 (Distance-to-abstraction).** As psychological distance increases, people represent targets at a higher level of construal (Trope and Liberman 2010).
- **P2 (Weighting).** At greater distance, choices weight high-level features (desirability, central attributes, values) more heavily and low-level features (feasibility, peripheral attributes, practical constraints) less heavily (Liberman and Trope 1998).
- **P3 (Preference reversals).** Because the weights change with distance, options that are attractive in principle but hard in practice are preferred from afar and abandoned up close (Trope and Liberman 2003).
- **P4 (Interchangeability).** Different dimensions of distance are mentally associated and have similar effects. Making one dimension distant makes others feel more distant too (Trope and Liberman 2010).
- **P5 (Bidirectionality).** Construal level also influences perceived distance: thinking abstractly makes events feel further away (Trope and Liberman 2010).

In a well-known demonstration, Liberman and Trope (1998) asked students to choose between an assignment that was interesting but difficult and one that was uninteresting but easy. When the assignment was due in the distant future, students showed a stronger preference for the interesting, difficult option: they weighted desirability. When it was due soon, the easy option gained ground: feasibility loomed larger.

20.3.3 Mechanism

Figure 20.3 summarises the mechanism. Distance on any dimension raises the level of construal. Construal level changes the weights attached to desirability and feasibility, and these weights drive evaluation and choice. Self-control can be seen as one important downstream outcome: abstract representations keep long-term goals in view (Fujita et al. 2006).

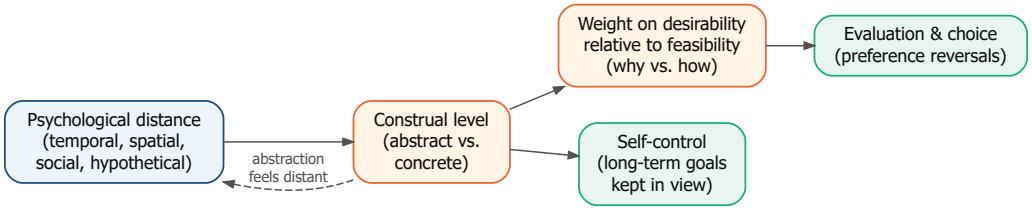


Figure 20.3. Mechanism of construal level theory: psychological distance raises the level of construal, which shifts weight from feasibility to desirability and changes evaluation, choice, and self-control. The dashed arrow indicates that abstract construal also makes targets feel more distant.

20.3.4 Evidence and extensions

CLT has generated a very large literature. Trope, Liberman, and Wakslak (2007) reviewed its implications for consumer psychology: distant purchase decisions are guided more by central attributes and brand values, near decisions more by price, convenience, and details of use. Fujita et al. (2006) showed that inducing high-level construal improved self-control, making people less drawn to immediate rewards and more persistent at an uncomfortable task that served a valued goal. Two meta-analyses by Soderberg et al. (2015) found a reliable effect of psychological distance on abstraction across many paradigms and all four dimensions, while also documenting considerable heterogeneity across studies.

DEBATE AND EVIDENCE

How robust and how general is construal level theory? CLT is among the most cited theories in consumer psychology, but its breadth invites scrutiny. Critics note that many studies use small samples and a wide variety of manipulations and dependent measures, which makes it difficult to know which effects would replicate and how large they are. Meta-analytic estimates from a literature with possible publication bias must be read cautiously (Soderberg et al. 2015). A second question concerns *mechanism*: some findings attributed to construal level might reflect other processes that covary with distance, such as reduced uncertainty about near events, mood, or simple differences in the information available. The most persuasive studies measure construal directly, manipulate it independently of distance, and show that it mediates the effect on choice.

20.3.5 Applications in business

CLT is especially useful whenever a business must communicate about events at different distances, launches, subscriptions, pre-orders, long-term contracts, international expansion, or must reconcile ambitious plans with operational reality.

BUSINESS CASE

Illustrative scenario: launch messaging for distant versus near events. A consumer-electronics firm announces a new device a year before release and opens pre-orders three weeks before shipping. CLT suggests that the two moments call for different messages. At announcement, customers construe the product abstractly, so communication should emphasise *why*: the vision, the core benefit, the identity the product expresses. As release approaches, customers switch to *how*: compatibility, set-up, delivery dates, and return policies. A campaign that leads with specification sheets a year out, or with visionary slogans the week of shipping, mismatches the customer's construal. The same logic applies to events: conferences sell the big idea months ahead and the logistics in the final weeks.

BUSINESS CASE

The planning fallacy in organizations. Strategic plans are made at a distance and executed up close. At the planning stage, desirability dominates and feasibility constraints are underweighted, so ambitious initiatives are approved readily. As implementation nears, the same initiatives encounter the concrete obstacles that were always there. CLT suggests two complementary remedies: deliberately bringing feasibility into early decisions (pre-mortems, concrete implementation plans, reference-class data on similar projects), and deliberately bringing desirability into late decisions (reminding teams of the purpose of the project when tactical problems threaten to derail it).

20.3.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Advertisements for products available in the distant future will be more effective when they emphasise desirability attributes, and advertisements for products available immediately will be more effective when they emphasise feasibility attributes.
- *H2.* Managers evaluating projects for a distant foreign subsidiary (spatial distance) will weight strategic fit more and implementation cost less than managers evaluating equivalent projects for a local unit.
- *H3.* Inducing high-level construal will increase employees' contributions to long-term savings plans, mediated by greater weight on future outcomes.

Designs. Laboratory experiments manipulating distance or construal directly (e.g., "why/how" ladders that ask participants to explain why or how they would pursue a goal). Field experiments varying message content by time-to-event. Archival studies of pre-order and early-booking behaviour.

Measures. Action-identification measures based on Vallacher and Wegner (1987), categorisation breadth, ratio of desirability to feasibility attributes in open-ended reasons, choices between desirable-but-difficult and easy-but-less-desirable options.

COMMON PITFALL

Assuming that distance always helps. High-level construal is not uniformly better. It promotes value-consistent choices and self-control, but it can also produce over-commitment,

under-planning, and neglect of practical constraints. Research and managerial advice should specify *which* features are central to the goal: if feasibility *is* the central issue (for example, safety-critical operations), abstract thinking may lead people astray.

20.4 Self-Control and Intertemporal Choice

THEORY CARD

Core claim. People often prefer larger, later rewards when both options are distant but switch to smaller, sooner rewards as the sooner option becomes imminent. This **dynamic inconsistency** creates a conflict between a far-sighted planner and a myopic doer. Individuals and institutions respond with rules, commitment devices, and choice architecture (Section 23.10).
Origins. Behavioural economics and psychology: the planner-doer model (R. H. Thaler and Shefrin 1981), hyperbolic discounting (Ainslie 1975, Laibson 1997), delay-of-gratification research (Mischel, Shoda, and Rodriguez 1989), ego depletion (Baumeister et al. 1998), reviewed in Frederick, Loewenstein, and O'Donoghue (2002).
Level of analysis. Individual and household, with organizational and policy implications (benefit design, product design, regulation).
Key constructs. Discount function (exponential, hyperbolic, quasi-hyperbolic), present bias, preference reversal, naivete versus sophistication, willpower, commitment device, default.
Boundary conditions. Present bias is strongest for immediately available, visceral rewards (food, leisure, spending) and for outcomes whose costs are delayed. It is weaker when both options lie in the future or when the decision is automated.
Related theories. Construal level theory, Mental accounting (Section 23.5), Bounded rationality (Section 23.1), Goal-gradient hypothesis (Section 19.5).

20.4.1 Origins and intellectual history

Standard economic models assume that people discount future utility **exponentially**, at a constant rate per period. Exponential discounting implies **time consistency**: if someone prefers 110 dollars in 31 days to 100 dollars in 30 days, she should also prefer 110 dollars tomorrow to 100 dollars today. Behaviour often violates this prediction. Ainslie (1975) showed that animals and people display **preference reversals** consistent with a **hyperbolic** discount function, under which value falls steeply over short delays and gently over long ones. Laibson (1997) introduced a tractable approximation, the **quasi-hyperbolic** or β-δ model, and showed that a present-biased consumer can benefit from illiquid assets that restrict future choices. Frederick, Loewenstein, and O'Donoghue (2002) provide the standard review of the evidence.

DEFINITION

Definition 20.2. Exponential discounting. The value today of a reward A delivered after delay t is $V = A\,\delta^t$, with $0 < \delta < 1$. The relative value of two rewards does not depend on when they are evaluated.
Hyperbolic discounting. $V = A/(1 + kt)$, with $k > 0$. Value declines steeply at short delays, so the ranking of a smaller-sooner and a larger-later reward can reverse as the sooner reward approaches.

Quasi-hyperbolic (β-δ) discounting. $V = A$ if $t = 0$ and $V = \beta\,\delta^t A$ if $t > 0$, with $0 < \beta < 1$ capturing a special premium on the present (Laibson 1997).

Psychology and economics converged on a common picture. R. H. Thaler and Shefrin (1981) modelled the individual as an organization with two parts: a far-sighted **planner** concerned with lifetime utility and a series of myopic **doers** who live only in the present. The planner can influence the doers either by exerting costly willpower or, more efficiently, by adopting **rules** and institutions, pension plans, payroll deductions, illiquid savings, that change the doers' options or incentives. O'Donoghue and Rabin (1999) added an important distinction: **sophisticated** agents know they are present-biased and seek commitment, whereas **naive** agents believe their future selves will be patient and therefore procrastinate.

20.4.2 Mechanism

Figure 20.4 shows why hyperbolic discounting produces preference reversals. A person can receive a smaller reward (50 units) at time 10 or a larger reward (100 units) at time 15. Viewed from far away, both rewards are heavily discounted and the larger, later reward is worth more. Under hyperbolic discounting, as the decision-maker approaches time 10, the value of the smaller, sooner reward rises steeply and overtakes the larger reward: the classic pattern of the dieter who orders dessert or the saver who raids the holiday fund. Under exponential discounting the two curves never cross: whatever is preferred from afar is preferred up close.

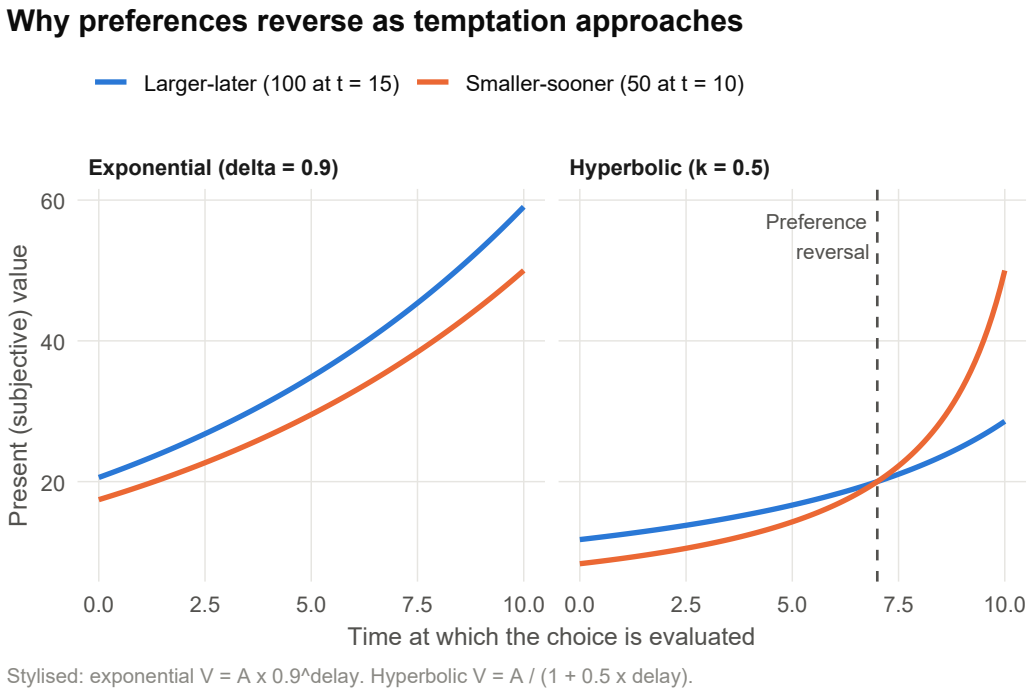


Figure 20.4. Stylised comparison of exponential and hyperbolic discounting. A smaller-sooner reward (50 units at t = 10) competes with a larger-later reward (100 units at t = 15). Under hyperbolic discounting the curves cross, producing a preference reversal as the sooner reward approaches. Under exponential discounting they do not.

Figure 20.5 places this conflict inside the planner-doer framework. Temptations act directly on the

doer. The planner can intervene through costly willpower in the moment or, more reliably, through rules and commitments adopted in advance.

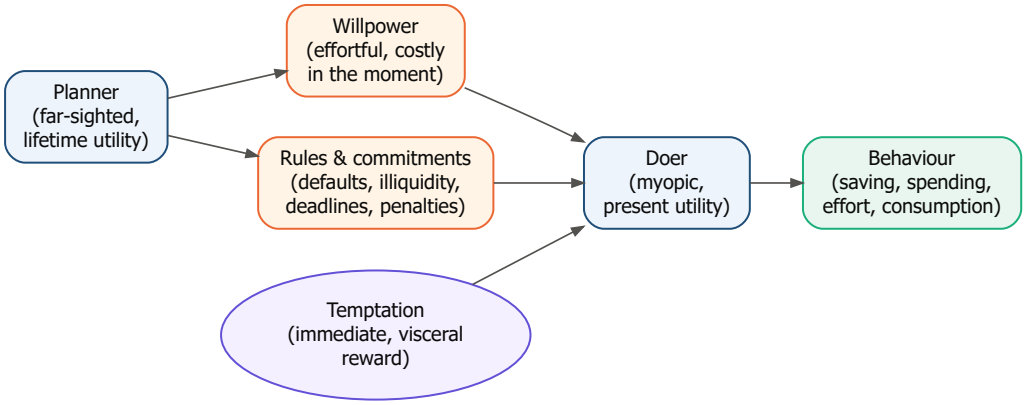


Figure 20.5. The planner-doer model of self-control (after Thaler & Shefrin, 1981): the far-sighted planner influences the myopic doer through costly willpower or through rules and commitment devices that change the doer's options and incentives.

20.4.3 Evidence and extensions

Delay of gratification. In the best-known psychological research on self-control, preschool children were offered a small treat immediately or a larger treat if they could wait alone for the experimenter to return. Mischel, Shoda, and Rodriguez (1989) reported that children who waited longer tended, years later, to be rated as more cognitively and socially competent in adolescence and to have higher scholastic aptitude scores, and that the strategies children used, for example, distracting themselves or thinking about the treat in abstract rather than "hot" terms, strongly affected how long they could wait. The latter finding anticipated construal level theory: cooling the representation of a temptation makes it easier to resist.

A conceptual replication by T. W. Watts, Duncan, and Quan (2018), using a larger and more socioeconomically diverse sample, found that early delay ability was associated with later achievement, but the association was considerably smaller than in the original studies and was substantially reduced once family background, early cognitive ability, and the home environment were controlled. The replication does not show that self-control is unimportant. It shows that a single childhood test is a noisy and confounded measure of it.

Ego depletion. Baumeister et al. (1998) proposed that self-control draws on a limited resource. In their studies, participants who had to resist tempting food (eating radishes while chocolates sat in front of them) subsequently gave up sooner on unsolvable puzzles than participants who had not exercised self-control. The **strength model** compared self-control to a muscle that tires with use and can be strengthened with practice (Muraven and Baumeister 2000). Experience-sampling work later documented that desires, and conflicts between desires and other goals, are a pervasive part of everyday life (Hofmann et al. 2012).

> **DEBATE AND EVIDENCE**
>
> **The rise and fall of ego depletion.** Ego depletion became one of the most replicated-seeming

findings in social psychology, with hundreds of published studies. Its standing changed when Hagger et al. (2016) coordinated a preregistered replication across some two dozen laboratories with more than 2,000 participants, using a standardised depletion task and outcome measure. The pooled effect was close to zero. Defenders responded that the chosen manipulation may have been too weak and that other paradigms might show the effect. Critics replied that a literature built largely on small, flexible studies could have produced a consistent picture through publication bias alone. In parallel, Inzlicht and Schmeichel (2012) proposed a **process model** in which prior exertion shifts motivation and attention, away from "have-to" goals and toward "want-to" rewards, rather than depleting a fuel-like resource. For applied researchers the safest conclusion is that "willpower fatigue" should not be assumed as a mechanism without direct evidence. The episode is discussed further in the chapter section on the replication crisis (Section 3.7).

Commitment devices. If people are at least partly sophisticated about their present bias, they should value ways to bind their future selves. Field evidence supports this. When a Philippine bank offered a savings account that restricted withdrawals until a self-chosen goal or date was reached, a meaningful share of clients took it up and savings rose relative to a control group (Ashraf, Karlan, and Yin 2006). Smokers who were offered a savings account they would forfeit if they failed a nicotine test were more likely to have quit at follow-up than a control group (Giné, Karlan, and Zinman 2010). A widely cited study of self-imposed deadlines in the classroom (Ariely and Wertenbroch 2002) has been retracted by *Psychological Science* (2026), and we do not rely on its findings. The broader case for deadlines and precommitment rests on the theoretical work of R. H. Thaler and Shefrin (1981) and O'Donoghue and Rabin (1999) and on field evidence such as the studies above.

Paying for self-control that never materialises. Naive present bias predicts that people will choose contracts that would be optimal for a more disciplined future self. DellaVigna and Malmendier (2006) found that gym members who chose flat-rate monthly contracts attended so infrequently that they paid more per visit than they would have on a pay-per-visit basis: consistent with overconfidence about future self-control.

Table 20.5 organises the main self-control tools by the mechanism through which they work.

Table 20.5. Self-control tools and the mechanisms through which they work.

Tool	Mechanism	Example	Key evidence
Defaults	Make the patient choice the path of least resistance	Automatic enrolment in a retirement plan	Madrian & Shea (2001)
Pre-commitment to future increases	Bind the future self when the cost is distant	Save More Tomorrow escalation	Thaler & Benartzi (2004)
Illiquidity / restricted access	Raise the cost of giving in to temptation	Locked savings accounts, pension rules	Laibson (1997), Ashraf et al. (2006)
Commitment contracts with penalties	Attach a loss to failure	Forfeitable deposit for quitting smoking	Gine et al. (2010)
Construal shifts	Represent the temptation abstractly	Reframing a purchase in terms of long-run goals	Fujita et al. (2006)
Rules and bright lines	Replace in-the-moment deliberation with a policy	Spending only from a separate account	Thaler & Shefrin (1981)

20.4.4 Applications in business

BUSINESS CASE

Save More Tomorrow. R. H. Thaler and Benartzi (2004) designed a retirement-savings programme that turns each insight of this section into a design feature. Employees who were saving too little were invited to commit *in advance* to raising their contribution rate, with each increase timed to coincide with a future pay rise. Because the increases lay in the future, present bias worked *for* the plan rather than against it. Because they coincided with pay rises, take-home pay never fell, which avoided the sting of a perceived loss, and because the escalation continued automatically unless employees opted out, inertia kept them in. In the first implementation, average saving rates among participants rose from 3.5 percent to 13.6 percent over 40 months. Together with evidence on automatic enrolment (Madrian and Shea 2001), the programme helped shape the automatic-escalation features now common in employer-sponsored retirement plans.

BUSINESS CASE

Designing products for (and against) present bias. Many business models interact with self-control problems. Subscriptions, free trials that convert automatically, and "buy now, pay later" schemes exploit the gap between a present-biased doer and a planner who expects to cancel or repay on time. The gym contracts studied by DellaVigna and Malmendier (2006) are a canonical example. Other firms sell commitment as a product: savings apps that lock funds until a goal date, commitment-contract platforms that donate a user's stake if a goal is missed, and employer wellness programmes with pre-committed incentives. The managerial question is ethical as well as strategic: designs that profit from customers' predictable failures invite regulatory scrutiny and erode trust, whereas designs that help customers achieve their own long-run goals can build durable loyalty.

20.4.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Offering employees the option to pre-commit future bonuses to a savings account will increase total savings more than an equivalent option to save from current pay.
- *H2.* Customers who select annual over monthly subscription plans will show lower usage relative to their stated intentions, and the gap will be larger for customers with stronger present bias.
- *H3.* Inducing high-level construal at the point of sale will reduce impulsive add-on purchases.

Designs. Field experiments with randomised access to commitment products. Choice tasks that elicit discount rates at different horizons to detect present bias. Regression discontinuity around default changes. Panel data on contract choice and subsequent usage.

Measures. Elicited β and δ from intertemporal choice tasks. Demand for commitment (willingness to pay for restrictions). Trait self-control scales. Behavioural outcomes such as savings balances, attendance, completion, and cancellation.

COMMON PITFALL

Inferring present bias from any impatience. A high discount rate is not the same as present bias. Liquidity constraints, uncertainty about whether a future reward will be paid, and differences in the utility of money over time can all produce apparent impatience. Evidence for dynamic inconsistency requires showing that preferences *change* as outcomes approach, or that people *demand* commitment, not merely that they prefer sooner rewards.

20.5 Integrating the Theories

The four theories are best seen as complementary lenses on the same broad question: how do people translate goals into action? Regulatory focus describes the *orientation* with which a goal is pursued. Regulatory fit describes the added value that arises when means match that orientation. Construal level theory describes the *level of abstraction* at which goals and options are represented, and self-control theories describe what happens when the abstract, long-term goal collides with a concrete, immediate temptation.

The connections are more than metaphorical. Construal level provides one account of *why* self-control fails: near temptations are represented concretely and vividly, while the long-term goal they threaten is distant and abstract (Fujita et al. 2006). Commitment devices work in part by moving decisions to a moment when the planner, thinking abstractly, is in charge: the logic of Save More Tomorrow (R. H. Thaler and Benartzi 2004). And regulatory focus shapes which commitment devices appeal: a prevention-focused saver may respond to a message about protecting against old-age poverty, a promotion-focused saver to a message about the freedom that savings will buy. Researchers who combine these lenses can design interventions that match orientation, construal, and timing.

20.6 Chapter Summary

KEY TAKEAWAYS

- **Regulatory focus theory** distinguishes a **promotion** system (ideals, gains, eager strategies) from a **prevention** system (oughts, non-losses, vigilant strategies) (Higgins 1997, Crowe and Higgins 1997).
- Focus has chronic and situational sources, including self-construal and culture, and it shapes persuasion, risk-taking, and emotion (A. Y. Lee, Aaker, and Gardner 2000, J. L. Aaker and Lee 2001).
- **Regulatory fit** adds value from the *manner* of goal pursuit: messages and choices feel right, are processed more fluently, and are valued more when means match orientation (Higgins 2000, Cesario, Grant, and Higgins 2004, A. Y. Lee and Aaker 2004).
- **Construal level theory** links psychological distance to abstraction: distant events are represented in terms of desirability and "why", near events in terms of feasibility and "how" (Liberman and Trope 1998, Trope and Liberman 2010).
- **Self-control** problems arise from dynamic inconsistency. The planner-doer model and hyperbolic discounting explain preference reversals (R. H. Thaler and Shefrin 1981, Laibson 1997).
- Evidence on the marshmallow test and ego depletion has been substantially revised by

large replications (T. W. Watts, Duncan, and Quan 2018, Hagger et al. 2016). Commitment devices, defaults, and pre-commitment designs such as Save More Tomorrow have stronger field support (R. H. Thaler and Benartzi 2004, Madrian and Shea 2001, Ashraf, Karlan, and Yin 2006).

Table 20.6 presents the comparison of the chapter's theories.

Table 20.6. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Typical business use	Main caution
Regulatory focus	Promotion vs. prevention orientation, eager vs. vigilant means	Risky vs. conservative bias, distinct emotions	Segmentation, product positioning, safety and innovation climates	Do not equate with gain/loss wording
Regulatory fit	Feeling right, processing fluency	Crossover interaction of orientation x frame	Message-context matching, A/B testing of copy	Requires both factors and the interaction
Construal level	Distance raises abstraction, desirability vs. feasibility	Preference reversals across distance	Launch timing and messaging, planning, international strategy	Heterogeneous effects, measure construal
Self-control and intertemporal choice	Present bias, planner vs. doer	Preference reversal as rewards approach, demand for commitment	Retirement design, subscriptions, commitment products	Impatience is not the same as present bias

20.7 Discussion Questions

1. Regulatory focus theory claims that prevention is a form of *approach*, not avoidance. Explain the distinction and design an experiment that could discriminate between the two interpretations.
2. A bank wants to increase uptake of income-protection insurance among young professionals, a segment it believes is promotion-focused. Should it reframe the product in promotion terms or try to induce a prevention focus? Justify your answer using regulatory fit.
3. Using construal level theory, explain why corporate sustainability pledges with distant target dates are easy to announce and hard to deliver. What governance mechanisms would you recommend?
4. The marshmallow test and ego depletion were both revised by large replications. What features of the original research designs made the initial findings fragile, and how would you design a study on self-control today?
5. Save More Tomorrow combines pre-commitment, loss aversion (Section 24.4), and inertia. Identify another domain (health, energy, learning) where the same three design principles could be combined, and specify how you would evaluate the result.
6. When does designing for customers' present bias become exploitation? Propose criteria a firm could use to distinguish helpful commitment products from harmful ones.

20.8 Further Reading

- Higgins (1997): the foundational statement of regulatory focus theory, with its argument for moving "beyond pleasure and pain".
- Higgins (2000): introduces value from fit and explains why the manner of goal pursuit matters for decision value.
- A. Y. Lee and Aaker (2004): a clear consumer-research demonstration of regulatory fit, linking it to processing fluency and persuasion.
- Trope and Liberman (2010): the integrative review of construal level theory and the four dimensions of psychological distance.
- R. H. Thaler and Shefrin (1981): the planner-doer model, an early and still illuminating economic theory of self-control.
- Frederick, Loewenstein, and O'Donoghue (2002): the standard review of time discounting and its anomalies.
- R. H. Thaler and Benartzi (2004): a model of how behavioural theory can be turned into a working product design.
- Hagger et al. (2016): the multilab replication of ego depletion and a case study in how cumulative science corrects itself.

Chapter 21

Attitudes, Attribution, and Social Learning

Why do buyers of an expensive car read more of the manufacturer's advertising *after* they have bought it than before? Why does a customer who waited two hours for a delayed flight forgive the airline when a storm is to blame but vow never to fly with it again when the crew simply failed to show up? Why do people who sign a small petition become more willing, weeks later, to put a large and ugly sign on their front lawn, and why does a "limited time only" banner sometimes drive shoppers away rather than toward the checkout? Each puzzle concerns the way people make sense of their own behaviour, of other people's behaviour, and of the pressures placed on them by firms.

This chapter brings together six classic theories from social psychology that remain workhorses of research in marketing, organizational behaviour, and information systems. **Cognitive dissonance theory** and **self-perception theory** explain how behaviour shapes attitudes, rather than the reverse. **Attribution theory** explains how people infer the causes of events and how those inferences drive emotion, blame, and loyalty. **Balance theory** explains how attitudes toward people and objects are organized into consistent structures, which is the logic behind celebrity endorsement. **Social cognitive theory** explains how people learn by watching others and why beliefs about one's own capabilities (self-efficacy) predict performance. **Reactance theory** explains why pressure, scarcity, and unsolicited advice can backfire. Together they form a toolkit for anyone who designs persuasion, service recovery, training, or choice environments (Table 21.1).

Table 21.1. Chapter at a glance: theories of attitudes, attribution, and social learning.

Theory	Core question	Level of analysis	Key source
Cognitive dissonance theory	How do people reduce the discomfort of inconsistent cognitions?	Individual	Festinger (1957)
Self-perception theory	How do people infer their attitudes from their own behaviour?	Individual	Bem (1967, 1972)
Attribution theory	How do people explain the causes of behaviour and outcomes?	Individual / dyad	Heider (1958), Kelley (1967), Weiner (1985)
Balance theory	How are attitudes toward people and objects kept consistent?	Triad (person, other, object)	Heider (1946)
Social cognitive theory	How do observation and self-efficacy shape learning and action?	Individual in context	Bandura (1977, 1986)
Reactance theory	Why do threats to freedom provoke resistance?	Individual	Brehm (1966)

21.1 Cognitive Dissonance Theory

THEORY CARD

Core claim. Holding two psychologically inconsistent cognitions produces an aversive state, **dissonance**, that motivates people to restore consistency, often by changing the attitude that is easiest to change rather than the behaviour that caused the problem.
Origins. Social psychology. Festinger (1957), with the classic experiments of J. W. Brehm (1956), Festinger and Carlsmith (1959), and Aronson and Mills (1959). Later revisions include Cooper and Fazio (1984), Steele (1988), and Harmon-Jones, Harmon-Jones, and Levy (2015).
Level of analysis. Individual.
Key constructs. Consonant and dissonant cognitions, magnitude of dissonance, importance of the cognitions, perceived choice, insufficient justification, effort, irrevocability, dissonance reduction (attitude change, adding consonant cognitions, trivialization, selective exposure, behaviour change).
Boundary conditions. Dissonance requires that people feel personally responsible for a behaviour (high perceived choice) and that the inconsistency matters to them. It is weakened when external justification is ample, when arousal can be attributed to something else, or when the self is affirmed in another domain.
Related theories. Self-perception theory, Escalation of commitment (Section 14.2), IKEA effect (Section 29.7), Balance theory, Expectancy disconfirmation (Section 27.1).

21.1.1 Origins and intellectual history

Leon Festinger developed dissonance theory in the 1950s, shortly after his work on social comparison (Section 22.1) (Festinger 1954). His starting observation was that people dislike inconsistency among their beliefs, attitudes, and knowledge of their own behaviour. A smoker who knows that smoking causes cancer holds two cognitions that do not fit together. Festinger proposed that such inconsistency produces a state of psychological tension, analogous to hunger, that people are motivated to reduce (Festinger 1957).

The theory became famous because of its counterintuitive predictions. In the best-known experiment, Festinger and Carlsmith (1959) had participants perform an hour of tedious tasks (turning pegs and loading spools) and then paid them either one dollar or twenty dollars to tell a waiting student that the tasks were interesting. Reinforcement theory predicted that a larger payment would produce more positive attitudes toward the task. Instead, participants paid one dollar later rated the tasks as more enjoyable than those paid twenty dollars. Twenty dollars was ample justification for a small lie. One dollar was not, so participants reduced dissonance by persuading themselves that the tasks had not been so dull after all. This **insufficient justification** effect became the template of the **induced-compliance paradigm**.

Two other paradigms followed. In the **free-choice paradigm**, J. W. Brehm (1956) asked participants to rate household products, choose one of two as a gift, and rate the products again. After choosing, participants rated the chosen product more favourably and the rejected product less favourably than before, a pattern called the **spreading of alternatives**. In the **effort-justification paradigm**, Aronson and Mills (1959) found that people who underwent a severe and embarrassing initiation to join a discussion group later rated a deliberately dull group discussion more favourably

than people who underwent a mild initiation or none. Having suffered to obtain something, people value it more.

21.1.2 Core logic and propositions

Dissonance theory specifies when dissonance arises, how large it is, and how it is reduced. Its central propositions can be stated as follows:

- **P1 (Inconsistency).** Two cognitions are dissonant when, considered alone, the opposite of one would follow from the other. Dissonance is a motivational state that people seek to reduce.
- **P2 (Magnitude).** The magnitude of dissonance increases with the importance of the cognitions and with the ratio of dissonant to consonant cognitions.
- **P3 (Justification).** The less external justification (reward, threat, instruction) there is for a counterattitudinal act, the greater the dissonance and the greater the subsequent attitude change.
- **P4 (Choice and responsibility).** Dissonance arises mainly when people feel that they chose freely to act and could have foreseen the consequences (Cooper and Fazio 1984).
- **P5 (Path of least resistance).** People reduce dissonance through whichever cognition is least resistant to change. Because past behaviour is fixed, attitudes, beliefs about consequences, and the perceived importance of the issue usually give way.
- **P6 (Selective exposure).** To prevent or reduce dissonance, people seek information that supports their decisions and avoid information that challenges them.

Table 21.2 summarises the experimental paradigms that operationalize these propositions and their business analogues.

Table 21.2. Five paradigms for studying cognitive dissonance.

Paradigm	Classic study	What creates dissonance	Typical reduction	Business analogue
Induced compliance	Festinger & Carlsmith (1959)	Saying or doing something counterattitudinal for little reward	Attitude moves toward the behaviour	Employees who advocate a policy they doubt, brand ambassadors
Free choice	Brehm (1956)	Giving up the attractive features of a rejected option	Chosen option rated higher, rejected option lower	Post-purchase evaluation of close alternatives
Effort justification	Aronson & Mills (1959)	Suffering or effort spent on an outcome that disappoints	Outcome is valued more	Demanding loyalty tiers, self-assembly, hard-won memberships
Belief disconfirmation	Festinger (1957)	Evidence that contradicts a strongly held belief	Reinterpretation, seeking social support	Investors and managers explaining away bad news
Hypocrisy	Stone et al. (1994)	Being reminded that one does not practise what one preaches	Behaviour changes to match the advocated position	Pledge-based sustainability and safety campaigns

The **hypocrisy paradigm** deserves special mention because it reverses the usual direction of change. Stone et al. (1994) had students make a persuasive message advocating condom use and then reminded them of their own past failures to use condoms. Facing a gap between what they preached and what they practised, participants were more likely to buy condoms afterwards.

When the attitude is desirable and public, people may reduce dissonance by changing *behaviour*.

21.1.3 Mechanism

Figure 21.1 summarizes the process. An inconsistency between behaviour and attitude produces dissonance only when moderating conditions are met: the behaviour must feel freely chosen, the consequences foreseeable and unwanted, and the issue important to the self. Dissonance is experienced as psychological discomfort (Elliot and Devine 1994) and as arousal. In a classic demonstration, Zanna and Cooper (1974) gave participants a placebo pill and told some of them that it would make them feel tense. Those who could blame their tension on the pill showed no attitude change after writing a counterattitudinal essay, which suggests that attitude change depends on interpreting arousal as caused by one's own inconsistent behaviour.

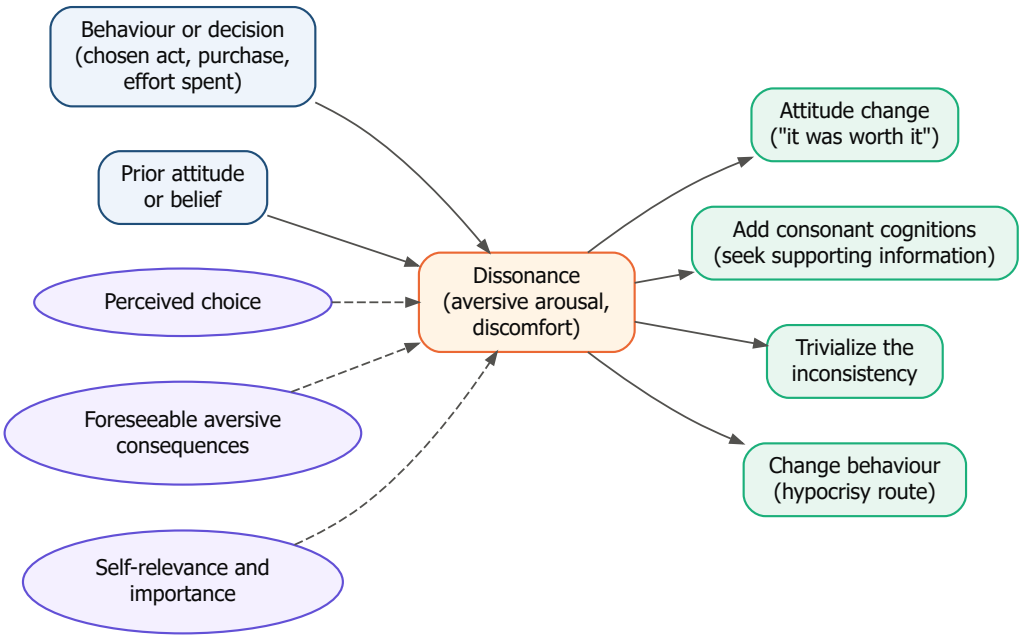


Figure 21.1. Mechanism of cognitive dissonance. An inconsistency between behaviour and attitude produces aversive arousal when the act is freely chosen and consequential. People then reduce dissonance through the least resistant route. Synthesized from Festinger (1957), Cooper and Fazio (1984), and Elliot and Devine (1994).

The induced-compliance prediction is easiest to see in contrast to a simple reinforcement account. Figure 21.2 plots the stylised predictions. Reinforcement theory expects more attitude change as the reward for advocacy grows. Dissonance theory expects the opposite for freely chosen acts, because a large reward supplies its own justification. The dissonance prediction applies only under high choice. When people are simply instructed to comply, they bear no responsibility and attitude change is small at any payment level.

21.1.4 Evidence and extensions

Dissonance theory generated hundreds of experiments and several important revisions (Harmon-Jones and Mills 1999). Three refinements stand out.

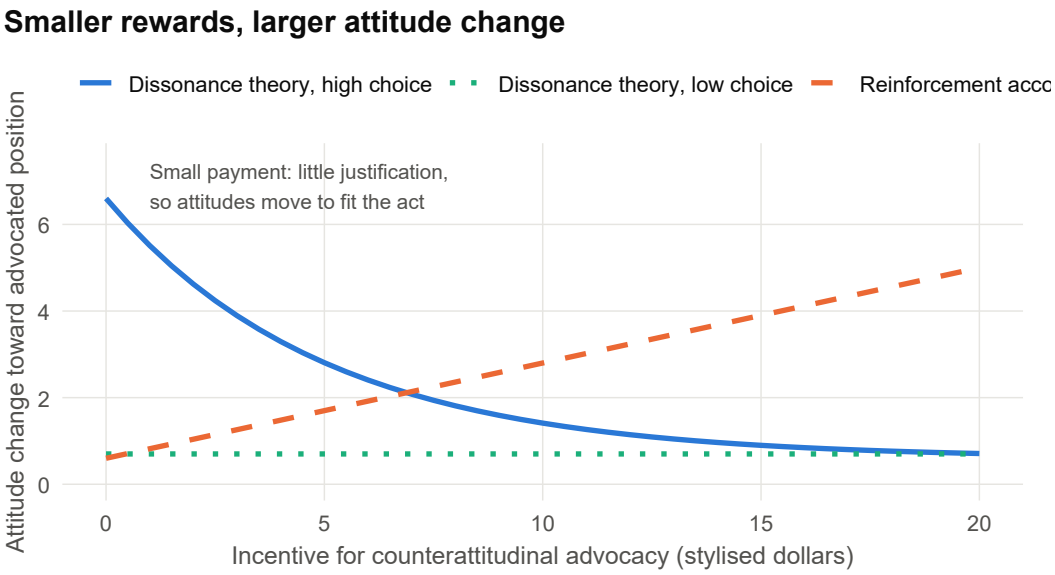


Figure 21.2. Stylised predictions for the induced-compliance paradigm. Under high perceived choice, dissonance theory predicts that attitude change toward a counterattitudinal position falls as the incentive for advocacy rises. Under low choice, attitude change is small at every incentive level. A reinforcement account predicts the opposite slope. Illustrative curves, not empirical data.

- **The "new look".** Cooper and Fazio (1984) argued that inconsistency per se matters less than responsibility for producing a foreseeable aversive consequence. On this view, a lie that harms no one produces little dissonance.
- **Self-affirmation.** Steele (1988) proposed that dissonance threatens the global integrity of the self. People who have affirmed an important value in an unrelated domain show less attitude change, as though their self-worth has been restored by other means.
- **The action-based model.** Harmon-Jones, Harmon-Jones, and Levy (2015) argue that dissonance exists because inconsistent cognitions interfere with effective action. Reducing dissonance after a decision helps people commit to and carry out the chosen course of action.

Choice-induced preference change has also been observed in young children and in capuchin monkeys (Egan, Santos, and Bloom 2007), which suggests that the tendency to devalue rejected options does not require sophisticated self-concepts or cultural norms about consistency.

DEBATE AND EVIDENCE

How robust is the evidence? Two lines of critique have reshaped the field. First, M. K. Chen and Risen (2010) showed that the classic free-choice paradigm contains a statistical artifact. Because choices reveal information about preferences that initial ratings measure with error, the chosen item will appear to rise and the rejected item to fall even if no attitude changes at all. Valid tests require control conditions in which ratings are collected before and after a choice that participants have not yet made, or designs that hold the informational value of the choice constant. Figure 21.3 illustrates the problem.

Second, a large multilab replication of the induced-compliance paradigm across 39 laboratories in 19 countries (N = 4,898) did *not* find the core predicted difference between high-choice

and low-choice conditions, although writing a counterattitudinal essay did shift attitudes relative to writing a neutral essay (Vaidis et al. 2024). The authors concluded that the paradigm, at least in that form, does not provide robust evidence for dissonance. Other paradigms have received less large-scale scrutiny. A fair summary is that the broad idea that people rationalize their choices is well supported, while specific classic effects are less robust and smaller than textbooks long implied.

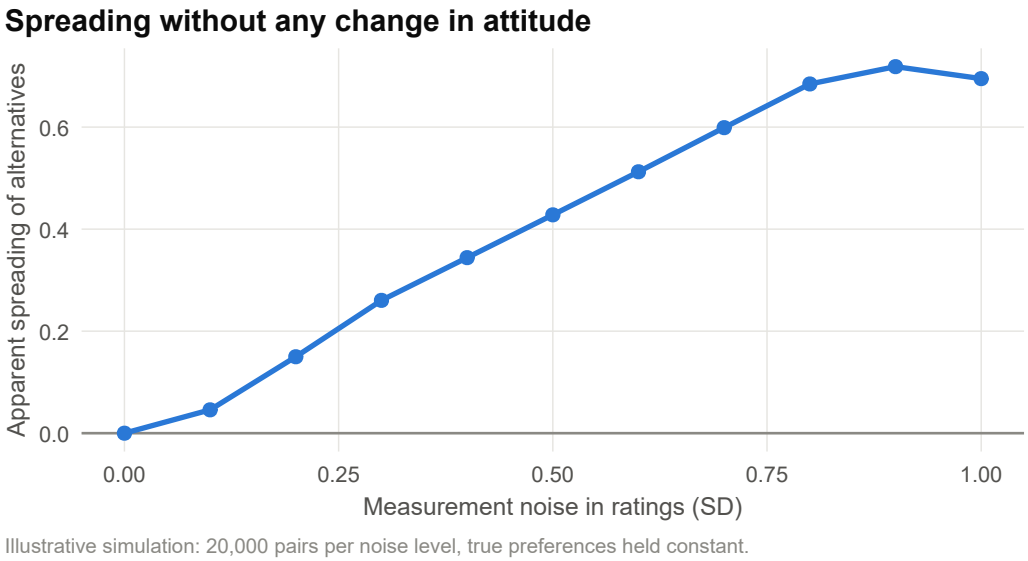


Figure 21.3. Illustrative simulation of the Chen and Risen (2010) artifact. Preferences are fixed and never change. Ratings measure them with noise, and people choose between items that received similar first ratings. The apparent 'spread' between chosen and rejected items grows with measurement noise even though no attitude has changed.

21.1.5 Applications in business

Marketing research adopted dissonance theory early, especially to explain **post-purchase dissonance**: the doubt that follows an important, costly, and irreversible purchase among close alternatives (Cummings and Venkatesan 1976). Sweeney, Hausknecht, and Soutar (2000) developed a scale that separates an emotional component (psychological discomfort after purchase) from cognitive components concerning the wisdom of the purchase and concern over whether one got a good deal. Table 21.3 maps common reduction strategies to managerial levers.

Table 21.3. Post-purchase dissonance reduction and managerial levers.

Reduction strategy	Consumer example	Managerial lever	Caution
Bolster the chosen option	Reading positive reviews of the car just bought	Welcome messages, owner communities, usage tips	Reassurance must be credible or it signals doubt
Derogate the rejected option	Recalling the rival's weaknesses	Comparative information at the point of sale	Overt attacks on rivals can trigger reactance
Increase perceived effort or commitment	Valuing a hard-won status tier	Earned loyalty tiers, onboarding rituals	Effort that feels pointless breeds resentment
Trivialize the decision	"It was only a small upgrade"	Low-risk trials, easy downgrade options	May reduce engagement with the product

Table 21.3. Post-purchase dissonance reduction and managerial levers. *(continued)*

Reduction strategy	Consumer example	Managerial lever	Caution
Reverse the decision	Returning the product	Generous return policies reduce anticipated regret	Returns are costly if dissonance is not addressed

BUSINESS CASE

Post-purchase reassurance. Many durable-goods and financial-services firms invest in communication that begins *after* the sale: welcome packs, onboarding calls, owner magazines, and first-month check-ins. Dissonance theory explains why. A customer who has just committed a large sum to one of several similar options is motivated to confirm the wisdom of the choice and is unusually receptive to supporting information (selective exposure). Reassurance that arrives in this window lowers returns, cancellations, and negative word of mouth. The same logic explains regulatory cooling-off periods for high-pressure sales contexts: they allow buyers to reverse decisions taken before dissonance reduction has set in.

Illustrative scenario. A subscription mattress retailer notices that most returns occur in the third and fourth weeks. It tests a sequence of messages that explains the adjustment period, shares sleep-quality tips, and invites the customer to rate specific features. If dissonance is the mechanism, the messages should reduce returns most among customers who chose between two close models and least among customers who faced no real alternative.

BUSINESS CASE

Effort justification and loyalty. Programs that make customers *work* for status, by accumulating qualifying activity, completing challenges, or assembling products, can raise valuation of the reward. The IKEA effect (Section 29.7) shows that people value products they assembled themselves more than identical pre-assembled products (Norton, Mochon, and Ariely 2012), a finding consistent with effort justification. The boundary condition matters for design: effort raises valuation only when the task is completed and the outcome is attributed to one's own choice. Effort that feels imposed, pointless, or that ends in failure produces frustration rather than commitment. Dissonance also helps explain escalation of commitment (Section 14.2): managers who chose a failing project may justify their choice by investing more rather than admitting error.

21.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Customers who choose between two close alternatives will show greater post-purchase selective exposure to brand communication than customers who choose a clearly dominant option.
- *H2.* The effect of a post-purchase reassurance message on product returns will be stronger for high-price, low-reversibility purchases.
- *H3.* Employees who publicly advocate a new policy under high perceived choice will report more favourable attitudes toward the policy than employees instructed to advocate

it, and the difference will be smaller when a self-affirmation opportunity is provided.
Designs. Preregistered experiments with adequate power. Free-choice designs with Chen and Risen controls. Field experiments on post-purchase communication with returns or cancellations as outcomes. Panel data linking choice difficulty to later satisfaction and repurchase.
Measures. Attitude before and after the act, perceived choice (manipulation check), dissonance discomfort (Elliot and Devine 1994), post-purchase dissonance scale (Sweeney, Hausknecht, and Soutar 2000), information-search logs.

COMMON PITFALL

Calling every rationalization "dissonance". Many studies infer dissonance from the outcome (attitudes moved toward behaviour) without measuring the mediating discomfort or manipulating its moderators. The same outcome can arise from self-perception, from new information, or from a statistical artifact. Test the moderators that are unique to dissonance (choice, foreseeable consequences, misattribution of arousal) and measure discomfort directly.

21.2 Self-Perception Theory

THEORY CARD

Core claim. When internal cues are weak or ambiguous, people infer their own attitudes and feelings by observing their behaviour and the circumstances in which it occurs, much as an outside observer would.
Origins. Social psychology, with roots in radical behaviourism. Bem (1967) and Bem (1972).
Level of analysis. Individual.
Key constructs. Behavioural cues, situational cues (external justification), internal cues, inference, overjustification, commitment.
Boundary conditions. Strongest when prior attitudes are weak, ambiguous, or absent, and when behaviour is close to existing attitudes (within the latitude of acceptance). Weaker when people hold strong, accessible attitudes.
Related theories. Cognitive dissonance, Attribution theory, Self-determination theory (Section 19.6), Social information processing (Section 14.10).

21.2.1 Core logic

Bem (1967) offered a radical alternative to dissonance theory. He argued that the participants in Festinger and Carlsmith's study did not need to experience any aversive tension. They simply asked themselves what their behaviour implied about their attitudes, just as an observer would. A person who says the task was interesting for twenty dollars was obviously doing it for the money. A person who says the same thing for one dollar must, on reflection, have found the task somewhat interesting. To support this view, Bem ran **interpersonal simulations**: observers who merely read a description of the procedure predicted the same pattern of attitudes that actual participants had reported.

Bem (1972) generalized the argument into a theory of self-knowledge. People come to know their

own attitudes, emotions, and other internal states partly by inferring them from their overt behaviour and its context, to the extent that internal cues are weak, ambiguous, or uninterpretable. The theory therefore applies the logic of attribution to the self. When behaviour occurs without obvious external pressure, people attribute it to their own dispositions. When strong external reasons are present, they discount internal causes.

21.2.2 Mechanism

Figure 21.4 contrasts the two routes. In self-perception, attitude change is a cool inference without arousal. In dissonance, it is a hot process driven by discomfort. Fazio, Zanna, and Cooper (1977) proposed a resolution that has stood the test of time: dissonance governs behaviour that falls clearly outside one's existing attitudes (the latitude of rejection), whereas self-perception governs behaviour that is only mildly discrepant or attitude-congruent (the latitude of acceptance).

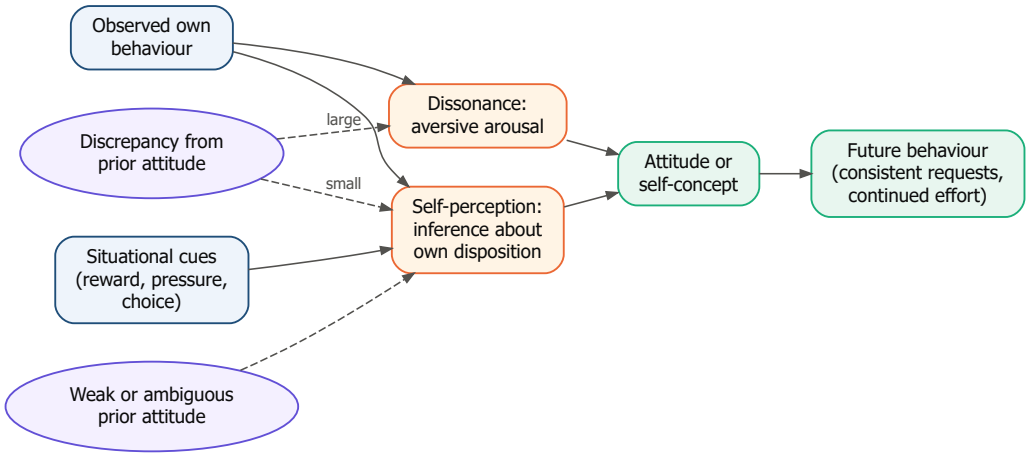


Figure 21.4. Self-perception and dissonance as complementary routes from behaviour to attitude. Self-perception is an inference that requires no arousal and dominates when behaviour is close to existing attitudes. Dissonance involves aversive arousal and dominates when behaviour is clearly counterattitudinal (after Fazio, Zanna, and Cooper, 1977).

Table 21.4 summarises how the two theories differ.

Table 21.4. Cognitive dissonance versus self-perception.

Feature	Cognitive dissonance	Self-perception
Nature of process	Motivational (aversive arousal)	Inferential (cool observation)
Role of prior attitude	Strong attitude creates the inconsistency	Weak or absent attitude leaves room for inference
Domain of application	Clearly counterattitudinal behaviour	Attitude-congruent or mildly discrepant behaviour
Diagnostic evidence	Arousal, discomfort, misattribution effects	Observers reproduce actors' inferences
Key business uses	Post-purchase reassurance, effort justification	Foot-in-the-door, rewards design, small commitments
Classic sources	Festinger (1957), Cooper & Fazio (1984)	Bem (1967, 1972), Fazio et al. (1977)

21.2.3 Evidence and extensions

Two research streams carry the theory's most practical predictions.

The foot-in-the-door effect. Freedman and Fraser (1966) asked homeowners to agree to a small request, such as signing a petition or displaying a small sign about safe driving. About two weeks later, a different person asked them to place a large, unattractive "Drive Carefully" sign on their lawn. Homeowners who had complied with the small request were far more likely to agree to the large one than homeowners approached only with the large request. The self-perception explanation is that complying with the first request changes how people see themselves ("I am the kind of person who supports good causes"), which makes the later request consistent with that self-view.

Overjustification. If people infer their attitudes from behaviour and its causes, then a salient external reward for an activity that was already enjoyable should lead them to attribute their behaviour to the reward and lower their intrinsic interest. Lepper, Greene, and Nisbett (1973) found that preschool children who expected and received a reward for drawing later spent less free time drawing than children who were not rewarded. A meta-analysis confirmed that expected, tangible, performance-contingent rewards tend to undermine free-choice intrinsic motivation for interesting tasks (Deci, Koestner, and Ryan 1999). This finding is developed in the chapter on self-determination theory (Section 19.6).

DEBATE AND EVIDENCE

How large is the foot-in-the-door effect? Meta-analyses show that the effect is real but modest and highly conditional. Beaman et al. (1983) found an average effect that was small and varied widely across studies. Burger (1999) reviewed the literature through a multiple-process lens and concluded that the effect is most reliable when the first request is actually performed (not merely agreed to), when it is prosocial, when people are given a self-relevant label, and when the second request is a natural continuation of the first. Other processes, such as reciprocity or the desire to appear consistent, can work against the effect, which explains the many null findings. Practitioners who cite "foot-in-the-door" as a universal law overstate the evidence.

21.2.4 Applications in business

BUSINESS CASE

Foot-in-the-door in sales and fundraising. Charities often start with a low-cost action, such as signing a petition, sharing a post, or answering a short survey, before asking for a donation. Software firms use the same logic in onboarding: small, completed actions (importing a contact list, inviting one colleague) precede requests for larger commitments such as a paid upgrade. Self-perception theory predicts that these sequences work best when the first step is freely chosen, actually performed, and framed in identity terms ("you are a supporter", "you are a power user"). It also predicts that paying people to take the first step can backfire, because the payment supplies an external explanation for the behaviour.

Illustrative scenario. A wildlife charity compares three email sequences: (a) an immediate donation request, (b) a petition followed one week later by a donation request, and (c) the same petition with a thank-you message that labels signers as "conservation supporters". Self-perception theory predicts (c) > (b) > (a) in donation rates, with the largest gap among people

who had no prior relationship with the charity, because their attitudes are weakest.

21.2.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Users who complete a small, voluntary in-app action will be more likely to accept a later premium offer than users who are shown the offer directly, and the effect will be stronger for users with no prior brand experience.
- *H2.* Introducing a cash reward for writing product reviews will reduce the rate of unrewarded reviews after the reward ends.

Designs. Sequential-request field experiments, A/B tests of onboarding flows, interpersonal-simulation studies, and experiments that cross behaviour with salient external justification.
Measures. Compliance with target requests, self-reported attitudes and self-labels, free-choice persistence after rewards end, and measures of attitude strength (certainty, accessibility).

COMMON PITFALL

Ignoring attitude strength. Self-perception theory applies mainly when people do not already know what they think. Applying it to customers or employees with strong, well-formed attitudes predicts effects that will not appear. Measure prior attitude strength and test it as a moderator.

21.3 Attribution Theory

THEORY CARD

Core claim. People act as intuitive scientists who infer the causes of behaviour and outcomes. The causes they infer, internal or external, stable or unstable, controllable or uncontrollable, determine their emotions, expectations, and actions toward others.
Origins. Social psychology. Heider (1958) founded the field. Kelley (1967) and Kelley (1973) formalized the covariation principle. Weiner (1985) developed the dimensional model of attribution, emotion, and motivation. L. Ross (1977) named the fundamental attribution error. Folkes (1984) and Folkes (1988) brought the theory into consumer research.
Level of analysis. Individual perceiver, applied to dyads (customer-firm, manager-employee, investor-executive).
Key constructs. Internal (dispositional) and external (situational) causes, consensus, distinctiveness, consistency, discounting and augmenting, locus, stability, controllability, correspondence bias, self-serving bias.
Boundary conditions. Causal analysis is most active after unexpected, negative, or important events. Routine, positive outcomes often trigger little causal search. Culture shapes the relative weight placed on persons versus situations.
Related theories. Self-perception theory, Organizational justice (Section 15.9), Expectancy disconfirmation (Section 27.1), Sensemaking (Section 14.11), Self-construal (Section 18.4).

21.3.1 Origins and intellectual history

Heider (1958) argued that ordinary people practise a "naive psychology": they try to understand the social world by identifying the stable dispositions that lie behind changing behaviour. His key distinction was between **personal** causes (ability, effort, intention) and **impersonal** or environmental causes (task difficulty, luck, situational pressure). How a person explains an event matters because it determines what they expect to happen next and whom they hold responsible.

Two research programs built on Heider. Kelley (1967) and Kelley (1973) asked *how* people arrive at causal conclusions and proposed that they reason like analysts of variance. Weiner asked *what follows* from causal conclusions and showed that the dimensions of a cause shape emotion and motivation (Weiner 1985). A third line, beginning with E. E. Jones and Harris (1967) and synthesized by L. Ross (1977), documented systematic biases in how people actually make attributions.

21.3.2 Core logic: Kelley's covariation model

Kelley's **covariation principle** holds that people attribute an effect to the cause with which it covaries over time. Three kinds of information matter when someone explains why an actor responded to a stimulus in a particular way (for example, why a customer complained about a restaurant):

- **Consensus.** Do other people respond to the same stimulus in the same way?
- **Distinctiveness.** Does the actor respond this way only to this stimulus, or to many stimuli?
- **Consistency.** Does the actor respond this way to this stimulus across time and situations?

Table 21.5 shows the three combinations that receive the clearest attributions.

Table 21.5. Kelley's covariation model: three information patterns and the attributions they support.

Consensus	Distinctiveness	Consistency	Likely attribution	Example: a customer complains about a restaurant
High	High	High	Stimulus (external: the entity)	Everyone complains about it. She rarely complains elsewhere. She complains every visit. The restaurant is the problem.
Low	Low	High	Person (internal: the actor)	No one else complains. She complains about every restaurant. She always complains. She is a chronic complainer.
Low	High	Low	Circumstance (a particular occasion)	No one else complains. She rarely complains elsewhere. She has not complained here before. Something unusual happened that night.

When people have only a single observation, Kelley proposed that they rely on **causal schemata**, including two principles that matter in business. Under the **discounting principle**, the role of a given cause is discounted if other plausible causes are present. A glowing review from a paid influencer is discounted because payment is an alternative explanation. Under the **augmenting principle**, a cause is judged stronger if the effect occurred despite an inhibiting factor. An employee who delivers results while understaffed is credited with more ability.

21.3.3 Mechanism: Weiner's dimensional model

Weiner (1985) showed that causes can be classified along three dimensions, each linked to distinct psychological consequences:

- **Locus** (internal or external to the actor) affects self-esteem and pride.
- **Stability** (permanent or temporary) affects expectations of future outcomes, and therefore hope or hopelessness.
- **Controllability** (could the actor have prevented it?) affects responsibility judgements and other-directed emotions: anger when another party could have controlled a negative outcome, sympathy when it could not, and guilt when the self could have.

Folkes (1984) applied this framework to product failure. Consumers who attributed a failure to the firm (locus) felt more entitled to a refund and an apology. Those who saw the cause as stable expected future failures and preferred refunds to exchanges. Those who saw the cause as controllable by the firm felt more anger and more desire to hurt the firm's business. A field study of delayed airline passengers found the same pattern: delays seen as controllable by the airline produced more anger and lower intentions to fly the airline again (Folkes, Koletsky, and Graham 1987). Figure 21.5 summarises the chain.

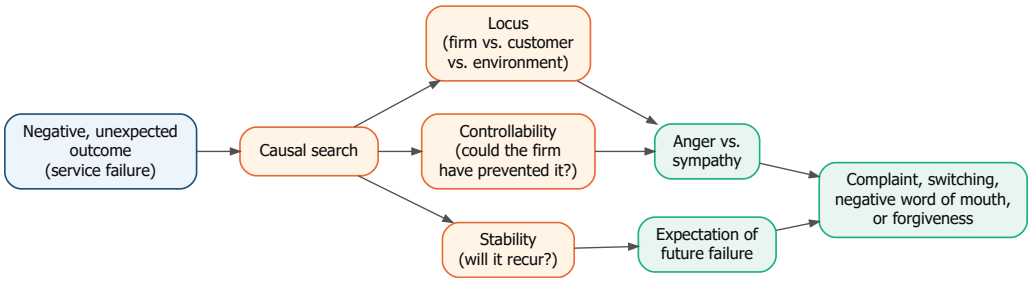


Figure 21.5. Weiner's attributional chain applied to a service failure. A negative and unexpected outcome triggers causal search. The perceived locus, stability, and controllability of the cause shape expectations and emotions, which in turn drive complaint, switching, word of mouth, and forgiveness (after Weiner, 1985, Folkes, 1984, 1988).

Figure 21.6 illustrates the managerial implication with simulated data. The same missed connection produces very different repurchase intentions depending on whether customers believe the firm could have prevented it, and an apology with compensation narrows, but does not close, the gap.

21.3.4 Evidence and extensions: biases in attribution

Kelley's model describes how a careful analyst *could* reason. Real perceivers take shortcuts. In a classic study, E. E. Jones and Harris (1967) asked participants to read essays that either supported or opposed Fidel Castro's government in Cuba. Even when participants were told that the writer had been *assigned* the position, they inferred that the writer's true attitude corresponded to the essay. L. Ross (1977) called the general tendency to underestimate situational influences and overestimate dispositional ones the **fundamental attribution error**. Gilbert and Malone (1995) renamed it the **correspondence bias** and identified four causes: lack of awareness of situational constraints, unrealistic expectations about how situations affect behaviour, inflated categorizations of behaviour, and incomplete correction of an initial dispositional inference.

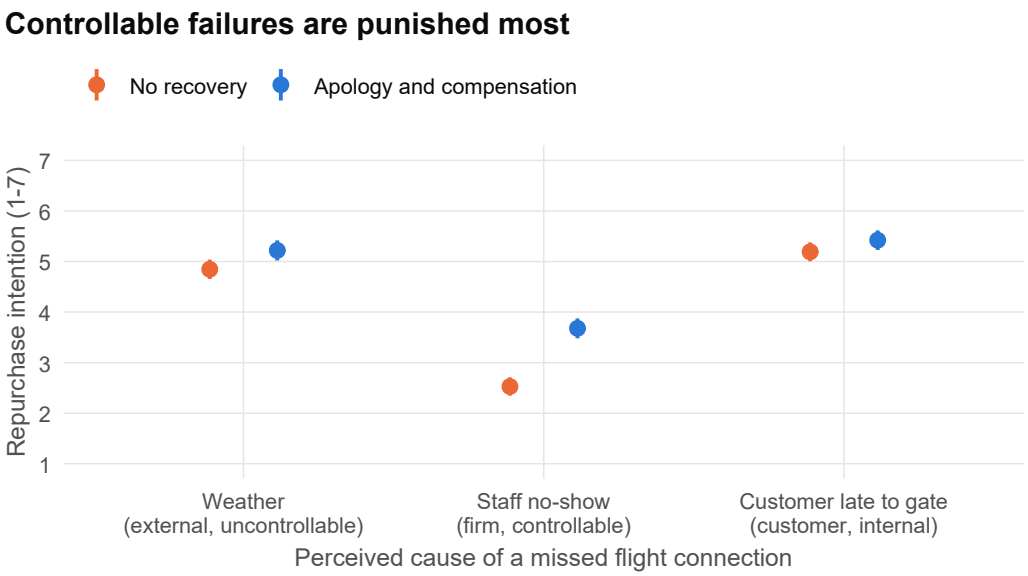


Figure 21.6. Illustrative simulated experiment on service-failure attributions. Repurchase intention (1-7 scale) is lowest when customers attribute a missed connection to a cause the firm controlled, and recovery effort helps most in that condition. Simulated data (n = 150 per cell), with a qualitative pattern consistent with Folkes (1984) and Folkes, Koletsky, and Graham (1987). Bars show 95 percent confidence intervals.

The **self-serving bias** is a second well-documented pattern: people attribute success to themselves and failure to circumstances. Organizations do the same. Studies of letters to shareholders found that firms tended to credit good results to internal factors such as strategy and management and to blame poor results on external factors such as the economy or regulation (Bettman and Weitz 1983, Staw, McKechnie, and Puffer 1983).

DEBATE AND EVIDENCE

Universal or cultural? Robust or fragile? Three critiques qualify the classic picture. First, the actor-observer asymmetry, the claim that people explain their own behaviour situationally but others' dispositionally, is much weaker than once believed. A meta-analysis found an average asymmetry close to zero, with meaningful effects only under specific conditions (Malle 2006). Second, the strength of the correspondence bias varies across cultures. Morris and Peng (1994) found that American participants explained the same events more in terms of individual dispositions than Chinese participants, who gave more weight to situational and contextual factors (see Self-construal (Section 18.4)). Third, the Jones and Harris effect itself has proved robust in many replications, but its magnitude depends on how diagnostic the behaviour appears and on whether perceivers are motivated and able to correct their first impressions (Gilbert and Malone 1995). The covariation model is better regarded as a normative benchmark than as a description of everyday reasoning.

21.3.5 Applications in business

Folkes (1988) and Weiner (2000) show that attribution theory explains satisfaction, complaining, word of mouth (Section 26.3), product choice, and responses to endorsers. In services, the most

important application is **service recovery**. Customers evaluate a recovery in light of who caused the failure and whether it could have been prevented. Compensation, speed, and apology all matter, but their effects depend on attributions (A. K. Smith, Bolton, and Wagner 1999). A recovery that ignores blame (for example, offering a voucher while insisting nothing went wrong) may leave the anger produced by a controllable failure intact.

BUSINESS CASE

JetBlue and the February 2007 ice storm. In February 2007 an ice storm on the US East Coast led JetBlue to cancel a large share of its flights, and some passengers were left on aircraft on the tarmac at New York's JFK airport for many hours. The storm was external and uncontrollable, but the length of the delays and the breakdown in rebooking were widely blamed on the airline's own operations. The company's response was framed in attributional terms: the founder and chief executive, David Neeleman, apologized publicly, accepted responsibility rather than blaming the weather, and the airline introduced a **Customer Bill of Rights** that promised specified compensation for future delays and cancellations. Attribution theory explains the logic. Accepting responsibility for the controllable part of the failure addresses anger directly, and a public commitment to compensation signals that the cause is being made unstable (it will not recur).

BUSINESS CASE

Crisis communication and the Tylenol recall. In 1982, several people in the Chicago area died after taking Extra-Strength Tylenol capsules that had been laced with cyanide. Johnson & Johnson was a victim of criminal tampering, an external and uncontrollable cause, yet it withdrew the product nationally, cooperated with authorities, and reintroduced Tylenol in tamper-resistant packaging. The case is a standard example in crisis communication because it matches the prescriptions of **situational crisis communication theory** (Coombs 2007), which builds directly on attribution theory. Stakeholders attribute crisis responsibility to an organization based on the type of crisis (victim, accidental, or preventable), and organizations should match their response to that attribution. Victim crises call for instruction and care, preventable crises for full apology and corrective action. Firms that deny responsibility for a crisis that stakeholders see as preventable tend to amplify reputational damage.

21.3.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The negative effect of a service failure on repurchase intention will be stronger when customers perceive the cause as controllable by the firm, and an apology will attenuate this effect more than compensation alone.
- *H2.* Firms will use more external attributions in letters to shareholders after poor performance than after good performance, and this tendency will be stronger when the CEO is new.
- *H3.* Consumers will discount positive product reviews when the reviewer is known to have been compensated, and the discount will be smaller when the review mentions a product weakness (augmenting).

Designs. Scenario and field experiments manipulating cause information. Text analysis of complaints, reviews, earnings calls, and annual reports using attribution dictionaries. Panel data linking attributions to churn.
Measures. Perceived locus, stability, and controllability (multi-item scales derived from Weiner's dimensions), anger, sympathy, blame, repurchase intention, complaint behaviour, word of mouth.

COMMON PITFALL

Assuming customers share the firm's causal model. Managers often explain failures in terms of root causes they know (a supplier outage, an IT migration) and assume customers will see the cause as external. Customers attribute based on what they see and on covariation they can observe. If competitors did not fail, consensus is low and the firm will be blamed. Measure customers' attributions directly rather than inferring them from the operational cause.

21.4 Balance Theory

THEORY CARD

Core claim. People prefer consistent configurations of sentiment among themselves, another person, and an object. A triad is balanced when the product of its three relations is positive. Unbalanced triads create tension that people resolve by changing one relation.
Origins. Social psychology. Heider (1946), extended to signed networks by Cartwright and Harary (1956).
Level of analysis. Triad (person P, other O, object X). In network form, any group of three or more actors.
Key constructs. Sentiment relations (liking, disliking), unit relations (belonging, association), balanced and unbalanced triads, pressure toward balance, structural balance.
Boundary conditions. Balance pressures are strongest when the relations are salient and important to the perceiver. They are weaker for loose associations and when people tolerate ambivalence.
Related theories. Cognitive dissonance, Homophily (Section 30.2), Structural holes (Section 9.4), Social identity theory (Section 22.2).

21.4.1 Core logic

Heider (1946) analysed the cognitive field of a person (P) who perceives another person (O) and an object or issue (X). Each relation is either positive (liking, approving, belonging together) or negative (disliking, rejecting). A triad is **balanced** when all three relations are positive or when exactly two are negative. "My friend likes what I like" is balanced. So is "my enemy likes what I dislike". A triad with exactly one negative relation is unbalanced: "I like the endorser, the endorser likes the brand, but I dislike the brand". The theory predicts that people experience strain in unbalanced configurations and restore balance by changing whichever relation is weakest. Heider treated the all-negative triad as ambiguous. Cartwright and Harary (1956) formalized balance with signed graphs, in which a cycle is balanced if the product of its signs is positive. This made the all-negative

triad unbalanced and allowed balance to be assessed in networks of any size. Figure 21.7 shows the four configurations.

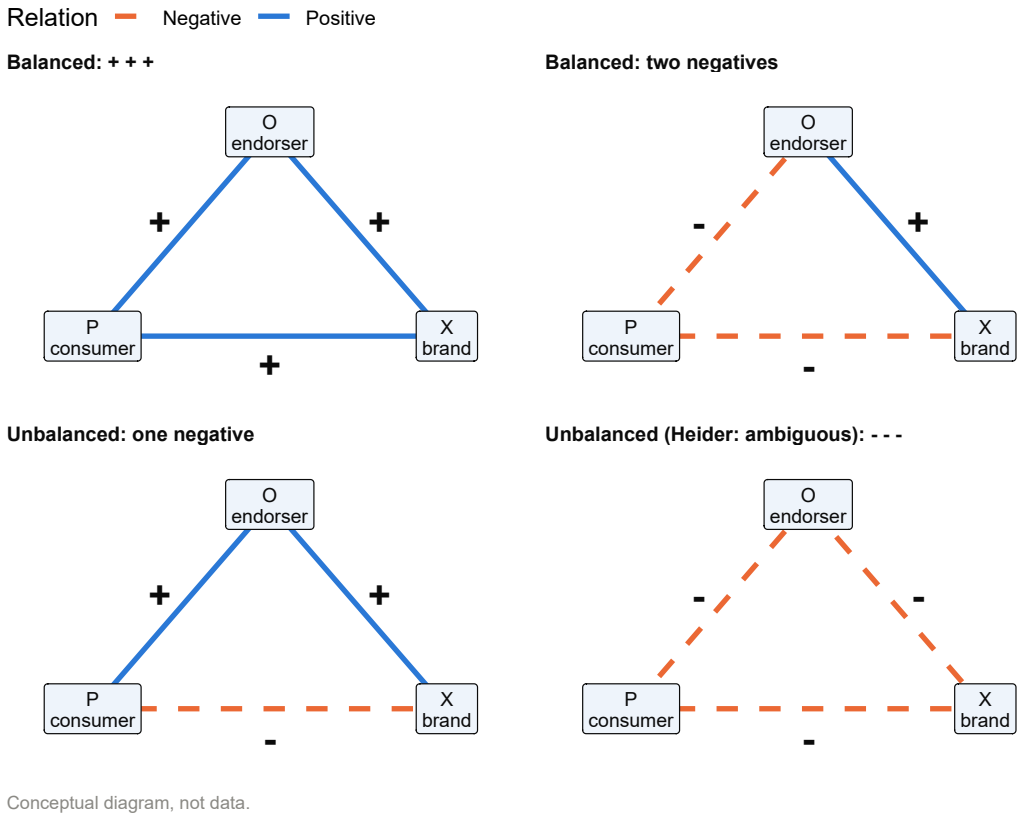


Figure 21.7. Heider's P-O-X triads applied to celebrity endorsement. P is the consumer, O the endorser, and X the brand. Solid blue lines are positive relations, dashed orange lines negative relations. A triad is balanced when the product of the three signs is positive. Heider treated the all-negative triad as ambiguous. Cartwright and Harary (1956) classify it as unbalanced.

21.4.2 Evidence and extensions

Balance theory has two lives. As a theory of individual cognition, it is a member of the family of **cognitive consistency** theories, alongside dissonance theory, and its predictions about preference for balanced configurations are well supported in judgement tasks. As a theory of social structure, generalized by Cartwright and Harary (1956), it predicts that groups of friends and enemies will tend to sort into internally positive and mutually negative clusters. Large-scale studies of signed online networks (in which users mark others as trusted or distrusted, or vote for or against each other) found that balance explains some patterns of positive and negative ties, but that a status-based account fits directed relationships better (Leskovec, Huttenlocher, and Kleinberg 2010). Balance is therefore one tendency among several, not a law of social structure.

21.4.3 Applications in business

Celebrity endorsement is the most direct application. The endorser (O) is chosen because consumers (P) like the endorser. The endorsement creates a positive unit relation between O and the brand (X). Balance pressures then push consumers toward liking the brand. McCracken (1989)

offered a complementary account in which endorsers transfer cultural meanings to products, but the balance logic predicts a specific risk: if the endorser becomes disliked, the same unit relation now pushes the brand down. Experiments show that negative information about an endorser can lower evaluations of the endorsed brand under some conditions, particularly when the association between them is strong (Till and Shimp 1998).

BUSINESS CASE

Endorser scandals and the balance dilemma. When the golfer Tiger Woods became the subject of a widely reported personal scandal in late 2009, several sponsors, including the consulting firm Accenture, ended their relationships with him within weeks, while Nike kept him under contract. Balance theory frames the decision. For a brand whose positioning depended on the endorser's personal integrity, a newly negative P-O relation threatened to turn the P-O-X triad unbalanced and pull brand attitudes down, so severing the O-X unit relation restored balance. For a brand associated with athletic performance, where many consumers' liking of the athlete's sporting excellence persisted, the triad could remain balanced on the relevant dimension. The theory does not say which decision was correct. It identifies what each decision bet on: which relation in the triad consumers would change.

BUSINESS CASE

Illustrative scenario: cause-related marketing. A coffee chain partners with a polarizing advocacy group. Among customers who like the group, the partnership creates a balanced triad and raises brand liking. Among customers who dislike the group, the triad becomes unbalanced, and they can restore balance either by disliking the brand or by discounting the partnership as insincere. Balance theory predicts polarization of brand attitudes, which is why firms test cause partnerships separately within customer segments rather than on average.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Negative news about an endorser will reduce brand attitudes more when the endorser has been associated with the brand for longer (stronger unit relation). *H2.* Ending an endorsement contract promptly after a scandal will attenuate the decline in brand attitudes among consumers who dislike the endorser. **Designs.** Event studies around endorser scandals, experiments that manipulate P-O and O-X relations, signed-network analysis of online trust and distrust ties. **Measures.** Attitudes toward endorser and brand, perceived association strength, signed tie data.

COMMON PITFALL

Treating the triad as closed. Real consumers hold many attitudes at once, and they can restore balance by differentiating ("I dislike his conduct but admire his golf") rather than by changing a whole relation. Measure the specific attitude dimension that the endorsement activates.

21.5 Social Cognitive Theory

THEORY CARD

Core claim. People learn not only by doing but by observing others, and they act on what they have learned when they expect it to pay off and believe they can do it. Beliefs in one's own capability, **self-efficacy**, are a central determinant of effort, persistence, and performance.
Origins. Psychology. Bandura, Ross, and Ross (1961) on observational learning, Bandura (1977) on self-efficacy, Bandura (1986) for the integrated social cognitive theory, and Bandura (1997) for the full treatment of self-efficacy.
Level of analysis. Individual in a social environment, with extensions to collective efficacy in groups.
Key constructs. Observational learning (attention, retention, production, motivation), vicarious reinforcement, self-efficacy, outcome expectancies, triadic reciprocal causation, self-regulation.
Boundary conditions. Self-efficacy predicts performance most strongly for tasks that are well defined and within the person's control. Efficacy beliefs are domain-specific: confidence in public speaking does not imply confidence in financial modelling.
Related theories. Expectancy theory (Section 19.2), Goal-setting theory (Section 19.4), Technology acceptance model (Section 10.4), Theory of planned behavior (Section 10.3), Self-control (Section 20.4).

21.5.1 Origins and intellectual history

Mid-century learning theory explained behaviour through direct reinforcement. Bandura, Ross, and Ross (1961) showed that this was incomplete. Children who watched an adult model behave aggressively toward an inflatable "Bobo" doll later imitated the specific aggressive acts, without having been reinforced for them, whereas children who watched a non-aggressive model did not. People can acquire complex behaviour simply by observing others and the consequences others receive.

Bandura (1977) added the construct that made the theory influential well beyond psychology. He distinguished **outcome expectancies** (beliefs that a behaviour will produce an outcome) from **efficacy expectancies** (beliefs that one can successfully perform the behaviour). A salesperson may be convinced that cold calling produces sales but still avoid it if she doubts her ability to do it well. Bandura (1986) integrated these ideas into **social cognitive theory**, which explains human functioning through **triadic reciprocal causation**: personal factors (cognition, affect, efficacy beliefs), behaviour, and the environment all influence one another.

21.5.2 Core logic and mechanism

Observational learning involves four subprocesses: **attention** (noticing the model's behaviour), **retention** (remembering it), **production** (being able to reproduce it), and **motivation** (expecting reward for performing it). Models are more influential when they are similar to the observer, competent, and visibly rewarded. Self-efficacy then shapes which behaviours people attempt, how much effort they invest, and how long they persist after setbacks. Figure 21.8 depicts the triadic structure and the place of self-efficacy within it.

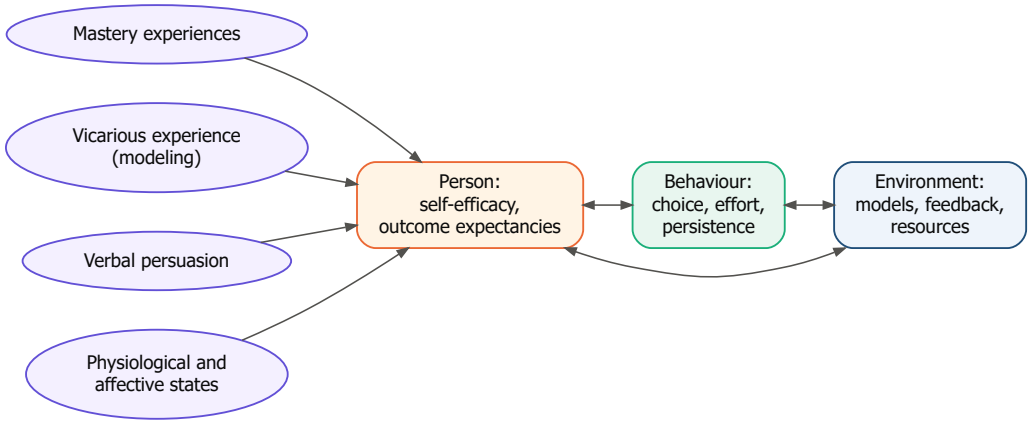


Figure 21.8. Social cognitive theory. Personal factors, behaviour, and environment influence one another (triadic reciprocal causation, Bandura, 1986). Self-efficacy beliefs are built from four sources and shape choice of activities, effort, and persistence (Bandura, 1977).

Bandura identified four sources of efficacy beliefs, summarized in Table 21.6 with their implications for training and management.

Table 21.6. Four sources of self-efficacy (Bandura, 1977, 1997) and how organizations can use them.

Source	How it builds efficacy	Relative strength	Training and management lever
Mastery experiences	Succeeding at a task, especially after effort and setbacks	Strongest	Graduated practice, early wins, simulations
Vicarious experience	Seeing similar others succeed	Moderate, strongest when the model is similar	Peer demonstrations, behaviour modeling videos, shadowing
Verbal persuasion	Credible encouragement and feedback	Weaker, easily undone by failure	Coaching that is specific and realistic
Physiological and affective states	Interpreting arousal as readiness rather than inability	Variable	Stress management, reframing nervousness

21.5.3 Evidence and extensions

The evidence for self-efficacy as a predictor of work performance is substantial. Stajkovic and Luthans (1998) meta-analysed 114 studies and found a weighted average correlation of about .38 between self-efficacy and work-related performance. The relationship was stronger for less complex tasks and in laboratory than in field settings, which is consistent with the boundary condition that efficacy matters most when performance is within the person's control. In information systems, Compeau and Higgins (1995) developed a measure of **computer self-efficacy** and showed that it shaped expectations, affect, and use of computers, a construct that now appears in many technology acceptance (Section 10.4) models.

Behaviour modeling training, which applies observational learning by showing trainees a model performing a skill, followed by practice and feedback, is one of the best-supported training methods. A meta-analysis found that its effects were largest on knowledge and skills measured at the end of training and smaller, though still positive, on job behaviour. Transfer to the job was greater when trainees saw both positive and negative models and when the work environment supported and

rewarded the new behaviour (P. J. Taylor, Russ-Eft, and Chan 2005).

DEBATE AND EVIDENCE

Can self-efficacy be too high? A debate in applied psychology asks whether the efficacy-performance correlation reflects a causal effect of efficacy or a reflection of past performance. Using within-person designs, Vancouver et al. (2002) found that higher self-efficacy could *reduce* subsequent performance under some conditions, apparently because confident people invest less effort or allocate it less carefully. Bandura and Locke (2003) replied that between-person evidence, experimental manipulations, and the full set of studies support a positive causal role, and that the negative effects arise under narrow conditions. The practical resolution is that well-calibrated efficacy helps, while inflated efficacy that is disconnected from feedback can breed complacency. The behaviour modeling literature, by contrast, rests on randomized evaluations and is among the more robust findings in training research. The original Bobo doll studies used small samples and laboratory measures of imitation, and their generalization to real-world aggression remains contested.

21.5.4 Applications in business

BUSINESS CASE

Training and self-efficacy. Social cognitive theory underpins much of modern corporate training. Sales enablement programs pair new representatives with high performers (vicarious experience), use role-play and simulations that build in early successes (mastery), and provide specific coaching (verbal persuasion). The theory also explains why training that transmits knowledge often fails to change behaviour: knowing what to do is not the same as believing one can do it under pressure.

Illustrative scenario. A contact-centre operator introduces a new complaint-handling protocol. Version A uses a slide presentation and a knowledge quiz. Version B shows video models of agents handling difficult calls (including one who makes and corrects a mistake), followed by graduated role-plays from easy to hard calls and specific feedback. Social cognitive theory predicts similar quiz scores but higher self-efficacy and better protocol adherence on live calls in Version B, with the advantage largest for new agents whose efficacy beliefs are least formed.

BUSINESS CASE

Customer self-efficacy and digital adoption. Customers also need efficacy to adopt new technologies and services. Banks and software firms that move customers to self-service channels use tutorials that show users like the customer completing tasks, guided first transactions that guarantee success, and progress indicators that make mastery visible. These are direct applications of the four sources of efficacy to customer onboarding, and they connect social cognitive theory to the facilitating conditions (Section 10.5) construct in adoption research.

21.5.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Employees who complete behaviour modeling training with mixed positive and negative models will show greater on-the-job use of a new skill than employees who see only positive models, mediated by task-specific self-efficacy.
- *H2.* The positive effect of self-efficacy on sales performance will be weaker for complex, long-cycle enterprise sales than for transactional sales.
- *H3.* Customers who complete a guided first transaction will show higher self-efficacy and greater continued use of a self-service app than customers who receive only written instructions.

Designs. Randomized training evaluations, within-person longitudinal designs that separate efficacy from prior performance, field experiments in customer onboarding.

Measures. Task-specific self-efficacy scales built following Bandura's guidance (strength and level of confidence for graded tasks), computer self-efficacy (Compeau and Higgins 1995), objective performance, transfer measures.

COMMON PITFALL

Measuring general confidence instead of task-specific efficacy. Self-efficacy is defined relative to a specific task and level of performance. Generic items such as "I am confident in my abilities" measure something closer to self-esteem and weaken predictions. Build items that refer to the actual tasks and difficulty levels in the setting studied, and measure efficacy before the performance it is meant to predict.

21.6 Reactance Theory

THEORY CARD

Core claim. When people perceive that a freedom they value is eliminated or threatened, they experience **psychological reactance**, a motivational state directed at restoring that freedom. Reactance can make a forbidden or scarce option more attractive and can push people to do the opposite of what a persuader wants.

Origins. Social psychology. J. W. Brehm (1966), S. S. Brehm and Brehm (1981). Consumer applications by Clee and Wicklund (1980). Recent review by Steindl et al. (2015).

Level of analysis. Individual.

Key constructs. Free behaviours, threat to freedom, elimination of freedom, magnitude of reactance, restoration (direct, by implication, through increased attractiveness of the threatened option), trait reactance.

Boundary conditions. People must believe they had the freedom in the first place and must value it. Reactance is greater when the threatened freedom is important, when the threat is strong or illegitimate, and when many freedoms are implicated. When people believe that control cannot be regained, reactance can give way to helplessness.

Related theories. Persuasion knowledge model (Section 26.2), Elaboration likelihood model (Section 26.1), Self-determination theory (Section 19.6), Loss aversion (Section 24.4), Choice

> architecture (Section 23.10).

21.6.1 Core logic

J. W. Brehm (1966) began from the premise that people believe they are free to engage in a set of behaviours. When an external force threatens or eliminates one of those freedoms, people are motivated to restore it. They can do so directly, by performing the threatened behaviour, or indirectly, by watching others perform it, by derogating the source of the threat, or by increasing their liking for the threatened option. S. S. Brehm and Brehm (1981) elaborated the determinants of reactance magnitude: the importance of the threatened freedom, the proportion of freedoms threatened, and the strength of the threat.

The theory explains several marketing phenomena that puzzle simpler models. Clee and Wicklund (1980) reviewed consumer applications including scarcity, high-pressure selling, and restrictions on product availability. In a well-known experiment, Worchel, Lee, and Adewole (1975) found that cookies from a jar containing two cookies were rated more desirable than identical cookies from a jar containing ten, and that the effect was strongest when scarcity resulted from other people's demand. A meta-analysis of scarcity research concluded that scarcity generally enhances the perceived value of objects (Lynn 1991). Reactance is one of several explanations for this effect, alongside quality inferences and uniqueness motives.

21.6.2 Mechanism

Contemporary research treats reactance as a measurable state. Dillard and Shen (2005) proposed and supported an **intertwined model**, in which reactance consists of anger and negative cognitions (counterarguments) that are fused together. A meta-analysis supported the intertwined model over alternatives (Rains 2013). Figure 21.9 summarizes the process.

Reactance implies that persuasive pressure has diminishing and eventually negative returns. Figure 21.10 sketches the stylised pattern. Moderate pressure informs and motivates, but beyond some point each increase in pressure adds more reactance than persuasion. Explicitly affirming the audience's freedom to decline shifts the curve upward and to the right.

21.6.3 Evidence and extensions

Reactance research has grown well beyond the laboratory (Steindl et al. 2015).

- **Recommendations and advice.** Fitzsimons and Lehmann (2004) found that unsolicited recommendations, especially those that contradicted consumers' initial impressions, could lead consumers to choose *against* the recommended option.
- **Personalization.** White et al. (2008) showed that highly personalized email solicitations could provoke reactance when the personalization was not justified by a clear benefit to the recipient, reducing response rather than increasing it.
- **Health and public communication.** Controlling, dogmatic language ("you must", "you have to") produces more reactance than autonomy-supportive language, which reduces the persuasiveness of health messages (Dillard and Shen 2005, Rains 2013).
- **Freedom affirmation.** The "but you are free" technique, adding a reminder that the target may refuse, produces a modest but reliable increase in compliance across studies (C. J. Carpenter 2013).

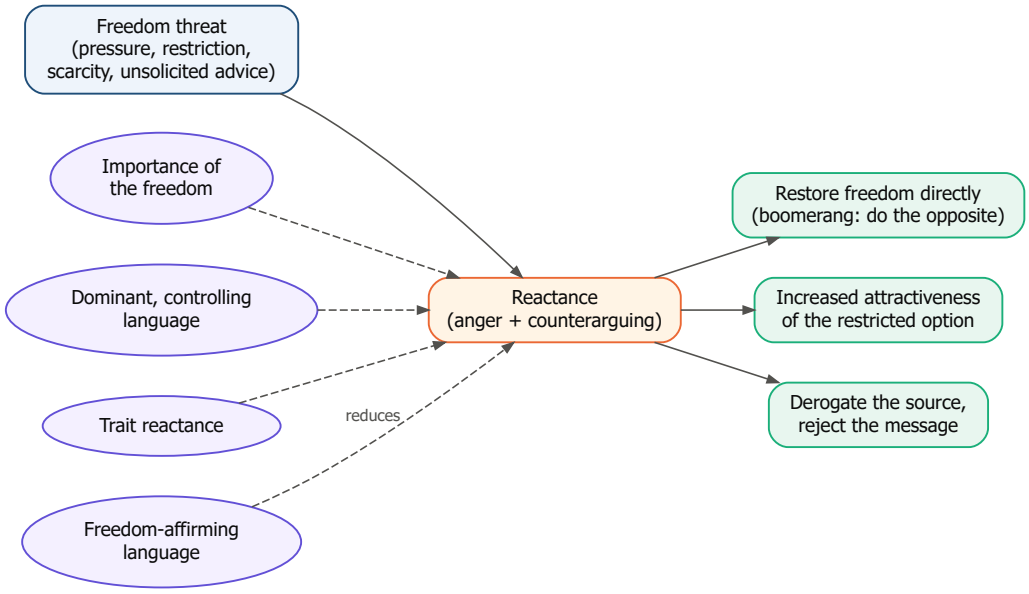


Figure 21.9. Mechanism of psychological reactance. A message or restriction that threatens a valued freedom produces reactance, measured as intertwined anger and counterarguing (Dillard and Shen, 2005). Reactance leads people to restore the freedom directly or indirectly. Freedom-affirming language and legitimate, autonomy-supportive framing reduce the threat.

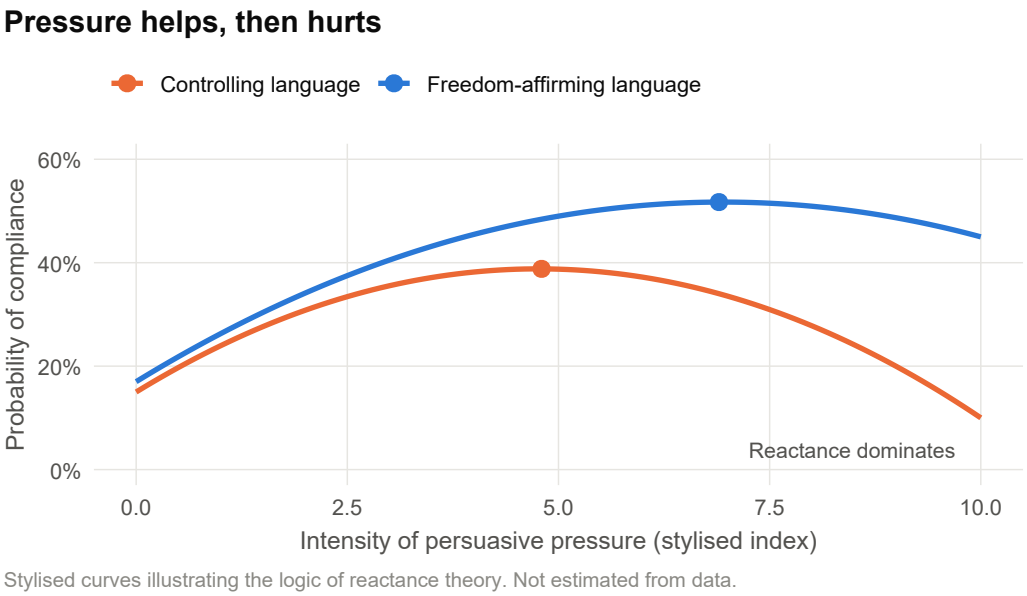


Figure 21.10. Stylised relationship between the intensity of persuasive pressure and compliance. Compliance first rises and then falls as reactance grows. Freedom-affirming language ('but you are free to decline') reduces the reactance penalty. Illustrative curves, not empirical estimates.

DEBATE AND EVIDENCE

Measurement, culture, and replication. For decades reactance was inferred from its out-

comes, which made the theory hard to falsify: almost any boomerang effect could be labelled reactance after the fact. The intertwined-model measures of anger and counterarguing have improved matters (Dillard and Shen 2005, Rains 2013), as have trait reactance scales. Steindl et al. (2015) review evidence that reactance can be experienced vicariously (when others' freedoms are threatened) and that its triggers vary across cultures, with threats to collective rather than individual freedoms more potent in some collectivistic contexts. Large preregistered replications of the classic reactance paradigms are scarce, and effect sizes in the "but you are free" literature are small and heterogeneous. The safest claim is directional: controlling pressure carries a reactance cost that persuaders systematically underestimate.

21.6.4 Applications in business

Table 21.7 maps common commercial tactics to the freedom they may threaten and to lower-reactance alternatives.

Table 21.7. Commercial tactics, the freedoms they threaten, and lower-reactance alternatives.

Tactic	Freedom threatened	Likely reactance response	Lower-reactance alternative
Countdown timers and "only 2 left"	Freedom to deliberate	Distrust, abandonment of the basket	Accurate availability information, price holds
Aggressive upselling	Freedom to choose the basic option	Choosing the cheapest option, complaint	Side-by-side comparison with a clear default
Removing a popular product	Freedom to buy the familiar option	Increased desire for the removed product, protest	Phased transition, retaining a legacy version
Unsolicited recommendations	Freedom to decide independently	Choosing against the recommendation	Recommendations on request, with reasons
Highly personal targeting	Freedom from being monitored	Ignoring or unsubscribing	Transparent explanation of why an offer is shown
Mandatory policies at work	Autonomy over how work is done	Minimal compliance, workarounds	Explain the rationale, allow choice in implementation

BUSINESS CASE

New Coke and the value of a lost freedom. In April 1985 the Coca-Cola Company replaced its flagship formula with a sweeter reformulation that had performed well in blind taste tests. The withdrawal of the original product provoked a wave of consumer protest, and within about three months the company brought the original formula back as "Coca-Cola Classic". Reactance theory offers one explanation for why taste tests misled. The tests measured preference between two *available* options, but the launch *eliminated* a familiar option, and eliminated freedoms are valued more than freely available ones. Brand attachment and identity motives also mattered, which is why the case is best read as a combination of reactance and loss aversion (Section 24.4) rather than as a pure test of either.

BUSINESS CASE

Scarcity and "you're free to choose" messaging. Online retailers and travel sites use scarcity cues and countdown timers to accelerate purchase. Genuine scarcity can raise perceived value (Lynn 1991), but cues that customers recognize as manipulative trigger reactance and erode

trust. In 2019 the UK Competition and Markets Authority secured commitments from several large hotel booking sites to stop giving false impressions of room availability and popularity, showing that regulators also treat such pressure as a consumer protection issue. Reactance theory suggests a different design: give accurate availability information, let customers hold a price or save a choice, and explicitly affirm their freedom to decide later. Such framing sacrifices some immediate conversion but reduces the reactance penalty in abandonment, trust, and repeat purchase.

21.6.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Countdown timers will increase immediate conversion but reduce repeat purchase, and the negative effect will be mediated by state reactance.
- *H2.* Adding a freedom-affirming statement to a donation request will increase donations more for recipients high in trait reactance.
- *H3.* Unsolicited product recommendations will reduce choice of the recommended item when they contradict the consumer's stated preference.

Designs. Online experiments manipulating message language and restriction, field experiments on e-commerce pages, natural experiments around product withdrawals or regulatory restrictions.

Measures. State reactance as anger and counterarguing (Dillard and Shen 2005), trait reactance scales, perceived threat to freedom, choice share of restricted options, unsubscribe and complaint rates.

COMMON PITFALL

Explaining every backfire as reactance. Reactance is easy to invoke after a campaign fails. Unless the study shows that a freedom was perceived as threatened and measures the reactance state, alternative explanations (irritation with repetition, distrust from persuasion knowledge (Section 26.2), or simply weak arguments) remain open. Manipulate the freedom threat directly and measure anger and counterarguing as mediators.

21.7 Chapter Summary

KEY TAKEAWAYS

- **Behaviour shapes attitudes.** People rationalize freely chosen, effortful, and irrevocable acts (Festinger 1957) and infer their attitudes from their behaviour when internal cues are weak (Bem 1972).
- Classic dissonance paradigms are less robust than once thought. The free-choice paradigm contains a statistical artifact (M. K. Chen and Risen 2010), and a large multilab test did not reproduce the core induced-compliance effect (Vaidis et al. 2024).
- **Attributions drive emotions and actions.** Covariation information shapes whether people blame persons or situations (Kelley 1973), and the locus, stability, and control-

lability of a cause shape anger, sympathy, and expectations (Weiner 1985).
- In service failures and crises, perceived **controllability** is the attribution that matters most for customer anger and switching (Folkes 1984, Coombs 2007).
- **Balance theory** explains endorsement and its risks: liked endorsers lift brands, and disliked endorsers drag them down (Heider 1946, Till and Shimp 1998).
- **Self-efficacy** is a robust predictor of performance, built mainly through mastery and modeling (Bandura 1977, Stajkovic and Luthans 1998).
- **Reactance** makes pressure, scarcity cues, and unsolicited advice backfire. Affirming freedom and explaining rationale reduce it (J. W. Brehm 1966, Rains 2013).

Table 21.8 presents the comparison of the chapter's theories.

Table 21.8. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Replication status	Typical business use
Cognitive dissonance	Aversive inconsistency	Small rewards and high effort produce more attitude change	Mixed: core idea supported, classic paradigms weaker	Post-purchase reassurance, loyalty design
Self-perception	Inference from own behaviour	Small commitments raise later compliance	Supported but effects modest and conditional	Onboarding sequences, fundraising, rewards design
Attribution theory	Causal inference along locus, stability, control	Controllable failures provoke anger and exit	Core dimensions robust, some biases weaker than claimed	Service recovery, crisis communication, investor relations
Balance theory	Preference for consistent triads	Attitudes move to balance P-O-X triads	Supported in judgement, partial in networks	Endorsements, sponsorships, cause marketing
Social cognitive theory	Observational learning and self-efficacy	Efficacy predicts effort and performance	Robust, with debate on causal direction	Training, sales enablement, customer onboarding
Reactance theory	Restoring threatened freedom	Pressure and restriction backfire	Directionally supported, few large replications	Scarcity messaging, personalization, change management

21.8 Discussion Questions

1. A large multilab study did not reproduce the core induced-compliance effect. Does this undermine cognitive dissonance theory as a whole, or only one paradigm? What evidence would you need to decide?
2. Design a field experiment that distinguishes a dissonance explanation from a self-perception explanation for the effect of a small first request on later donations.
3. A cloud-software outage was caused by a third-party data centre. Use Kelley's covariation model and Weiner's dimensions to predict how customers will attribute blame, and draft the first two sentences of the firm's statement.
4. Using balance theory, analyse the risks and benefits for a brand that partners with a politically outspoken athlete. Which consumers will change which relation?
5. Self-efficacy correlates with performance, but performance also builds self-efficacy. Propose a research design that isolates the causal effect of efficacy on sales performance.
6. A retailer is considering countdown timers on its product pages. Using reactance theory, spec-

ify the conditions under which the timers will help or hurt, and how you would measure the long-run effect.

21.9 Further Reading

- Festinger (1957): the original statement of dissonance theory, still readable and full of ideas that were never tested.
- Harmon-Jones and Mills (1999): an edited volume that surveys the major revisions of dissonance theory and the debates among them.
- Vaidis et al. (2024): a model of a modern multilab replication and an essential corrective to textbook accounts of induced compliance.
- Bem (1972): the fullest statement of self-perception theory.
- Kelley (1973): a concise review of the covariation and causal-schema models.
- Weiner (2000): a short essay that connects attributional theory to consumer behaviour.
- Bandura (1977): the paper that introduced self-efficacy and remains among the most cited in psychology.
- Steindl et al. (2015): an accessible review of reactance research since Brehm.

Chapter 22

Social Comparison, Identity, and Influence

A household receives its monthly utility statement. Alongside the usual numbers is a small chart: the household's own electricity use, the average of a hundred similar homes nearby, and the use of the most efficient neighbours. Nothing about the price of electricity has changed, and the chart contains no technical advice, yet households that receive such reports reliably use less energy. A hotel guest reads a card saying that most guests who stayed in this room reused their towels, and hangs hers back on the rail. A teenager wants the same brand of trainers as her friends, but in a colour none of them owns. An engineer notices a serious flaw in a project plan, looks around a meeting room in which no one else seems worried, and says nothing.

Each scene turns on the same fact: people do not evaluate themselves, form preferences, or decide how to act in isolation. They look to others: to learn what is true, to know where they stand, to belong, and to stand out. This chapter assembles the core theories of social influence that operate at the level of the individual in a social context. We begin with **social comparison theory**, which explains how people evaluate themselves by reference to others, and **social identity theory**, which explains how group memberships become part of the self. **Optimal distinctiveness theory** reconciles the competing needs to belong and to be unique. We then examine **conformity** and **social norms**, the processes by which groups shape individual judgement and behaviour, before turning to the effects of the mere presence of others on performance (**social facilitation** and **distraction-conflict theory**) and, finally, to the dynamics by which perceived majorities silence minorities (**the spiral of silence**) (Table 22.1).

Table 22.1. Chapter at a glance: theories of social comparison, identity, and influence.

Theory	Core question	Level of analysis	Key source
Social comparison theory	How do people evaluate their abilities and opinions?	Individual	Festinger (1954)
Social identity theory	How do group memberships shape the self and intergroup behaviour?	Individual / group	Tajfel & Turner (1979)
Optimal distinctiveness theory	How do people balance belonging and uniqueness?	Individual / group	Brewer (1991)
Conformity	Why do individuals adopt the judgements of a group?	Individual / group	Asch (1956), Deutsch & Gerard (1955)
Social norms (focus theory)	When do beliefs about what others do and approve guide behaviour?	Individual / community	Cialdini, Reno, & Kallgren (1990)
Social facilitation / inhibition	How does the presence of others change performance?	Individual	Triplett (1898), Zajonc (1965)

Table 22.1. Chapter at a glance: theories of social comparison, identity, and influence. *(continued)*

Theory	Core question	Level of analysis	Key source
Distraction-conflict theory	Why does the presence of others raise arousal?	Individual	Sanders, Baron, & Moore (1978), Baron (1986)
Spiral of silence	Why do perceived minorities stop expressing their views?	Individual / public	Noelle-Neumann (1974)

22.1 Social Comparison Theory

THEORY CARD

Core claim. People have a drive to evaluate their opinions and abilities. When objective standards are unavailable, they evaluate themselves by comparison with other people, preferably similar others.
Origins. Social psychology, Festinger (1954), later developments on downward comparison (Wills 1981), upward comparison (Collins 1996), and self-evaluation maintenance (Tesser 1988), reviewed in Suls, Martin, and Wheeler (2002) and Gerber, Wheeler, and Suls (2018).
Level of analysis. Individual (the comparer) in relation to comparison targets (individuals or groups).
Key constructs. Comparison target, similarity, upward versus downward comparison, assimilation versus contrast, relevance, attainability, closeness.
Boundary conditions. Comparison is most likely when objective standards are absent or ambiguous, when the dimension matters to the self, and when information about others is salient. Its effects depend on whether the comparer sees the target as similar and the target's standing as attainable.
Related theories. Social norms, Equity theory (Section 19.3), Social identity theory, Goal-gradient hypothesis (Section 19.5).

22.1.1 Origins and core logic

Festinger (1954) began from a simple premise: people need to know whether their opinions are correct and how good their abilities are. Where physical reality provides an answer, whether a piece of furniture will fit through a doorway, people consult it. Where it does not, whether one is a good writer, or whether one's political views are reasonable, they turn to **social reality**, comparing themselves with others. Festinger further proposed that people prefer to compare with **similar others**, because comparison with someone very different is uninformative, and that for abilities (in Western cultures at least) there is a **unidirectional drive upward**: a pressure to do somewhat better than others.

Later work distinguished the *direction* of comparison:

- **Downward comparison**, comparing with someone worse off, can protect or restore self-esteem, especially when people are under threat (Wills 1981).
- **Upward comparison**, comparing with someone better off, can be either deflating or inspiring. It deflates when the comparer feels *contrasted* with the target, and it inspires when the comparer *assimilates* toward the target, seeing the target's success as a glimpse of what is

possible (Collins 1996).

Lockwood and Kunda (1997) identified a key moderator: **attainability**. An outstanding role model inspires when the comparer believes similar success is achievable and demoralises when it seems out of reach. Tesser (1988) added **relevance** and **closeness**: being outperformed by a close friend on a dimension central to one's self-definition is threatening (a *comparison* process), whereas being outperformed on an irrelevant dimension lets one bask in the friend's success (a *reflection* process). A meta-analysis covering six decades of research supported Festinger's core idea that people actively seek comparison information while showing that the choice of targets and the consequences of comparison depend strongly on context (Gerber, Wheeler, and Suls 2018) (Table 22.2).

Table 22.2. Upward and downward social comparison (based on Wills, 1981, Collins, 1996, Lockwood & Kunda, 1997).

Feature	Upward comparison	Downward comparison
Target	Someone better off or more able	Someone worse off or less able
Typical motive	Self-improvement, accurate self-evaluation	Self-enhancement, coping with threat
Assimilative effect	Inspiration ('I could be like that')	Worry ('I could become like that')
Contrastive effect	Deflation, envy, lower self-evaluation	Relief, higher self-evaluation
Key moderators	Attainability, similarity, relevance	Threat, perceived control
Business example	Leaderboards showing top performers	Messages that one is 'doing better than most'

22.1.2 Mechanism

Figure 22.1 summarises how a comparison produces its effects. The comparison target and its direction set up the comparison. Whether the comparer assimilates or contrasts depends on attainability, similarity, and the relevance of the dimension, and the outcome is a change in self-evaluation, affect, and motivation.

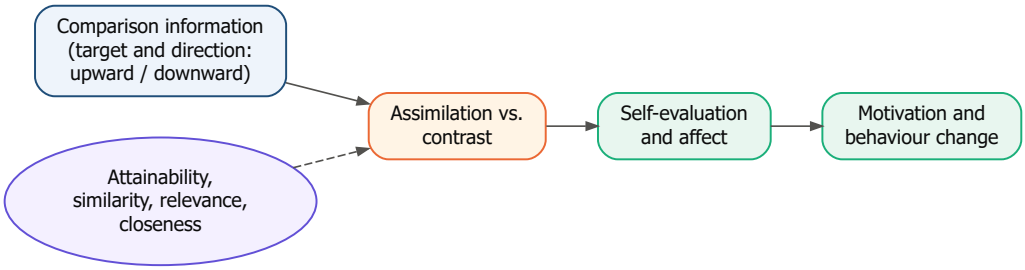


Figure 22.1. Mechanism of social comparison: a comparison target produces assimilation or contrast depending on attainability, similarity, and relevance, which in turn shape self-evaluation, affect, and motivation.

22.1.3 Applications in business

Organizations create comparison information constantly: sales leaderboards, customer-satisfaction rankings, peer benchmarks in performance reviews, "customers like you" panels, and fitness apps that show where a user ranks among friends. Social comparison theory predicts that such information motivates when it highlights attainable, relevant comparisons and backfires when it exposes large, apparently unbridgeable gaps.

BUSINESS CASE

Opower and the power of position. The utility-reporting firm Opower built its business on a comparison. Its home energy reports did not try to teach customers the science of energy efficiency. They told customers where they stood relative to their neighbours: whether they were better or worse at conserving energy than similar households nearby (Cuddy, Doherty, and Bos 2010). The insight was that *position relative to others* is a more powerful motivator than abstract facts. The reports also illustrate the theory's boundary conditions: the comparison group was chosen to be similar (homes of comparable size in the same area), and the reports paired the comparison with an evaluative message to prevent already-efficient households from drifting toward the average. We return to the evidence on these reports in the section on social norms.

COMMON PITFALL

Assuming that comparison always motivates. Ranking employees or customers can demoralise as easily as it inspires. People far below the top may contrast rather than assimilate, disengage from the dimension, or question the fairness of the metric. Designs should specify the comparison group (similar others), make improvement appear attainable (e.g., showing the next rung rather than the top), and anticipate effects on those who are already ahead.

22.2 Social Identity Theory

THEORY CARD

Core claim. Part of a person's self-concept derives from membership in social groups. People are motivated to achieve a positive social identity, which they do by favourably distinguishing their ingroups from relevant outgroups.
Origins. Social psychology, minimal group experiments (Tajfel et al. 1971), Tajfel and Turner (1979), extended by self-categorization theory (Turner et al. 1987), reviewed in Hornsey (2008).
Level of analysis. Individual and group. Extended to organizations (Ashforth and Mael 1989) and to consumer-brand relationships (Section 27.5) (C. B. Bhattacharya and Sen 2003).
Key constructs. Social categorization, social identification, social comparison between groups, positive distinctiveness, ingroup favouritism, permeability, stability, and legitimacy (Section @ref(legitimacy (Section 7.3))) of group boundaries, depersonalization, prototypes.
Boundary conditions. Effects depend on the salience of the category in the situation and on how strongly people identify with the group. When personal identity is salient, behaviour is guided more by individual than by group considerations.
Related theories. Optimal distinctiveness, Identity signaling (Section 8.5), Social comparison, Institutional theory (Section 7.2).

22.2.1 Origins and core logic

Social identity theory began with a surprising experimental result. In the **minimal group paradigm**, Tajfel et al. (1971) assigned participants to groups on the basis of trivial criteria, such as a preference between paintings by two abstract artists, and asked them to allocate rewards to

anonymous ingroup and outgroup members. Even though the groups had no history, no interaction, and no conflict of interest, participants favoured their own group, and in some allocation choices preferred to maximise the *difference* between groups rather than the absolute reward to their own group. Mere categorization was enough to produce intergroup discrimination.

Tajfel and Turner (1979) explained this with a three-step process. **Social categorization** sorts people, including oneself, into groups. **Social identification** makes group membership part of the self-concept, so that the group's fortunes become one's own. **Social comparison** between groups then motivates people to seek **positive distinctiveness**: a favourable difference between ingroup and relevant outgroups. When a group's identity is unsatisfactory, members choose among strategies whose viability depends on the social structure, as summarised in Table 22.3.

Table 22.3. Identity-management strategies in social identity theory (Tajfel & Turner, 1979).

Strategy	What members do	When it is likely	Business illustration
Individual mobility	Leave the group or dissociate from it	Group boundaries are permeable	Employees of a struggling firm move to a competitor
Social creativity	Change the comparison dimension or reframe its value	Boundaries impermeable. Status relations seen as stable and legitimate	A challenger brand positions itself on authenticity rather than scale
Social competition	Compete directly for status or resources	Status relations seen as unstable and illegitimate	Rival professions lobby over regulatory recognition

Self-categorization theory extended the account by specifying when people think of themselves as individuals or as group members (Turner et al. 1987). When a social category becomes salient, people undergo **depersonalization**: they see themselves less as unique individuals and more as interchangeable exemplars of the group **prototype**, and they adopt the group's norms as their own. The theory explains why identity-based influence is strongest from ingroup members who are seen as prototypical, and why the same person can behave very differently as a "marketing professional", a "parent", or a "fan" depending on which identity the situation activates (Hornsey 2008).

22.2.2 Evidence and extensions

Ashforth and Mael (1989) imported social identity theory into organization studies, defining **organizational identification** as the perception of oneness with or belongingness to an organization. Identification helps explain loyalty, cooperation, and support for organizational goals, and also resistance to mergers that threaten a valued identity. In marketing, C. B. Bhattacharya and Sen (2003) proposed **consumer-company identification**: consumers identify with companies whose identities are distinctive, prestigious, and similar to their own, and such identification fosters loyalty, advocacy, and resilience to negative information.

BUSINESS CASE

Illustrative scenario: post-merger identity. Two engineering firms of similar size merge. Employees of each legacy firm continue to describe themselves by their old affiliation, sit with former colleagues, and attribute problems to "them". Social identity theory interprets this as a predictable response to a threat to valued identities rather than as mere resistance to change. Remedies suggested by the theory include creating a superordinate identity that both groups can embrace, preserving valued elements of each legacy identity, ensuring that integration is

seen as fair (legitimacy), and avoiding status signals, office location, leadership appointments, systems choices, that mark one side as the "acquired" group.

COMMON PITFALL

Measuring membership instead of identification. Being a member of a group is not the same as identifying with it. Many studies use membership (for example, working for a firm or owning a brand) as a proxy for social identity. The theory's predictions apply to identification and to situations in which the identity is salient. Both should be measured or manipulated.

22.3 Optimal Distinctiveness Theory

THEORY CARD

Core claim. Social identities satisfy two opposing needs: a need for **assimilation** (inclusion, belonging) and a need for **differentiation** (uniqueness, distinctiveness). People identify most strongly with groups that reconcile these needs: groups inclusive enough to provide belonging but exclusive enough to provide distinctiveness.
Origins. Social psychology, Brewer (1991), reviewed and extended in Leonardelli, Pickett, and Brewer (2010), consumer applications in Chan, Berger, and Boven (2012).
Level of analysis. Individual and group.
Key constructs. Need for assimilation, need for differentiation, inclusiveness of the social category, optimal distinctiveness, need for uniqueness.
Boundary conditions. Distinctiveness is context-specific. The optimum shifts with the frame of reference and with the relative strength of the two needs, which vary across cultures, individuals, and recent experience.
Related theories. Social identity theory, Identity signaling (Section 8.5), Snob effect (Section 29.5), Conformity.

22.3.1 Origins and core logic

Brewer (1991) started from a puzzle in social identity theory: if group membership provides self-esteem, why do people not simply identify with the largest, most inclusive, highest-status groups available? Her answer was that social identity is the reconciliation of two opposing needs. **Personal identity** is the individuated self: the characteristics that differentiate one person from others. **Social identities** are categorizations of the self into more inclusive social units that depersonalize the self-concept, so that "I" becomes "we". Too much individuation produces isolation. Too much deindividuation means that self-definition is lost. People therefore seek a balance between **assimilation** and **differentiation**: a state of optimal distinctiveness. Brewer also noted that while some group memberships are imposed, the identities people embrace among the groups available to them are partly a matter of choice.

Brewer formalised the theory in four assumptions (Brewer 1991, 478):

> **A1.** "Social identification will be strongest for social groups or categories at that level of inclusiveness which resolves the conflict between needs for differentiation of the self and assimilation with others."

A2. "Optimal distinctiveness is independent of the evaluative implications of group membership, although, other things being equal, individuals will prefer positive group identities to negative identities."

A3. "Distinctiveness of a given social identity is context-specific. It depends on the frame of reference within which possible social identities are defined at a particular time, which can range from participants in a specific social gathering to the entire human race."

A4. "The optimal level of category distinctiveness or inclusiveness is a function of the relative strength (steepness) of the opposing drives for assimilation and differentiation. For any individual, the relative strength of the two needs is determined by cultural norms, individual socialization, and recent experience."

A2 is a notable departure from social identity theory: Brewer argued that people may identify strongly with small, distinctive groups even when those groups have modest status. Figure 22.2 depicts the opposing-processes logic. As a category becomes more inclusive, the need for differentiation is increasingly activated. As it becomes more exclusive, the need for assimilation is increasingly activated. The optimum lies where the two forces balance, and, consistent with A4, it shifts toward more inclusive groups when the need to belong is temporarily heightened: for example, after an experience of rejection.

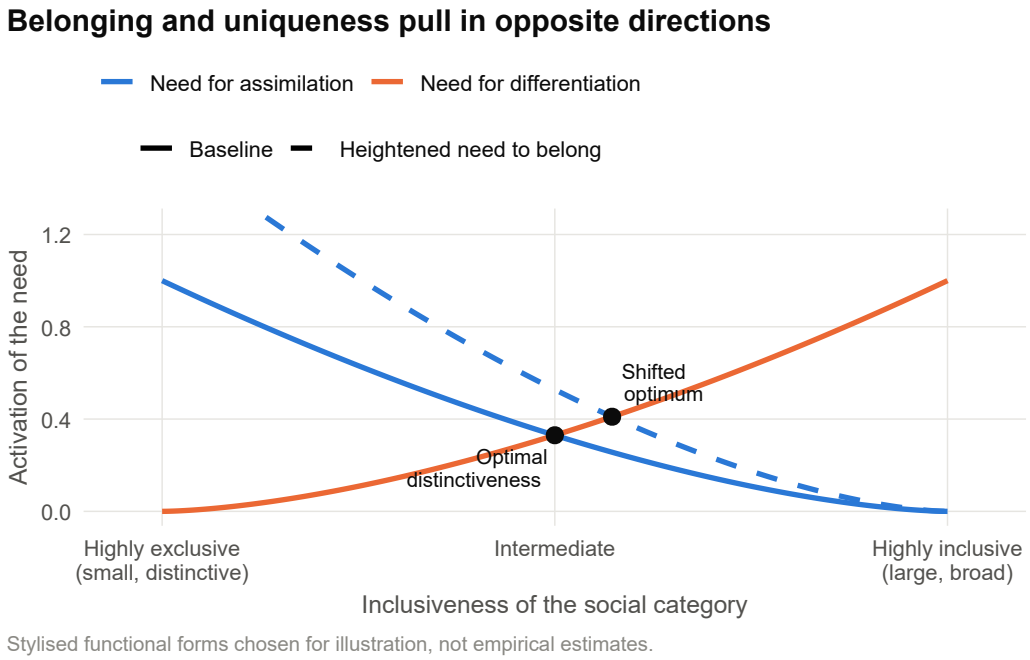


Figure 22.2. Stylised optimal distinctiveness model (after Brewer, 1991). The need for differentiation rises as a group becomes more inclusive and the need for assimilation rises as it becomes more exclusive. Identification peaks where the drives balance. A stronger assimilation drive (dashed line) shifts the optimum toward more inclusive groups.

22.3.2 Evidence and extensions: consumption as a marker of membership

Consumer research has made optimal distinctiveness one of its most productive theoretical lenses, because consumption gains symbolic meaning as a marker of group membership. The relevant literatures had long emphasised one motive or the other. On the assimilation side, people conform to others under informational and normative influence (Deutsch and Gerard 1955), and they make choices similar to those of their **aspiration groups** to signal desired identities (Englis and Solomon 1995, Berger and Heath 2007). On the differentiation side, people have a genuine **need for uniqueness**: they can experience negative emotion when they feel too similar to others, and they seek difference through possessions, opinions, and style (Snyder and Fromkin 1980, Lynn and Snyder 2002). Consumers' need for uniqueness can be measured as a stable individual difference (Tian, Bearden, and Hunter 2001), but it is also driven by situational factors. Research on social influence in networks likewise shows that influence can produce both convergence and divergence among individuals, depending on the structure of relationships and the identities at stake (W. A. Mason, Conrey, and Smith 2007).

Earlier work tended to assume that one motive prevails in any given decision. Chan, Berger, and Boven (2012) showed instead that consumers can satisfy **both** needs in a single choice by assimilating on one dimension and differentiating on another. Consumers conform to their group on attributes that signal group identity, such as the brand, while distinguishing themselves on attributes that signal individuality, such as colour. The desire to communicate social identity leads consumers to conform on dimensions strongly associated with their group, particularly in identity-relevant categories such as clothing, while higher need for uniqueness leads them to differentiate *within* the group by choosing less popular options among those associated with it. The result is a self-concept that is, in the title's phrase, identifiable but not identical.

A related line of work examines differentiation *between* groups. People diverge from members of dissimilar outgroups: when another social group adopts a taste, people may abandon it to avoid signalling an undesired identity and to avoid the costs of misidentification (Berger and Heath 2007, 2008). This mechanism is developed in the section on identity signaling (Section 8.5).

The creativity literature offers an intriguing parallel. Novel ideas tend to balance familiarity and novelty, recombining established elements in new ways, and averaged, prototypical forms are often judged attractive. Those ideas are developed elsewhere in the book, in the sections on the Geneplore model (Section 31.2) and beauty in averageness (Section 29.10) (Table 22.4).

Table 22.4. How consumers reconcile assimilation and differentiation (based on Brewer, 1991, Chan, Berger, & Van Boven, 2012, Berger & Heath, 2007).

Consumer strategy	Need served	Choice dimension	Example
Adopt the group's brand or category	Assimilation	Identity-signalling attribute	Buying the brand of trainers one's team wears
Choose a less common variant within it	Differentiation	Uniqueness attribute (colour, style, edition)	Choosing an unusual colourway of that brand
Customise a mainstream product	Both	Personalisation	Monogrammed or configurable products
Join a small, distinctive community	Both (optimal group size)	Group membership	A niche brand community or enthusiasts' club
Abandon a taste adopted by an outgroup	Differentiation (between groups)	Identity-relevant domain	Dropping a style once a dissimilar group adopts it

BUSINESS CASE

Identifiable but not identical: designing product lines. Optimal distinctiveness suggests that brands with strong group associations should offer variety along dimensions that do *not* dilute the identity signal. A sportswear brand can keep its logo, silhouette, and brand name constant, the assimilation cues, while offering many colours, limited editions, and personalisation: the differentiation cues. The theory also warns that variety on the identity-signalling dimension itself (for example, removing the logo) may weaken the reason consumers chose the brand. For managers, the practical question is to identify which attributes in their category carry group meaning and which carry individual meaning, and to design assortment accordingly (Chan, Berger, and Boven 2012).

COMMON PITFALL

Treating uniqueness and conformity as opposite ends of one scale. The theory's central claim is that both needs operate simultaneously and are satisfied on different dimensions or at different levels of inclusiveness. Measuring a single "conformity versus uniqueness" preference will miss the pattern. Research should distinguish the attributes or groups on which consumers assimilate from those on which they differentiate.

22.4 Conformity

THEORY CARD

Core claim. Individuals change their judgements and behaviour to match those of a group, either because they take others' responses as evidence about reality (**informational influence**) or because they want to meet others' expectations and be accepted (**normative influence**).
Origins. Social psychology, Sherif (1936), Asch (1956), Deutsch and Gerard (1955), reviewed in Cialdini and Goldstein (2004).
Level of analysis. Individual within a group, aggregated to markets and networks (W. A. Mason, Conrey, and Smith 2007, Salganik, Dodds, and Watts 2006).
Key constructs. Informational influence, normative influence, public compliance versus private acceptance, unanimity, group size, anonymity, social proof, herding.
Boundary conditions. Informational influence is strongest when the task is ambiguous and others seem knowledgeable. Normative influence is strongest when responses are public, the group is valued, and the majority is unanimous.
Related theories. Social norms, Optimal distinctiveness, Wisdom of the crowd (Section 30.5), Diffusion of innovations (Section 10.1).

22.4.1 Origins and core logic

Sherif (1936) showed that when people judged the apparent movement of a point of light in a dark room, an inherently ambiguous stimulus, their estimates converged on a group norm, which persisted when they later judged alone. Here conformity looks like rational learning: when reality is uncertain, others' judgements are informative.

Asch (1956) asked a sharper question: would people conform even when the right answer was obvious? Participants judged which of three comparison lines matched a standard line, a task on which

people working alone almost never erred. In the critical trials, a unanimous group of confederates gave the same wrong answer before the participant responded. Roughly a third of participants' responses on these trials conformed to the erroneous majority, and most participants conformed at least once, although a substantial minority never did. Variations showed that a single dissenting ally sharply reduced conformity, highlighting the power of unanimity.

Deutsch and Gerard (1955) distinguished the two forces at work. **Informational influence** is the tendency to accept information from others as evidence about reality. **Normative influence** is the tendency to conform to the positive expectations of others. Their experiments showed that conformity in an Asch-type task declined when participants responded anonymously, evidence of a normative component, but did not disappear, consistent with informational influence. Cialdini and Goldstein (2004) integrate the literature by linking conformity to three goals: **accuracy** (informational), **affiliation** (normative), and the maintenance of a **positive self-concept** (Table 22.5).

Table 22.5. Informational versus normative social influence (based on Deutsch & Gerard, 1955, Cialdini & Goldstein, 2004).

Dimension	Informational influence	Normative influence
Underlying goal	Accuracy: be right about reality	Affiliation: be liked and accepted
Type of change	Private acceptance (internalisation)	Often public compliance only
Strongest when	Task is ambiguous, others seem expert	Responses are public, group is valued, majority unanimous
Weakened by	Clear objective evidence, own expertise	Anonymity, a dissenting ally
Classic study	Sherif (1936) autokinetic effect	Asch (1956) line judgements
Business example	Best-seller lists and review counts	Dress codes, visible participation in team rituals

22.4.2 Evidence and extensions

A meta-analysis of Asch-type studies found that conformity levels were higher in collectivist than in individualist cultures and had declined in the United States since the 1950s (R. Bond and Smith 1996), suggesting that conformity is shaped by cultural values (see Hofstede (Section 18.2)). In markets, conformity aggregates into **herding** and **social proof**. In a large online experiment, Salganik, Dodds, and Watts (2006) created parallel "worlds" in which participants could download songs. When participants could see how often others had downloaded each song, popularity became more unequal and less predictable across worlds than when they could not. A randomized experiment on a social news site found that giving a comment an artificial positive vote made subsequent positive votes more likely and raised the comment's final rating, whereas artificial negative votes tended to be corrected by later users (Muchnik, Aral, and Taylor 2013). And in a field experiment in a restaurant chain, displaying the most popular dishes increased demand for those dishes, consistent with observational learning (Cai, Chen, and Fang 2009).

BUSINESS CASE

Social proof in e-commerce. Online retailers and platforms routinely display "best-seller" badges, review counts, "N people are viewing this item", and "customers also bought" panels. Each exploits informational influence: under uncertainty about quality, others' choices are a signal. The research above suggests two managerial cautions. First, social proof amplifies early random advantages, so displayed popularity may reflect luck as much as quality (Sal-

ganik, Dodds, and Watts 2006). Platforms that care about matching customers to the best products should guard against self-reinforcing rankings, for example by mixing popularity with quality signals or by randomising exposure for new items. Second, herding is asymmetric: positive signals snowball more readily than negative ones (Muchnik, Aral, and Taylor 2013), which makes rating systems vulnerable to early manipulation.

COMMON PITFALL

Interpreting all similarity of behaviour as conformity. People who behave alike may have been influenced by one another, may share unobserved preferences (homophily (Section @ref(homophily (Section 30.2)))), or may face common shocks. Inferring social influence from correlated behaviour in observational data is notoriously difficult. Randomised manipulations of what others are seen to do, as in the studies above, provide the clearest evidence.

22.5 Social Norms

THEORY CARD

Core claim. Norms influence behaviour mainly when they are salient. Two kinds of norm must be distinguished: **descriptive norms** (what most people do) and **injunctive norms** (what most people approve or disapprove of). Messages that activate the wrong norm can backfire.
Origins. Social psychology, focus theory of normative conduct (Cialdini, Reno, and Kallgren 1990), field applications in energy (Schultz et al. 2007, Allcott 2011), hotels (N. J. Goldstein, Cialdini, and Griskevicius 2008), and taxation (Hallsworth et al. 2017).
Level of analysis. Individual within a community. Designed interventions at the level of firms and public agencies.
Key constructs. Descriptive norm, injunctive norm, norm salience (focus), reference group, boomerang effect, pluralistic ignorance.
Boundary conditions. Norm messages work best when the reference group is similar and identifiable, the norm is credible, and the behaviour is feasible. Descriptive norms can backfire for people who are already doing better than the norm or when they reveal that an undesirable behaviour is common.
Related theories. Conformity, Social comparison, Social identity theory, Spiral of silence.

22.5.1 Core logic

Cialdini, Reno, and Kallgren (1990) argued that the concept of norms had lost explanatory power because it was invoked after the fact to explain almost anything. Their **focus theory of normative conduct** restores predictive content by distinguishing two kinds of norm and by holding that a norm guides behaviour only when attention is focused on it. In field experiments on littering, people were least likely to litter when they saw someone drop litter into a *clean* environment, an act that drew attention to the descriptive norm that people here do not litter, and most likely to litter when the environment was already littered. Seeing someone *pick up* litter, which focused attention on the injunctive norm, reduced littering regardless of the state of the environment (Table 22.6).

Table 22.6. Descriptive versus injunctive norms (based on Cialdini, Reno, & Kallgren, 1990).

Dimension	Descriptive norm	Injunctive norm
Definition	Perception of what most others do	Perception of what most others approve or disapprove
Motivational basis	Information: what is likely effective or adaptive	Social sanctions: approval and disapproval
Typical message	'Most guests reuse their towels'	'Wasting energy is frowned upon'
Risk	Backfires if it reveals that the bad behaviour is common, or pulls low users up to the average	Can feel preachy or controlling
Best use	When the desired behaviour is already common	When the desired behaviour is uncommon, or combined with descriptive feedback

Norm perceptions are often inaccurate. Prentice and Miller (1993) showed that university students systematically overestimated their peers' comfort with heavy drinking: a case of **pluralistic ignorance**, in which most members of a group privately reject a norm but believe that most others accept it. Correcting such misperceptions is one of the main levers of norm-based interventions. A striking feature of normative influence is that people underestimate it: in a study of energy conservation, residents rated information about their neighbours' behaviour as the *least* motivating reason to conserve, yet it produced larger changes in behaviour than appeals to environmental protection, social responsibility, or saving money (Nolan et al. 2008).

22.5.2 Mechanism and the boomerang effect

Schultz et al. (2007) provided the canonical demonstration of how descriptive norms can backfire. Households in a Californian community received feedback comparing their electricity use to the average in their neighbourhood. Households that had been using *more* than average subsequently reduced consumption, as intended. But households that had been using *less* than average increased their consumption: a **boomerang effect**, as the descriptive norm pulled them toward the average. When the feedback was accompanied by an injunctive cue, a smiling emoticon for below-average users and a frowning emoticon for above-average users, the boomerang effect among low users disappeared. Figure 22.3 reproduces the pattern with simulated data.

Hotel guests show the same sensitivity to descriptive norms. N. J. Goldstein, Cialdini, and Griskevicius (2008) found that a card stating that most guests reused their towels produced more towel reuse than the hotel industry's standard environmental appeal, and that a "provincial" norm referring to guests who had stayed in *the same room* was more effective still, even though that reference group is not more meaningful in any objective sense.

22.5.3 Applications in business and policy

BUSINESS CASE

Opower home energy reports. Opower, a software firm working with utilities, scaled the Schultz et al. design to millions of households. Its reports compared each home's energy use with that of similar neighbours and of the most efficient neighbours, added an injunctive element (smiling faces for good performance), and offered conservation tips (Cuddy, Doherty, and Bos 2010). Because utilities randomly assigned which households received reports, the programme became one of the best-evaluated behavioural interventions in the world.

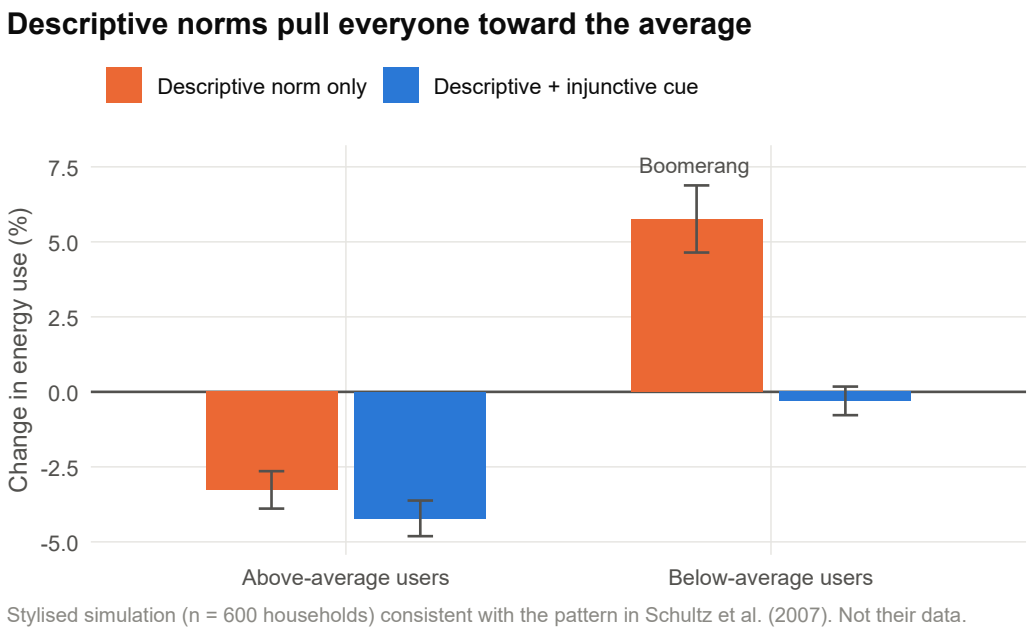


Figure 22.3. Stylised boomerang effect of descriptive-norm feedback. Households above the neighbourhood average reduce use under both message types. Households below the average drift upward when they receive descriptive feedback alone, but not when an injunctive cue (approval) is added. Simulated data. Bars show means with 95% confidence intervals.

Analysing randomised trials covering hundreds of thousands of households, Allcott (2011) estimated that the reports reduced energy consumption by about 2 percent on average, with larger effects among households that had used more energy at baseline. Follow-up research found that households responded to each report and then partially backslid, that continued reports produced more persistent changes in behaviour, and that a meaningful share of the effect persisted after reports were discontinued (Allcott and Rogers 2014). The case illustrates both the power of norms at scale and the importance of the injunctive element in preventing boomerang effects among already-efficient households.

BUSINESS CASE

Social-norm letters and tax compliance. In two large natural field experiments, the UK tax authority varied the wording of reminder letters sent to taxpayers who had not yet paid what they owed (Hallsworth et al. 2017). Adding a descriptive-norm statement, that the great majority of people pay their tax on time, increased payment rates relative to a standard letter, and messages that made the norm more specific to the recipient's situation tended to be more effective. Because the intervention cost almost nothing to implement, the additional revenue collected represented a very high return. The study has become a template for norm-based communications in public administration and, by extension, in business collections, compliance, and renewal reminders.

DEBATE AND EVIDENCE

Norms as policy: scale, persistence, and ethics. Norm-based nudges are cheap, scalable, and well-evidenced in field experiments, but three debates remain. First, effects are typically modest in size, a few percent in energy use, so their value depends on scale and low cost rather than on transforming behaviour. Second, persistence varies: effects can decay when messages stop, although repeated messaging can build habits (Allcott and Rogers 2014). Third, the ethics of "managing" people through social pressure is contested, especially when the norm statements are selectively framed. Transparency about the source and accuracy of normative claims is a practical safeguard: a norm message that turns out to be misleading destroys the credibility that makes such messages work.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Descriptive-norm feedback will reduce resource use among above-average users and increase it among below-average users. Adding an injunctive cue will eliminate the increase.
- *H2.* Norm messages that refer to a more proximal reference group (same neighbourhood, same team, same room) will be more effective than messages that refer to a broad population.
- *H3.* Correcting overestimates of colleagues' tolerance for unethical shortcuts (pluralistic ignorance) will increase employees' reporting of misconduct.

Designs. Randomised field experiments varying message content (the Opower and tax-letter designs are templates). Survey measures of perceived descriptive and injunctive norms with experimental corrections. Difference-in-differences analyses of staggered roll-outs.
Measures. Behavioural outcomes (consumption, payment, reuse, reporting). Perceived norms (estimated percentage of peers who perform a behaviour. Perceived approval). Heterogeneity by baseline position relative to the norm.

22.6 Social Facilitation and Inhibition

THEORY CARD

Core claim. The presence of others, as an audience or as co-actors, improves performance on simple or well-learned tasks (**social facilitation**) and impairs performance on complex or novel tasks (**social inhibition**).
Origins. Social psychology, Triplett (1898), the term "social facilitation" from Allport (1920), the drive theory synthesis of Zajonc (1965), evaluation apprehension (Cottrell et al. 1968), meta-analysis in C. F. Bond and Titus (1983).
Level of analysis. Individual performing in the presence of others.
Key constructs. Audience effect, coaction effect, dominant response, drive or arousal, task complexity or familiarity, mere presence, evaluation apprehension.
Boundary conditions. Effects depend on the task: facilitation requires that the correct response be dominant (well-learned). Inhibition arises when the dominant response is incorrect. Effects are generally small in magnitude.
Related theories. Distraction-conflict theory, Social comparison, Goal-setting theory (Sec-

> tion 19.4).

22.6.1 Origins and intellectual history

Social facilitation is often described as the topic of the first experiment in social psychology. Triplett (1898) observed from cycling records that racers rode faster when paced or in competition than when riding alone against the clock, and he followed this with a laboratory study in which children wound fishing reels either alone or alongside another child. He reported that many children worked faster in competition. A modern reanalysis of Triplett's data concluded that, by contemporary statistical standards, the evidence in the original experiment was considerably weaker than textbook accounts suggest (Strube 2005): a useful reminder that founding studies are not always the strongest.

Allport (1920) coined the term **social facilitation**. Working on "the influence of the group upon association and thought", he found that people working alongside others produced more free associations than people working alone, but that the quality of their reasoning did not improve in the same way. Two research streams developed:

- **Audience effects**, in which passive spectators influence performance. Skilled participants performed an eye-hand coordination task better before a small audience (Travis 1925), whereas the presence of a spectator impaired the learning of lists of nonsense syllables (Pessin 1933).
- **Coaction effects**, in which others perform the same task independently. These were documented in animals as well as people: worker ants excavated more nest material in groups than alone (S. C. Chen 1937), and armadillos ate more in the presence of conspecifics (J. J. Platt, Yaksh, and Darby 1967).

By the 1960s the literature appeared contradictory: the presence of others sometimes helped and sometimes hurt.

22.6.2 Core logic: drive theory

Zajonc (1965) resolved the contradiction with an elegant argument from learning theory. The presence of members of one's own species increases **generalized drive** (arousal). Drive increases the likelihood of **dominant responses**: the responses that are most strongly established in a given situation. On simple or well-learned tasks the dominant response is usually correct, so performance improves. On complex or novel tasks the dominant response is often incorrect, so performance suffers. In short, an audience facilitates the emission of well-learned responses and impairs the acquisition of new ones. Zajonc and Sales (1966) provided direct support: after participants had practised nonsense words at different frequencies, the presence of an audience increased their tendency to give the most-practised (dominant) words in a later task (Table 22.7).

Table 22.7. Task familiarity moderates the effect of social presence (after Zajonc, 1965).

Social presence	Task	Dominant response	Consequence
Others present (audience or coactors)	Automatic, simple, well-learned	Usually correct	Social facilitation: faster, better performance
Others present (audience or coactors)	Difficult, novel, requires attention	Often incorrect	Social inhibition: slower, more errors

Table 22.7. Task familiarity moderates the effect of social presence (after Zajonc, 1965). *(continued)*

Social presence	Task	Dominant response	Consequence
Alone	Any	Baseline	Reference condition

Figure 22.4 shows the resulting interaction with illustrative data, and Figure 22.5 depicts the mechanism.

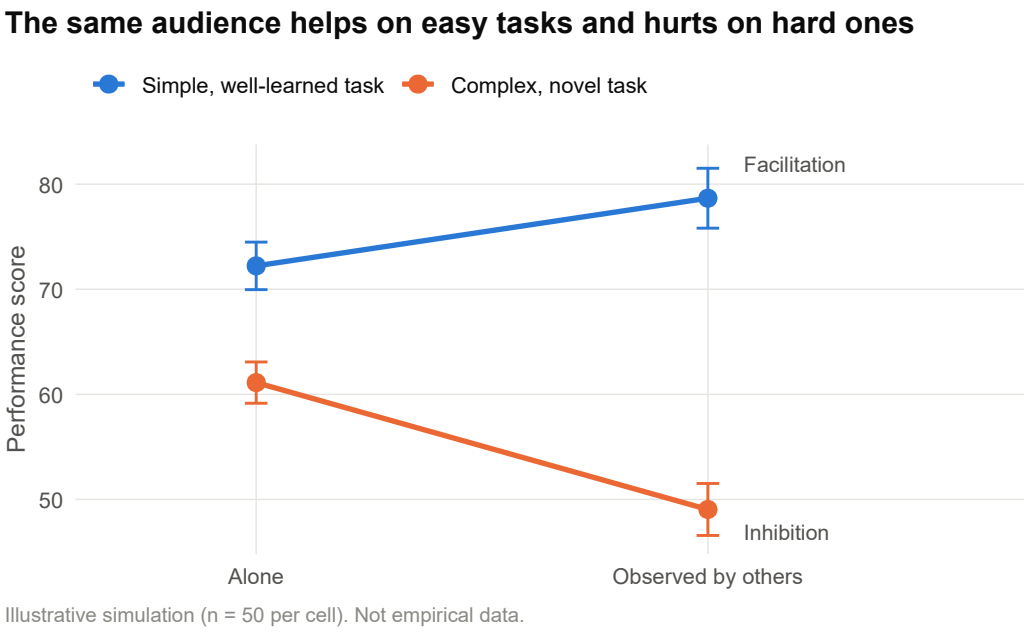


Figure 22.4. Illustrative social facilitation and inhibition: the presence of an audience improves performance on a simple, well-learned task and impairs performance on a complex, novel task. Simulated data. Points are means with 95% confidence intervals.

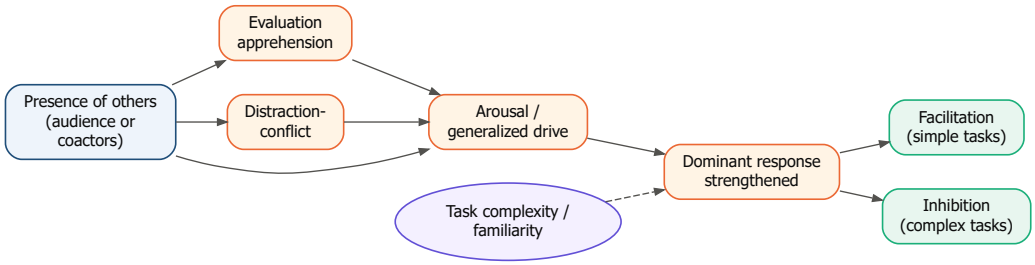


Figure 22.5. Mechanism of social facilitation and inhibition (drive theory, after Zajonc, 1965): the presence of others raises arousal, which strengthens dominant responses. Task complexity determines whether the dominant response is correct. Evaluation apprehension and distraction-conflict are alternative accounts of why presence raises arousal.

22.6.3 Evidence and extensions: why does presence arouse?

Drive theory left open *why* the presence of others should raise drive. Three answers have been proposed (Table 22.8).

- **Mere presence.** Zajonc held that the presence of conspecifics is innately arousing because others are unpredictable. H. Markus (1978) tested this unobtrusively: participants who dressed and undressed in familiar clothing did so faster, and in unfamiliar clothing more slowly, when an inattentive person was in the room than when they were alone.
- **Evaluation apprehension.** Cottrell et al. (1968) argued that people learn to associate others' presence with evaluation. In their study, an attentive audience enhanced dominant responses, but a blindfolded "audience" that could not evaluate did not. On this view the passive spectator is not really passive: what matters is the performer's knowledge of being evaluated, a concern closely related to impression management in organizations.
- **Distraction-conflict.** Others are a source of distraction that competes with the task for attention, and the resulting attentional conflict raises arousal (see the next section).

A meta-analysis of 241 studies concluded that the presence of others has small effects overall: it slightly increases the speed of simple performance and impairs both the speed and the accuracy of complex performance, and these effects were not fully explained by evaluation apprehension (C. F. Bond and Titus 1983).

Table 22.8. Explanations of social facilitation and a contrasting phenomenon.

Account	Why presence matters	Key evidence	Distinctive prediction
Drive / mere presence	Others are inherently arousing and unpredictable	Zajonc (1965), Markus (1978)	Effects even with an inattentive other
Evaluation apprehension	Others are associated with evaluation	Cottrell et al. (1968)	No effect if others cannot evaluate
Distraction-conflict	Others divide attention between task and audience	Sanders et al. (1978), Baron (1986)	Non-social distractions produce similar effects
Contrast: social loafing	Individual contributions are not identifiable in a collective task	Latane et al. (1979), Karau & Williams (1993)	Effort falls when outputs are pooled

Social facilitation should be distinguished from **social loafing (Section 30.9)**, the tendency to exert less effort when working collectively than individually (Latané, Williams, and Harkins 1979, Karau and Williams 1993). Facilitation arises when individual performance is visible to others. Loafing arises when individual contributions are pooled and cannot be identified. The same team can show either, depending on how work and evaluation are designed.

22.6.4 Applications in business

BUSINESS CASE

Open-plan offices, monitoring, and peer visibility. Many organizations have redesigned workspaces to increase visibility and interaction. Social facilitation theory predicts that visibility will help routine, well-practised work and hinder complex, novel work that demands concentration. Field evidence is consistent with a cautious view. In two studies of corporate headquarters that moved to open-plan layouts, face-to-face interaction fell sharply while elec-

tronic communication rose: the opposite of the intended effect (Bernstein and Turban 2018). In a laboratory study of electronic performance monitoring, monitoring raised the productivity of highly skilled participants but lowered that of less skilled participants, and monitored participants reported more stress (Aiello and Kolb 1995): the facilitation-inhibition pattern applied to skill rather than task. Visibility to peers can also raise effort: among supermarket cashiers, productivity rose when a highly productive co-worker joined the shift, especially when the cashier was in that co-worker's line of sight (Mas and Moretti 2009). The managerial implication is to match visibility to task: open, observable settings for practised, routine work. Quiet, protected settings for learning and complex problem solving.

22.7 Distraction-Conflict Theory

THEORY CARD

Core claim. Others (and other distractors) draw attention away from a task. When a person feels pulled between attending to the task and attending to the distraction, the resulting **attentional conflict** raises arousal and narrows attention, producing facilitation on simple tasks and impairment on complex ones.
Origins. Social psychology, Sanders, Baron, and Moore (1978), reviewed in R. S. Baron (1986).
Level of analysis. Individual.
Key constructs. Distraction, attentional conflict, drive, attentional overload, narrowing of attention (cue utilization).
Boundary conditions. Effects require that the distractor be salient and that the person feel pressure to attend to both the task and the distractor. Distractors that can be ignored easily produce little conflict.
Related theories. Social facilitation, Information processing model (Section 31.1), Social comparison.

22.7.1 Core logic

Distraction-conflict theory offers a mechanism for social facilitation that does not depend on anything uniquely social. Sanders, Baron, and Moore (1978) proposed that others are distracting, people glance at them, wonder what they think, and compare their own progress with that of co-actors, and that distraction creates a conflict between two attentional demands. In their experiments, co-actors produced the classic pattern of facilitation on simple tasks and impairment on complex tasks mainly when they were working on the *same* task, making them relevant targets for social comparison and thus more distracting. Participants in those conditions also reported greater distraction.

R. S. Baron (1986) developed the theory in two directions. First, attentional conflict can raise drive, producing Zajonc's dominant-response effects. Second, conflict can create **attentional overload**, which leads people to narrow their attention to a smaller set of cues. Narrowing helps on simple tasks that require attention to only a few central cues, irrelevant cues are screened out, but harms complex tasks that require attention to many cues. Because the account is about attention rather than about other people as such, it predicts that non-social distractions such as noise or flashing lights can produce effects similar to those of an audience, and Baron reviewed evidence consistent with that prediction. Figure 22.6 summarises the mechanism.

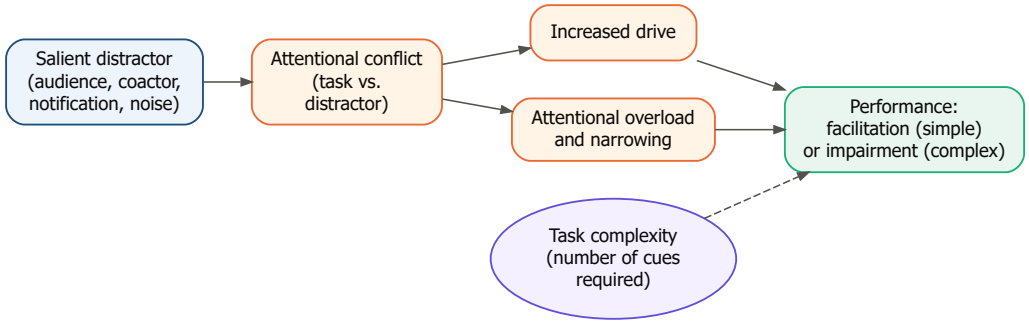


Figure 22.6. Mechanism of distraction-conflict theory (after Sanders, Baron, & Moore, 1978, Baron, 1986): a salient distractor creates attentional conflict with the task, which raises drive and produces attentional overload and narrowing. Task complexity determines whether narrowing helps or hurts.

BUSINESS CASE

Illustrative scenario: dashboards and interruptions in the contact centre. A customer-service centre installs wall-mounted screens showing each agent's live call statistics and ranking. Distraction-conflict theory predicts mixed effects. For experienced agents handling routine queries, the screens may narrow attention onto the core script and raise throughput. For newer agents, or for complex complaints that require attention to many cues, the customer's tone, account history, policy exceptions, the competing pull of the leaderboard may divide attention and increase errors. Similar logic applies to knowledge workers facing frequent notifications. A testable design would randomise the visibility of real-time metrics across teams and examine effects separately for routine and complex cases and for experienced and novice staff.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Real-time public performance displays will increase speed on routine tasks and reduce accuracy on complex tasks, with larger effects for less experienced employees.
- *H2.* The effect of observation on complex-task performance will be mediated by self-reported distraction and attentional narrowing.
- *H3.* Non-social interruptions (notifications) will reproduce the facilitation-inhibition pattern produced by an observer.

Designs. Laboratory experiments crossing presence (alone, audience, blindfolded audience, non-social distractor) with task complexity. Field experiments that vary monitoring or workspace openness. Natural experiments from office redesigns or remote-work transitions.
Measures. Speed and accuracy separately, physiological arousal, self-reported distraction, peripheral-cue recall as an index of attentional narrowing.

COMMON PITFALL

Ignoring task complexity. The most common error in applying social facilitation and distraction-conflict theory is to predict a main effect of observation, monitoring, or open workspaces on "performance". The theories predict an interaction: the same social presence

helps on simple, practised tasks and hurts on complex, novel ones. Studies and workplace policies that average across task types will find small or inconsistent effects.

22.8 Spiral of Silence

THEORY CARD

Core claim. Because people fear social isolation, they continuously monitor the climate of opinion. Those who believe their view is in the minority, or losing ground, become less willing to express it, which makes the view appear even weaker and further silences its holders, in a self-reinforcing spiral.
Origins. Political science and mass communication, Noelle-Neumann (1974), meta-analytic assessment in Glynn, Hayes, and Shanahan (1997).
Level of analysis. Individual expression aggregated to public opinion. Analogues in organizations (employee voice and silence).
Key constructs. Fear of isolation, quasi-statistical sense of the climate of opinion, perceived current and future majority, willingness to speak out, mass media as a source of climate perceptions, the "hard core" who speak out regardless.
Boundary conditions. The spiral is most likely for morally loaded or controversial issues, in public settings, and when people rely on media or other indirect cues to judge opinion. It is weaker among those with strong convictions and in settings where dissent is protected.
Related theories. Conformity, Social norms, Agenda setting (Section 26.5), Pack journalism (Section 30.7).

22.8.1 Core logic

Noelle-Neumann (1974) proposed that people possess a **quasi-statistical sense** with which they estimate the distribution of opinion in their environment and whether it is shifting. Driven by **fear of isolation**, people who perceive that their opinion is shared by the majority, or is gaining ground, speak out more confidently. Those who perceive their view to be in the minority, or losing ground, fall silent. Because public expression is itself one of the main cues that others use to judge the climate of opinion, silence makes the minority view seem even rarer than it is, which further suppresses its expression. The mass media play a central role: they are a principal source of information about what "most people" think, so media emphasis can shape perceptions of the dominant opinion and of the dominant culture. One implication is that a person's perceived social environment may differ substantially from reality. Noelle-Neumann also noted the existence of a **hard core** who continue to express their views regardless of the perceived climate (Figure 22.7).

22.8.2 Evidence and extensions

The spiral of silence has been tested mainly with surveys that measure people's perceptions of majority opinion and their stated willingness to express their own views in hypothetical public situations. A meta-analysis of such studies found a positive but small relationship between perceived support for one's opinion and willingness to speak out (Glynn, Hayes, and Shanahan 1997). The theory thus captures a real tendency but not a deterministic law, and its effect is easily swamped by individual differences in conviction, interest, and personality. The closely related concept of

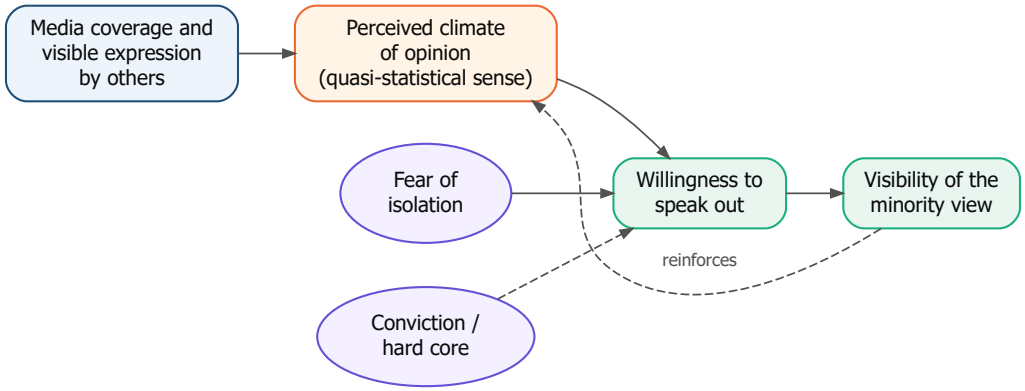


Figure 22.7. The spiral of silence (after Noelle-Neumann, 1974): perceptions of the climate of opinion, shaped by media and by others' expression, combine with fear of isolation to reduce minority expression, which in turn reinforces the impression that the view is rare.

pluralistic ignorance (Prentice and Miller 1993), widespread private rejection of a norm that everyone believes others accept, describes the end state that a spiral of silence can produce.

DEBATE AND EVIDENCE

Does the spiral survive the internet? Online platforms both weaken and strengthen the conditions for a spiral of silence. On one hand, anonymity and the ease of finding like-minded others should reduce fear of isolation and allow minority views to be expressed. On the other, visible metrics of approval, algorithmic amplification of popular content, and the permanence of online statements may make the climate of opinion more salient and the costs of deviation higher. Researchers also debate whether "willingness to speak out" in hypothetical survey scenarios captures actual expression. The theory's core mechanisms, perceived climate, fear of isolation, and expression as a signal, offer a framework for these questions, but the size and direction of effects in digital environments must be established empirically rather than assumed.

22.8.3 Applications in organizations: employee voice and silence

The organizational analogue of the spiral of silence is **organizational silence**: a collective-level phenomenon in which employees withhold information about problems and concerns (Morrison and Milliken 2000). Morrison and Milliken (2000) argued that silence develops when managers fear negative feedback and hold implicit beliefs, for example, that employees are self-interested or that management knows best, which shape structures and behaviours that discourage upward communication. Employees then form shared perceptions that speaking up is futile or dangerous. In interviews, most employees in one study reported having remained silent about issues that concerned them at work, frequently citing fear of being labelled negatively or damaging relationships (Milliken, Morrison, and Hewlin 2003). Leadership matters: managerial openness has been linked to higher employee voice through its effect on employees' sense that speaking up is safe (Detert and Burris 2007), and team **psychological safety (Section 14.8)**, a shared belief that the team is safe for interpersonal risk-taking, supports learning behaviour in work teams (A. Edmondson 1999).

BUSINESS CASE

Breaking the spiral in meetings and organizations. The engineer who notices a flaw but sees no one else worried is caught in a micro-spiral of silence: the silence of others is read as agreement, and the engineer's own silence becomes part of the cue for the next person. Organizations use several practices that map directly onto the theory's mechanisms. *Anonymous or simultaneous input* (written pre-reads, independent estimates before discussion, anonymous polling) prevents early speakers from defining the climate of opinion. *Designated dissent* (red teams, pre-mortems, a rotating "devil's advocate") lowers the isolation cost of disagreement by making it a role. *Leader behaviour*, senior people speaking last and visibly rewarding bad news, changes the perceived climate. And *measurement* through confidential surveys can reveal pluralistic ignorance, showing employees that concerns they believed were idiosyncratic are widely shared.

COMMON PITFALL

Equating silence with agreement. Managers and analysts often infer consensus from the absence of objections, and researchers sometimes infer public opinion from public expression. The spiral of silence predicts that both inferences are biased in favour of the perceived majority. Measure private opinions separately from public expression, and treat the gap between them as an outcome of interest.

22.9 Chapter Summary

KEY TAKEAWAYS

- People evaluate themselves through **social comparison** with similar others. Upward comparisons inspire when success seems attainable and deflate when it does not (Festinger 1954, Lockwood and Kunda 1997).
- **Social identity theory** explains how group memberships become part of the self and why people seek positive distinctiveness for their groups (Tajfel and Turner 1979, Turner et al. 1987).
- **Optimal distinctiveness theory** holds that identification is strongest for groups that balance assimilation and differentiation. Consumers often conform on identity attributes and differentiate on others (Brewer 1991, Chan, Berger, and Boven 2012).
- **Conformity** reflects informational influence (to be right) and normative influence (to be accepted). In markets it appears as social proof and herding (Deutsch and Gerard 1955, Asch 1956, Salganik, Dodds, and Watts 2006).
- **Descriptive norms** can backfire by pulling low users toward the average. Pairing them with **injunctive** cues prevents boomerang effects, as in Opower's home energy reports (Cialdini, Reno, and Kallgren 1990, Schultz et al. 2007, Allcott 2011).
- The presence of others facilitates simple, well-learned performance and impairs complex, novel performance. **Drive**, **evaluation apprehension**, and **distraction-conflict** offer complementary explanations (Zajonc 1965, Cottrell et al. 1968, R. S. Baron 1986).
- The **spiral of silence** describes how fear of isolation suppresses minority views. Its organizational analogue is employee silence, which psychological safety and process design can counteract (Noelle-Neumann 1974, Morrison and Milliken 2000).

Table 22.9 presents the comparison of the chapter's theories.

Table 22.9. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Typical business use
Social comparison	Self-evaluation against similar others	Assimilation vs. contrast depending on attainability	Leaderboards, peer benchmarks, customer feedback
Social identity	Categorization, identification, positive distinctiveness	Ingroup favouritism, identity-management strategies	Mergers, employer branding, brand communities
Optimal distinctiveness	Opposing needs for assimilation and differentiation	Conform on identity attributes, differentiate on others	Product-line and assortment design, customisation
Conformity	Informational and normative influence	Convergence on group judgements, herding	Social proof, ratings, best-seller lists
Social norms	Salient descriptive and injunctive norms	Boomerang without injunctive cue	Energy reports, tax letters, compliance messaging
Social facilitation	Arousal strengthens dominant responses	Help on simple tasks, harm on complex tasks	Workspace design, monitoring, team visibility
Distraction-conflict	Attentional conflict, overload, narrowing	Non-social distractors mimic audience effects	Notifications, dashboards, interruptions
Spiral of silence	Fear of isolation and perceived climate	Perceived minorities fall silent	Employee voice, meeting design, opinion research

22.10 Discussion Questions

1. Opower's reports combined social comparison, descriptive norms, and injunctive cues. Design an experiment that would isolate the contribution of each component, and state the predicted pattern of results.
2. Optimal distinctiveness theory suggests that consumers conform on some attributes and differentiate on others. Choose a product category and identify which attributes carry group meaning and which carry individual meaning. How would you test your classification?
3. Asch's participants conformed on an unambiguous task. Is conformity in online ratings better explained by informational or normative influence? What evidence would distinguish the two?
4. A firm plans to move all teams into an open-plan office to increase collaboration. Using social facilitation and distraction-conflict theory, advise which teams should and should not move, and propose metrics to evaluate the change.
5. Triplett's original findings appear weaker on reanalysis, and the overall effects of social presence are small. Why might a theory with small average effects still be useful to managers?
6. How would you measure a spiral of silence inside an organization? Specify the data you would collect and the pattern that would indicate a spiral rather than genuine consensus.

22.11 Further Reading

- Festinger (1954): the founding statement of social comparison theory, still worth reading for its compact, propositional style.
- Tajfel and Turner (1979): the classic integrative statement of social identity theory and its identity-management strategies.
- Brewer (1991): a short, influential article that reframes social identity as a balance between belonging and uniqueness.
- Chan, Berger, and Boven (2012): shows how consumers satisfy assimilation and differentiation motives on different dimensions of the same choice.
- Cialdini and Goldstein (2004): an authoritative review of compliance and conformity organized around the goals of accuracy, affiliation, and self-concept.
- Schultz et al. (2007): a compact field experiment demonstrating the boomerang effect of descriptive norms and its cure.
- C. F. Bond and Titus (1983): the meta-analysis that put the social facilitation literature in quantitative perspective.
- Noelle-Neumann (1974): the original statement of the spiral of silence. Read alongside Glynn, Hayes, and Shanahan (1997) for the meta-analytic evidence.

Chapter 23

Judgment and Decision Making

In the 1970s, the credit card industry lobbied hard over a question that a standard economic model would regard as trivial. If merchants were to be allowed to charge card users more than cash customers, should the price difference be called a "credit card surcharge" or a "cash discount"? The two labels describe exactly the same prices, yet the industry insisted on the second (R. Thaler 1980). Its instinct was sound. Paying a surcharge feels like a loss. Forgoing a discount feels like a missed gain, and people weigh losses more heavily than equivalent gains. The same insight explains why employees who are automatically enrolled in a retirement plan tend to stay enrolled, why a customer who has lived with a sofa for thirty days rarely returns it, and why a $50 gift card is spent differently from $50 in cash.

This chapter introduces the theories that explain how people *actually* judge and decide, as opposed to how an idealized rational agent would. We begin with **bounded rationality**, the idea that decision makers satisfice rather than optimize, and then turn to the **heuristics-and-biases** programme and its principal rival, **fast-and-frugal heuristics**. A brief overview of **prospect theory** **(Chapter 24)** (treated in full in Chapter 24) leads to two close relatives: the **endowment effect** and **mental accounting**. We then examine the cognitive architecture behind these phenomena (**dual-process theory**) and three theories about how subjective experience feeds into judgment: **feelings-as-information**, **processing fluency**, and the **discrepancy-attribution hypothesis**. We close with **choice architecture**, the applied science of designing decision environments, which draws on all of the preceding ideas (Table 23.1).

Table 23.1. Chapter at a glance: theories of judgment and decision making.

Theory	Core question	Level of analysis	Key source
Bounded rationality	How do decision makers choose when information and computation are limited?	Individual / organization	Simon (1955)
Heuristics and biases	Which mental shortcuts do people use, and when do they mislead?	Individual	Tversky & Kahneman (1974)
Prospect theory	How do people evaluate risky prospects relative to a reference point?	Individual / market	Kahneman & Tversky (1979)
Endowment effect	Why do owners value goods more than buyers do?	Individual / dyad	Kahneman, Knetsch & Thaler (1990)
Mental accounting	How do people categorize, track, and evaluate money?	Individual / household	Thaler (1985, 1999)

Table 23.1. Chapter at a glance: theories of judgment and decision making. *(continued)*

Theory	Core question	Level of analysis	Key source
Dual-process theory	How do intuitive and deliberative processes interact?	Individual	Evans & Stanovich (2013)
Feelings-as-information	When do feelings serve as evidence for judgments?	Individual	Schwarz & Clore (1983)
Processing fluency	How does the ease of processing shape liking, truth, and choice?	Individual	Reber, Schwarz & Winkielman (2004)
Discrepancy-attribution	Why does surprising fluency feel like familiarity?	Individual	Whittlesea & Williams (1998)
Choice architecture	How does the design of the choice environment shape decisions?	Individual / organization / policy	Thaler & Sunstein (2008)

23.1 Bounded Rationality

THEORY CARD

Core claim. Real decision makers have limited information, attention, and computational capacity. Instead of maximizing expected utility, they search sequentially and stop when they find an option that meets an aspiration level : they **satisfice**.
Origins. Economics, political science, and cognitive science, Simon (1955) and Simon (1956), developed for organizations by Cyert and March (1963).
Level of analysis. Individual decision makers and organizations.
Key constructs. Aspiration level, satisficing, sequential search, search costs, procedural (vs. substantive) rationality, task environment.
Boundary conditions. Matters most when options are numerous, information is costly to gather, and outcomes are uncertain. Least when options are few, attributes are transparent, and computation is cheap.
Related theories. Heuristics and biases, Dual-process theory, Choice architecture, Transaction cost economics (Section 6.3).

23.1.1 Origins and intellectual history

By the early 1950s, economics had settled on a model of the decision maker as an agent with complete and consistent preferences who chooses the alternative that maximizes expected utility. Simon (1955) did not dispute that people pursue goals. He disputed that they could possibly do so in the manner the model assumed. Global rationality requires knowing all alternatives, the consequences of each, and the probabilities of those consequences, and then performing the computations needed to compare them. Simon proposed instead a "behavioral model" in which the organism has a limited set of alternatives in view, a simplified payoff function that classifies outcomes as satisfactory or unsatisfactory, and an aspiration level that adjusts with experience.

In a companion paper, Simon (1956) introduced the term **satisficing** and emphasized that rationality is a property of the fit between a mind and its environment. An organism foraging in a sparse environment does not need an elaborate optimization routine if the structure of the environment allows a simple rule to achieve its goals. Simon's often-quoted image of rationality as a pair of scis-

sors, with one blade being the structure of the task environment and the other the computational capabilities of the actor, captures this point (discussed in Gigerenzer and Gaissmaier 2011).

The idea migrated quickly into organization theory. Cyert and March (1963) modelled firms as coalitions that set aspiration levels for goals such as sales and profit, engage in **problemistic search** when performance falls below aspiration, and accumulate **organizational slack** when performance is comfortably above it. In psychology, Payne, Bettman, and Johnson (1993) showed that individuals select among decision strategies (weighted-additive rules, lexicographic rules, elimination by aspects) by trading off cognitive effort against accuracy, making the choice of strategy itself an adaptive decision.

23.1.2 Core logic and propositions

DEFINITION

Definition 23.1. Satisficing. A decision procedure in which alternatives are examined sequentially and the first alternative whose value meets or exceeds an **aspiration level** is chosen. The aspiration level rises when satisfactory alternatives are easy to find and falls when they are hard to find (Simon 1955, 1956).

The theory can be summarized in four propositions:

- **P1 (Limited search).** Decision makers evaluate a subset of the available alternatives, and the size of that subset decreases as search costs rise.
- **P2 (Aspiration-based stopping).** Search stops when an alternative meets the current aspiration level, not when the best alternative has been found.
- **P3 (Adaptive aspirations).** Aspiration levels adjust to experience: they rise after easy success and fall after repeated failure (Simon 1955, Cyert and March 1963).
- **P4 (Procedural rationality).** Whether a decision rule is "good" depends on its fit with the structure of the task environment, not on its resemblance to optimization (Simon 1956, Gigerenzer and Gaissmaier 2011).

23.1.3 Mechanism

Satisficing is not merely a concession to cognitive weakness: when search is costly it can be the *better* policy. Figure 23.1 simulates a decision maker who inspects up to 50 options one at a time, pays a fixed cost per inspection, and stops at the first option that meets the aspiration level. The net payoff (value of the chosen option minus total search cost) peaks at an interior aspiration level, and the optimal aspiration falls as search becomes more expensive.

23.1.4 Evidence and extensions

Bounded rationality is less a single testable hypothesis than a research programme, and its influence is visible in at least three literatures. First, behavioural theories of the firm built on Cyert and March (1963) study how performance relative to aspirations drives search, risk taking, and organizational change. Second, research on adaptive decision making (Payne, Bettman, and Johnson 1993) documents how people shift from compensatory strategies (which trade off attributes) to non-compensatory strategies (which screen on one attribute at a time) as the number of alternatives and the time pressure increase. Third, Simon's emphasis on the fit between mind and envi-

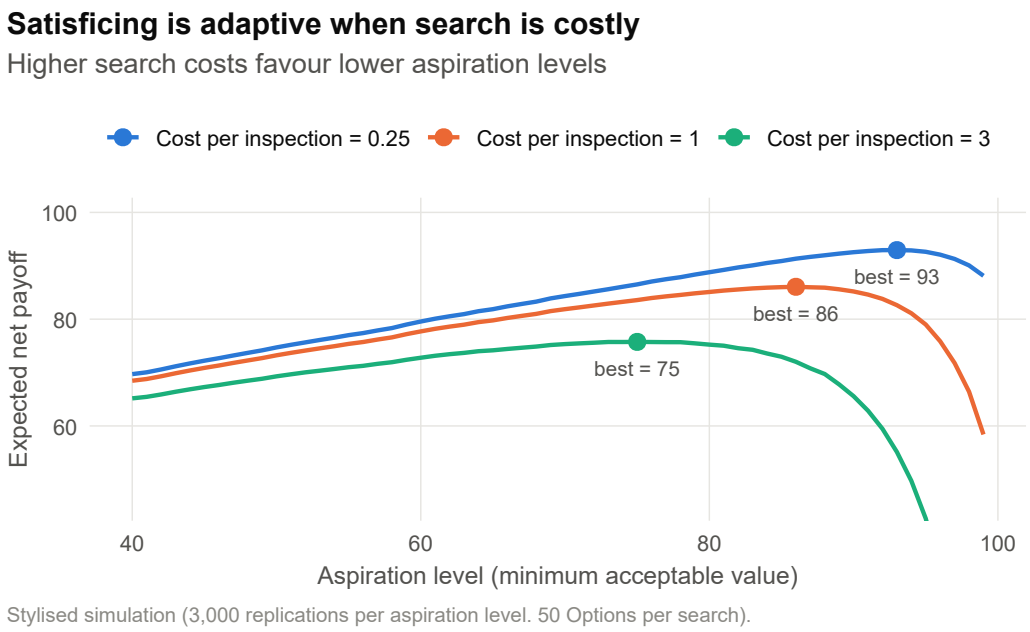


Figure 23.1. Illustrative simulation of satisficing. A decision maker inspects options sequentially (values drawn from Uniform[0, 100]) and stops at the first option meeting the aspiration level. Net payoff peaks at an interior aspiration level that falls as the cost of inspecting each option rises.

ronment became the foundation of the **ecological rationality** programme of Gigerenzer and colleagues, discussed in the next section (Gigerenzer and Goldstein 1996, Gigerenzer and Gaissmaier 2011).

BUSINESS CASE

Procurement and hiring as satisficing problems. Purchasing managers rarely solicit bids from every possible supplier, and recruiters rarely interview every qualified applicant. Instead, organizations codify aspiration levels as qualification thresholds ("pre-approved vendor", "minimum years of experience") and stop searching once a candidate clears them. Bounded rationality reframes the managerial question: rather than asking whether the chosen supplier was optimal, ask whether the threshold and the search budget were well calibrated to the cost of search and the variance of quality in the market. When the variance of quality is high, a longer search pays. When it is low, fast satisficing is efficient.

COMMON PITFALL

Equating bounded rationality with irrationality. Simon's argument was that decision rules should be evaluated relative to the decision maker's constraints and environment. A study that documents a departure from expected-utility maximization has not thereby shown that people are "boundedly rational" in Simon's sense. It must also show *why* the observed procedure is a sensible response to limited information or computation.

23.2 Heuristics and Biases

THEORY CARD

Core claim. People judge probabilities and frequencies by relying on a small number of heuristics. These heuristics are generally useful but produce systematic and predictable errors (biases).
Origins. Cognitive psychology, Tversky and Kahneman (1973) and Tversky and Kahneman (1974), later reformulated as **attribute substitution** by Kahneman (2003).
Level of analysis. Individual judgment.
Key constructs. Representativeness, availability, anchoring and adjustment, base-rate neglect, conjunction fallacy, attribute substitution.
Boundary conditions. Biases are largest when judgments are intuitive, uncertain, and made without feedback. They shrink when information is presented in natural frequencies, when incentives and expertise are aligned with accuracy, and when the task environment matches the heuristic's assumptions.
Related theories. Bounded rationality, Dual-process theory, Anchoring effect (Section 29.9), Feelings-as-information.

23.2.1 Core logic

Tversky and Kahneman (1974) proposed that three heuristics account for a wide range of intuitive judgments under uncertainty. Each substitutes an easy question for a hard one. Table 23.2 summarizes them.

Table 23.2. The three heuristics of Tversky and Kahneman (1974).

Heuristic	Hard question	Easier substitute	Typical bias	Business illustration
Representativeness	How probable is it that A belongs to category B?	How similar is A to the prototype of B?	Base-rate neglect, conjunction fallacy, insensitivity to sample size	Investors infer a great company from a great story, ignoring base rates of start-up failure
Availability	How frequent or likely is event X?	How easily do instances of X come to mind?	Overweighting vivid, recent, or publicized events	Insurance demand rises after a widely reported disaster
Anchoring and adjustment	What is the value of quantity Q?	Start from a salient number and adjust (insufficiently)	Estimates biased toward arbitrary anchors	List prices and first offers anchor negotiated prices

Representativeness produces the **conjunction fallacy**, in which a detailed, coherent description is judged more probable than a more general one that contains it (Tversky and Kahneman 1983). Availability treats ease of recall as evidence of frequency (Tversky and Kahneman 1973). Anchoring lets even arbitrary numbers, such as the result of a spin of a wheel of fortune, shift estimates (Tversky and Kahneman 1974). See Anchoring effect (Section 29.9).

Kahneman (2003) later unified these heuristics under a single mechanism, **attribute substitution**: when a target attribute is hard to assess, people answer on the basis of a related heuristic attribute that comes more readily to mind, often without noticing the substitution. The same reformulation links heuristics to the dual-process architecture discussed later in this chapter.

23.2.2 The fast-and-frugal alternative

The heuristics-and-biases programme defined good judgment by the norms of probability theory and studied heuristics mostly as sources of error. A rival programme, led by Gerd Gigerenzer, argues that heuristics should be studied as **adaptive tools** whose accuracy depends on the environment. Gigerenzer and Goldstein (1996) proposed the **Take-the-Best** algorithm for inferring which of two objects scores higher on a criterion (e.g., which of two German cities has the larger population): search cues in order of validity and stop at the first cue that discriminates. In their simulations this one-reason rule matched or exceeded the accuracy of more complex strategies such as multiple regression, while using far less information. A related **recognition heuristic** infers that a recognized object has the higher value, which can yield a "less-is-more" effect in which partial ignorance improves inference (D. G. Goldstein and Gigerenzer 2002) (Table 23.3).

Table 23.3. Two research programmes on heuristics.

Dimension	Heuristics and biases	Fast-and-frugal heuristics
Key sources	Tversky & Kahneman (1974), Kahneman (2003)	Gigerenzer & Goldstein (1996), Gigerenzer & Gaissmaier (2011)
Normative benchmark	Probability theory and expected utility	Ecological rationality: success in the actual environment
View of heuristics	Useful but error-prone shortcuts	Adaptive tools that can outperform complex models
Typical method	Word problems revealing deviations from norms	Simulation and prediction in real-world data
Characteristic finding	Systematic biases (base-rate neglect, anchoring)	Less-is-more effects, one-reason decisions
Managerial lesson	De-bias decisions, use statistical aids	Match simple rules to stable, predictable environments

DEBATE AND EVIDENCE

Cognitive illusions or adaptive tools? The two programmes clashed directly in a 1996 exchange in *Psychological Review*. Gigerenzer argued that many "biases" disappear when problems are presented in frequency formats and that heuristics were too vaguely specified to be tested. Kahneman and Tversky (1996) replied that cognitive illusions are real and robust, that frequency formats reduce but do not eliminate them, and that their programme had never claimed people are generally irrational. Much of the disagreement concerns the benchmark. If the standard is coherence with probability theory, heuristics produce errors. If the standard is predictive accuracy in a particular environment, simple heuristics can be superior, especially when data are scarce and noisy (Gigerenzer and Gaissmaier 2011). A productive synthesis asks *which* heuristic is used, *when*, and *in which environments* it succeeds or fails.

23.2.3 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Analysts' earnings forecasts will be more strongly anchored on the prior consensus when uncertainty about the firm is high.
- *H2.* Media coverage of product recalls will increase consumers' perceived risk of the

product category more than objective recall frequency does (availability).
- *H3.* In stable demand environments, a simple lexicographic sourcing rule will predict supplier performance as accurately as a multi-attribute scoring model, but not in volatile environments.

Designs. Vignette and incentive-compatible experiments. Field data with exogenous anchors (e.g., random list prices, lottery numbers). Out-of-sample prediction contests comparing heuristics with statistical models.
Measures. Deviation from Bayesian or actuarial benchmarks, anchoring indices, recall ease, cue-use patterns from process tracing.

23.3 Prospect Theory in Brief

THEORY CARD

Core claim. People evaluate outcomes as gains and losses relative to a reference point. They are risk averse for gains, risk seeking for losses, and losses loom larger than gains. Probabilities are transformed into decision weights that overweight small probabilities and underweight moderate and large ones.
Origins. Psychology and economics, Kahneman and Tversky (1979), cumulative version in Tversky and Kahneman (1992), framing in Tversky and Kahneman (1981) and Kahneman and Tversky (1984).
Level of analysis. Individual choice, with aggregate implications for markets (finance, insurance, labour, pricing).
Key constructs. Reference point, value function, loss aversion, diminishing sensitivity, probability weighting, framing, editing.
Boundary conditions. Requires an identifiable reference point. Loss aversion is weaker or absent for goods held for exchange, such as money spent in routine transactions (Novemsky and Kahneman 2005).
Related theories. Prospect theory and reference-dependent preferences (full treatment) (Chapter 24), Endowment effect, Mental accounting, Choice architecture.

Expected utility theory (Section 24.1) assumes that people evaluate prospects by their effect on final wealth and weight outcomes by their probabilities. Kahneman and Tversky (1979) documented choice patterns that violate both assumptions, among them the **certainty effect** (outcomes that are certain are overweighted relative to outcomes that are merely probable) and the **reflection effect** (preferences between gambles reverse when gains are replaced by losses). Their alternative evaluates prospects with a **value function** defined over gains and losses, x, relative to a reference point:

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -\lambda(-x)^{\beta} & \text{if } x < 0 \end{cases}$$

where $\alpha, \beta < 1$ capture **diminishing sensitivity** and $\lambda > 1$ captures **loss aversion (Section 24.4)**. Tversky and Kahneman (1992) estimated $\alpha = \beta = 0.88$ and $\lambda = 2.25$: a loss is weighted a little more than twice as heavily as a gain of the same size (Figure 23.2). Outcomes are then weighted by

an inverse-S **probability weighting function**, and because the reference point is malleable, logically equivalent descriptions of the same problem can produce different choices (Tversky and Kahneman 1981, Levin, Schneider, and Gaeth 1998).

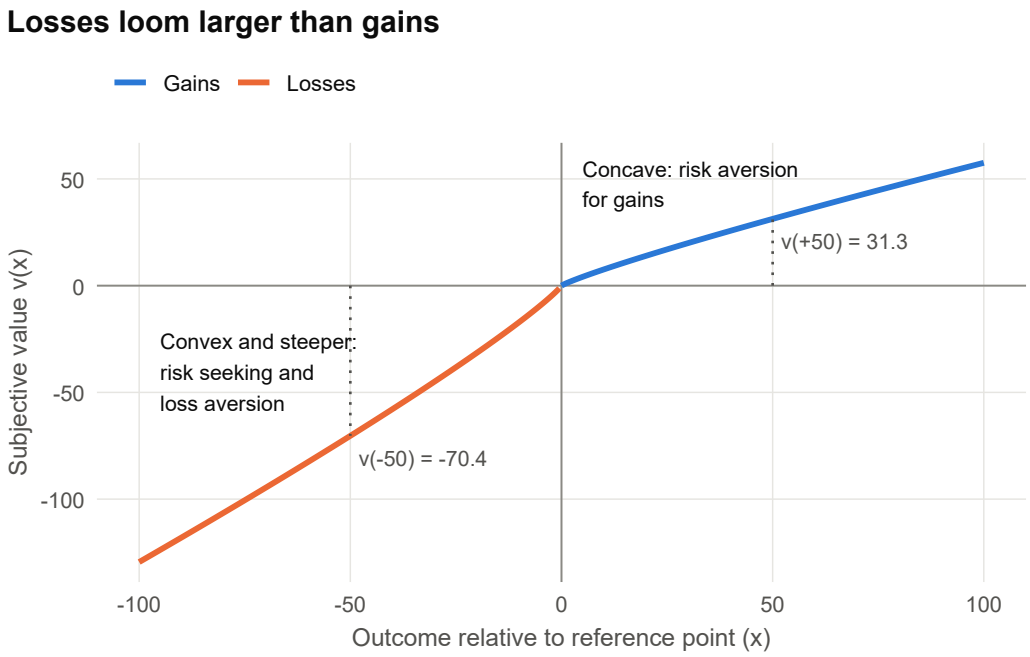


Figure 23.2. The prospect theory value function with the median parameter estimates of Tversky and Kahneman (1992): alpha = beta = 0.88 and lambda = 2.25. The function is concave for gains, convex for losses, and steeper for losses than for gains.

Prospect theory is the theoretical engine behind the two sections that follow: the endowment effect applies loss aversion to ownership, and mental accounting applies the shape of the value function to the way people group gains and losses. Because the theory is so central to the behavioural sciences, it receives a chapter of its own. Chapter 24 (Chapter 24) develops it in full: the expected utility benchmark and its paradoxes, the original and cumulative versions of the model, the debate over loss aversion, expectation-based reference points, status quo bias (Section 24.6), myopic loss aversion (Section 24.7), the disposition effect (Section 24.8), applications across business and policy, and the evidence from large-scale replications.

23.4 Endowment Effect

THEORY CARD

Core claim. People demand more to give up an object they own than they would pay to acquire the same object.

Origins. Behavioural economics, named by R. Thaler (1980), definitive experiments by Kahneman, Knetsch, and Thaler (1990), reviewed by Kahneman, Knetsch, and Thaler (1991) and Morewedge and Giblin (2015).

Level of analysis. Individual valuation, dyadic exchange (buyer-seller).

Key constructs. Willingness to accept (WTA), willingness to pay (WTP), ownership, reference price (Section 27.11), loss aversion, transaction disutility.
Boundary conditions. Weak or absent for goods held for resale or exchange (money, tokens, inventory) and among experienced traders. Sensitive to elicitation procedures (Plott and Zeiler 2005).
Related theories. Prospect theory (Chapter 24), Mental accounting, Appraisal theory (Section 25.5).

23.4.1 Core logic

In the experiments of Kahneman, Knetsch, and Thaler (1990), students in a classroom market were randomly given a coffee mug bearing the university logo. Because the mugs were assigned at random, standard theory predicts that about half of them should change hands when trading opens: the mugs should flow to the students who value them most. Instead, trading volume was far below this prediction, and the median selling price demanded by owners was roughly twice the median price buyers were willing to pay. When the same procedure was run with induced-value tokens that could be redeemed for cash, no such gap appeared, showing that the effect is not a by-product of the market institution itself.

The classic explanation is loss aversion: giving up a mug is coded as a loss, acquiring one as a gain, and losses loom larger (Kahneman, Knetsch, and Thaler 1991). Ownership shifts the reference point, so the same object is valued differently depending on which side of the transaction one is on.

23.4.2 Competing accounts

The loss-aversion explanation is not the only one. Carmon and Ariely (2000) show that buyers and sellers focus on what they would forgo: sellers attend to the object, buyers to the money. Weaver and Frederick (2012) propose a **reference-price theory** in which the endowment effect reflects an aversion to *bad deals* rather than to losing the object itself. Sellers do not want to be "suckers" who sell for less than a salient reference price, such as the market price. Buyers do not want to pay more than the object is worth to them or more than they think it usually costs. On this account the WTA-WTP gap arises from **transaction disutility** relative to reference prices, and it should vary with the information people have about those prices. Weaver and Frederick (2012) report evidence that the size of the gap changes with reference-price information in the ways the theory predicts, a pattern that a pure loss-aversion account does not anticipate. Figure 23.3 contrasts the two mechanisms.

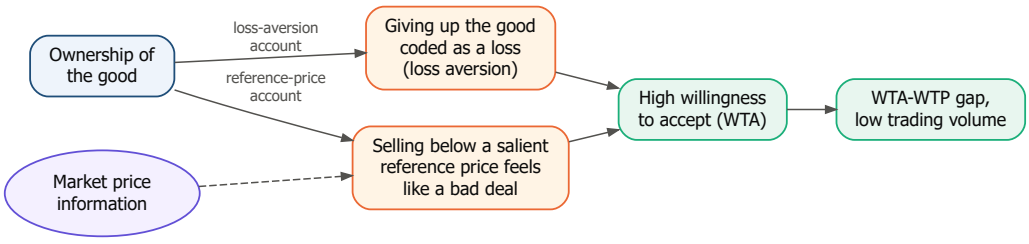


Figure 23.3. Two mechanisms for the endowment effect. In the loss-aversion account, ownership shifts the reference point so that giving up the good is coded as a loss. In the reference-price account (Weaver & Frederick, 2012), owners resist transactions that would be bad deals relative to salient market prices.

DEBATE AND EVIDENCE

Is the endowment effect an artefact? Plott and Zeiler (2005) argued that WTA-WTP gaps can be produced by subjects' misconceptions about elicitation procedures and showed that, with extensive training, anonymity, and practice rounds, the gap for mugs could be eliminated in their experiments. Defenders reply that elaborate procedures may remove the psychological conditions (such as a sense of ownership) that produce the effect in everyday life. An integrative review by Morewedge and Giblin (2015) concludes that no single mechanism explains all findings: loss aversion, ownership-based self-association, and reference-price considerations each contribute in different settings.

23.4.3 Applications in business

The endowment effect implies that anything which creates a **sense of ownership** before purchase raises the customer's valuation. Merely touching a product increases perceived ownership and valuation (Peck and Shu 2009). Free trials, test drives, configurators that let customers "build their own", and virtual try-on tools all exploit this mechanism.

BUSINESS CASE

Return policies and at-home trials. Several direct-to-consumer furniture and mattress retailers popularized extended at-home trial periods with free returns. The endowment effect suggests why such policies may be less costly than they appear: once the product is in the home, the customer's reference point shifts, and returning it is coded as giving up something owned. At the same time, the reference-price account cautions that customers who learn of a lower market price will feel they made a bad deal, which increases returns. **Illustrative scenario:** a retailer considering a trial extension from 30 to 100 days should test whether the gain in conversions outweighs the cost of returns, and track whether return rates vary with competitors' price promotions, a pattern the reference-price account predicts.

TRY IT YOURSELF

Think of an item you own but rarely use. Write down the lowest price at which you would sell it today, then the highest price you would pay to buy an identical item if you did not own it. What reference prices came to mind for each answer? Which account of the endowment effect better describes your reasoning?

23.5 Mental Accounting

THEORY CARD

Core claim. People organize, label, and evaluate financial activities in separate mental accounts. Because money in different accounts is not treated as fungible, the framing and timing of gains, losses, and payments change choices.
Origins. Behavioural economics and marketing, R. Thaler (1980), R. Thaler (1985), and R. H. Thaler (1999).
Level of analysis. Individual and household, firm pricing and promotion.

Key constructs. Mental accounts and budgets, hedonic editing, acquisition vs. transaction utility, coupling of payments and consumption, choice bracketing, fungibility.
Boundary conditions. Strongest when money is labelled (gift cards, rebates, windfalls), when budgets are explicit, and when payment and consumption are separated in time.
Related theories. Prospect theory (Chapter 24), Endowment effect, Self-control (Section 20.4).

23.5.1 Core logic

R. Thaler (1985) combined prospect theory's value function with the observation that consumers code and combine outcomes in systematic ways. Because the value function is concave for gains and convex for losses, the way outcomes are grouped affects their total subjective value. This yields four principles of **hedonic editing**, summarized in Table 23.4.

Table 23.4. Hedonic editing principles derived from mental accounting (Thaler, 1985).

Principle	Situation	Prescription	Why (value function)	Business application
Segregate gains	Two gains	Present them separately	Concavity: v(x) + v(y) > v(x + y)	Announce product benefits one by one, split a bonus into components
Integrate losses	Two losses	Combine them	Convexity: v(-x) + v(-y) < v(-(x + y))	All-inclusive pricing, bundled fees
Cancel small losses	Large gain, small loss	Integrate the loss into the gain	Small loss falls on the steep part of the loss curve	Deduct fees from a refund rather than charging them separately
Silver lining	Large loss, small gain	Segregate the small gain	Small gain falls on the steep part of the gain curve	Cash rebates on expensive purchases

Thaler also distinguished two kinds of utility in a purchase. **Acquisition utility** is the value of the good relative to its price. **Transaction utility** is the perceived merit of the deal itself: the difference between the price paid and a reference price. In his well-known illustration, people say they would pay more for the same beer if it were bought from a fancy resort hotel than from a run-down grocery store, because the reference price differs (R. Thaler 1985). Transaction utility links mental accounting to the reference-price account of the endowment effect (Weaver and Frederick 2012).

R. H. Thaler (1999) later organized the theory around three components: (1) how outcomes are perceived and experienced, including hedonic editing. (2) How activities are assigned to specific accounts, including budgets for categories of spending, and (3) how frequently accounts are evaluated, which relates to **choice bracketing**. Budgeting can lead to underconsumption when a category budget is exhausted and to overconsumption of items that are not easily categorized (Heath and Soll 1996).

23.5.2 Payment and consumption

Prelec and Loewenstein (1998) proposed that payments produce an immediate "pain of paying" and that consumers prefer to **decouple** payment from consumption. Prepaid vacations feel free at the time of consumption. Per-use fees spoil enjoyment. The flip side is **payment depreciation**: the psychological cost of a past payment fades over time. In a study of health-club members, Gourville and Soman (1998) found that attendance was highest in the months immediately following payment and declined as the payment receded, suggesting that members consumed to "use up"

a salient sunk cost.

23.5.3 Applications in business

BUSINESS CASE

Promotions, gift cards, and labelled money. If money were fungible, a $10 coupon for a grocer would be equivalent to $10 in cash and would raise grocery spending only slightly. Using data from an online grocer, Milkman and Beshears (2009) found that coupon redemption increased grocery spending by more than fungibility would predict and shifted purchases toward items that customers did not usually buy. Gift cards, store credit, and loyalty points work similarly: they create a dedicated account whose balance feels like a windfall and is spent more freely. Thaler's hedonic editing principles add a second lesson: rebates are an effective "silver lining" on large purchases, while fees are best integrated into a single price. Consistent with the latter point, partitioned prices (a base price plus a separate surcharge, such as shipping) can lower consumers' recalled total cost and increase demand (Morwitz, Greenleaf, and Johnson 1998), although such practices raise transparency concerns.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Customers who pay for an annual membership upfront will use the service more in the first month after payment than in later months (payment depreciation).
- *H2.* A promotional credit labelled for a product category will increase spending in that category more than an unlabelled credit of equal value.
- *H3.* Consumers will prefer an all-inclusive price to a partitioned price with the same total when the surcharge is large relative to the base price.

Designs. Field experiments on promotion labels, transaction data around payment dates, scanner panels with windfalls such as rebates.
Measures. Category spending, marginal propensity to consume out of labelled vs. unlabelled funds, usage relative to payment timing.

23.6 Dual-Process Theory

THEORY CARD

Core claim. Judgment and choice reflect two qualitatively different kinds of processing: fast, autonomous **Type 1** processes and slow, effortful **Type 2** processes that require working memory.
Origins. Cognitive and social psychology, Stanovich and West (2000) coined the System 1 / System 2 labels, J. St. B. T. Evans and Stanovich (2013) provide the canonical defence, popularized by Kahneman (2011).
Level of analysis. Individual cognition.
Key constructs. Type 1 (intuitive) and Type 2 (reflective) processing, working memory, cognitive decoupling, default-interventionist architecture, cognitive reflection.
Boundary conditions. Type 2 intervention requires capacity, motivation, and a cue that the intuitive response may be wrong. Under time pressure, cognitive load, or low motivation, Type

1 responses dominate.
Related theories. Heuristics and biases, Elaboration likelihood model (Section 26.1), Self-control (Section 20.4).

23.6.1 Core logic

Dual-process theories hold that people have two ways of arriving at a judgment. J. St. B. T. Evans and Stanovich (2013) argue that the *defining* feature of Type 2 processing is its reliance on working memory, which enables **cognitive decoupling**: the ability to hold a representation of the world apart from reality in order to simulate hypothetical alternatives. Type 1 processes, by contrast, are autonomous: they run whenever their triggering conditions are present. Speed, consciousness, and accuracy are *typical correlates* rather than defining features. Table 23.5 summarizes the distinction.

Table 23.5. Type 1 and Type 2 processing, following the distinction between defining features and typical correlates in Evans and Stanovich (2013).

Feature	Type 1 (intuitive)	Type 2 (reflective)
Defining feature	Does not require working memory, autonomous	Requires working memory, cognitive decoupling and mental simulation
Typical correlates	Fast, high capacity, parallel, often nonconscious	Slow, capacity-limited, serial, conscious
Basis	Associations, experience, emotion	Rules, abstraction, hypothetical thinking
Relation to ability	Largely independent of cognitive ability	Correlated with cognitive ability and thinking dispositions
Typical output	Default response, often heuristic	Intervention that may override the default
Popular label	System 1	System 2

The most widely used architecture is **default-interventionist**: Type 1 processes generate a rapid default response, which Type 2 processes may endorse, adjust, or override when there is sufficient capacity and a reason to doubt the default (Figure 23.4). A well-known individual difference measure is the Cognitive Reflection Test (Frederick 2005). Its best-known item asks: if a bat and a ball cost $1.10 in total and the bat costs $1.00 more than the ball, how much does the ball cost? The intuitive answer (10 cents) comes quickly to mind and is wrong. Reaching the correct answer (5 cents) requires noticing the error and checking.

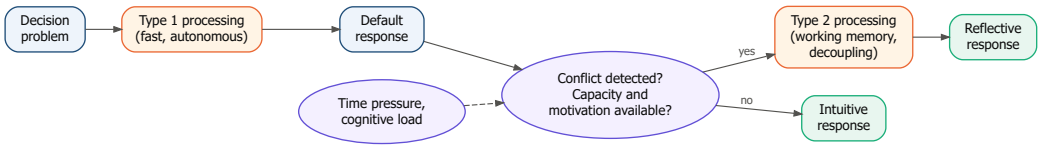


Figure 23.4. A default-interventionist model of dual processing. Type 1 processes produce a default response. Type 2 processes intervene only when capacity, motivation, and a conflict signal are present.

23.6.2 Evidence and extensions

Experimental evidence for the role of capacity comes from cognitive load manipulations. In a classic consumer study, participants who had to memorize a seven-digit number were more likely to choose chocolate cake over fruit salad than participants memorizing a two-digit number, consistent

with affective (Type 1) reactions dominating when deliberative resources are occupied (Shiv and Fedorikhin 1999). Performance on reasoning tasks also correlates with cognitive ability and thinking dispositions, a pattern Stanovich and West (2000) used to argue that normative and intuitive responses are generated by different processes.

DEBATE AND EVIDENCE

One system, two systems, or a continuum? Critics argue that the features attributed to the two systems do not reliably co-occur and that the theory is hard to falsify (Keren and Schul 2009). Kruglanski and Gigerenzer (2011) propose that intuitive and deliberate judgments are based on common rule-following principles and differ mainly in the difficulty of applying those rules. J. St. B. T. Evans and Stanovich (2013) accept some of these criticisms (the long lists of correlated features were misleading) but maintain that a qualitative distinction defined by working memory is supported by experimental, individual-difference, and neuroscientific evidence. For applied researchers, the safest course is to avoid treating "System 1" as an explanation in itself and instead to specify and measure the process (e.g., load, time pressure, conflict detection) that is assumed to operate.

BUSINESS CASE

Designing for both systems. Checkout pages, investment platforms, and insurance enrolment forms are often completed quickly and under distraction. Dual-process theory suggests two design levers. When the intuitive response is usually right (e.g., reordering a familiar product), reduce friction so that Type 1 processing can operate smoothly. When the intuitive response is likely to be wrong (e.g., choosing an insurance plan on the basis of the premium alone), insert a **reflection cue**: a summary of expected annual costs, a comparison table, or a brief "cooling-off" confirmation step.

23.7 Feelings-as-Information Theory

THEORY CARD

Core claim. People use their momentary feelings (moods, emotions, and metacognitive experiences) as information when making judgments, asking implicitly "How do I feel about it?" The impact of a feeling depends on its perceived informational value.
Origins. Social psychology, Schwarz and Clore (1983), reviewed in Schwarz (2012), consumer applications in Pham (1998) and Pham (2004) ("the logic of feeling"). Also known as the *affect-as-information* hypothesis.
Level of analysis. Individual judgment.
Key constructs. Mood, integral vs. incidental affect, misattribution, discounting, perceived relevance (representativeness) of feelings, processing style.
Boundary conditions. Feelings influence judgment when they are perceived as relevant to the target. Their influence is reduced or eliminated when people attribute them to an irrelevant source.
Related theories. Processing fluency, Discrepancy-attribution hypothesis, Appraisal theory (Section 25.5), Warm glow heuristic (Section 29.2).

23.7.1 Origins and core logic

Schwarz and Clore (1983) interviewed people by telephone about their life satisfaction on sunny or rainy spring days. Respondents reported higher satisfaction with their lives as a whole on sunny days. The decisive manipulation came next: when the interviewer casually asked about the local weather before the well-being questions, the effect of the weather disappeared. Drawing attention to an irrelevant cause of one's mood led people to **discount** the mood as information about their lives. The finding established a central principle: feelings are used as information unless their informational value is called into question.

Schwarz (2012) summarizes the theory in a set of postulates:

- **P1.** People attend to their feelings as a source of information, and different feelings convey different information (e.g., a positive mood signals a benign situation. Fear signals threat).
- **P2.** The impact of a feeling depends on its **perceived informational value**. When a feeling is attributed to a source unrelated to the target, its influence is attenuated or eliminated (misattribution and discounting).
- **P3.** Feelings are experienced as reactions to whatever is in the focus of attention, which is why incidental feelings are easily misread as reactions to the target.
- **P4.** Feelings inform processing style. Negative affect signals a problematic situation and fosters detail-oriented, systematic processing. Positive affect signals a benign situation and fosters reliance on heuristics and existing knowledge.
- **P5.** Feelings are more likely to be used when they seem *representative* of and *relevant* to the target (Pham 1998) and when other information is scarce or the judgment is made quickly (Greifeneder, Bless, and Pham 2011) (Figure 23.5).

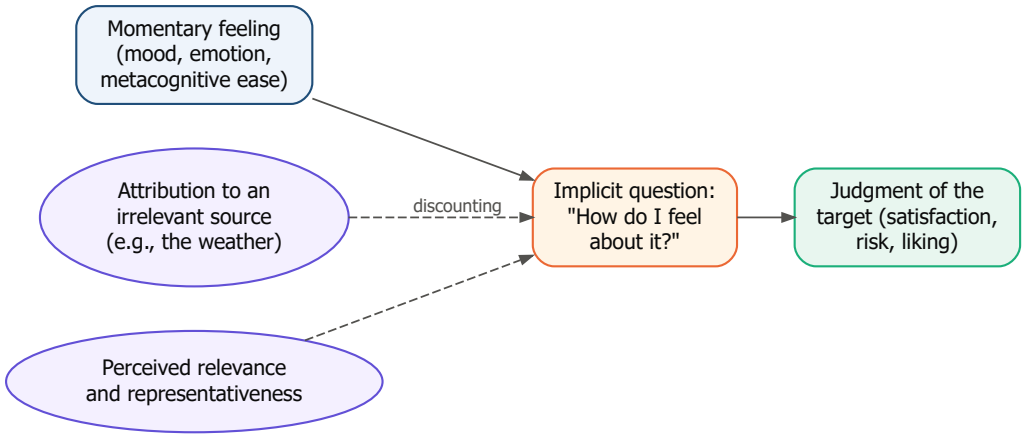


Figure 23.5. The feelings-as-information mechanism. Incidental or integral feelings are consulted as information about the target. The effect disappears when the feeling is attributed to an irrelevant source.

23.7.2 Evidence and extensions

Feelings-as-information has been extended in three important directions.

1. **Metacognitive experiences.** Not only moods but also feelings of ease or difficulty serve as information. In a classic study, participants who recalled six examples of their own assertive behaviour rated themselves as *less* assertive than those who recalled twelve, because recalling

twelve examples felt difficult and the difficulty was taken as evidence (Schwarz et al. 1991). This **ease-of-retrieval** effect refines the availability heuristic.

2. **Risk and affect.** The "risk-as-feelings" hypothesis holds that emotional reactions to risky situations often diverge from cognitive assessments and drive behaviour (Loewenstein et al. 2001). The related **affect heuristic** proposes that people consult an affective "pool" of positive and negative tags when judging risks and benefits (Slovic et al. 2007).
3. **Consumer judgment.** Pham (1998) showed that consumers rely on their feelings more when their motives are experiential (consummatory) than when they are instrumental, and Pham (2004) argued that reliance on feelings follows a coherent "logic": feelings are a fast, often valid summary of one's preferences.

BUSINESS CASE

Sunshine, sports, and stock prices. If moods are misread as information, mood-altering events that are irrelevant to fundamentals should leave a trace in market prices. Across 26 stock exchanges, morning sunshine in the city of a country's leading exchange was positively associated with daily market returns (Hirshleifer and Shumway 2003). Similarly, national stock markets have been shown to decline after the country's team loses an important international soccer match (Edmans, García, and Norli 2007). For managers, the lesson is not to trade on the weather but to recognize that customers, investors, and employees bring incidental feelings to every judgment. Retail atmospherics, music, and service interactions affect evaluations of the product itself, partly through misattributed mood.

COMMON PITFALL

Confusing mood congruence with information use. Many studies show that people in good moods make more positive judgments, but this alone does not distinguish feelings-as-information from memory-based accounts in which mood primes congruent material. The diagnostic test is **misattribution**: if the mood effect disappears when people are led to attribute their feelings to an irrelevant source, the feeling was being used as information (Schwarz and Clore 1983, Schwarz 2012).

23.8 Processing Fluency

THEORY CARD

Core claim. The subjective ease with which information is processed is itself a cue for judgment. Fluent processing is experienced as positive and increases judgments of liking, truth, familiarity, and confidence.

Origins. Cognitive and social psychology, L. L. Jacoby et al. (1989) on fluency misattribution, Reber, Schwarz, and Winkielman (2004) on aesthetic pleasure, reviewed by Alter and Oppenheimer (2009).

Level of analysis. Individual judgment.

Key constructs. Perceptual fluency, conceptual fluency, linguistic fluency, retrieval fluency, hedonic marking of fluency, prototypicality.

Boundary conditions. Fluency effects depend on how the experience is interpreted. They are reduced when people attribute fluency to an irrelevant source and can reverse when people

hold a lay theory linking difficulty to value.
Related theories. Feelings-as-information, Warm glow heuristic (Section 29.2), Mere exposure effect (Section 29.1), Beauty in averageness (Section 29.10).

23.8.1 Core logic

Reber, Schwarz, and Winkielman (2004) proposed that beauty is, in part, "in the perceiver's processing experience." Their account rests on four assumptions: objects differ in how fluently they can be processed. Fluency is **hedonically marked**, so high fluency feels good. This positive feeling is used as information when people judge their liking for the object, and the impact of fluency depends on expectations and attributions. Features that make stimuli easier to process (symmetry, figure-ground contrast, clarity, repetition, and prototypicality) should therefore increase aesthetic appreciation.

Prototypicality offers a clean test. Winkielman et al. (2006) showed that prototypical patterns are both processed more fluently and liked more, and that the fluency of processing accounts for part of the preference. Reber (2011) extends this reasoning to culturally shared taste, arguing that fluency explains why aesthetic preferences converge within groups that share perceptual experience. The same logic underlies the classic finding that averaged faces are attractive (see Beauty in averageness (Section 29.10)).

Fluency is not one thing. Alter and Oppenheimer (2009) catalogue a family of fluency experiences, summarized in Table 23.6, and argue that they have similar downstream effects regardless of source.

Table 23.6. Sources and consequences of processing fluency (organized following Alter & Oppenheimer, 2009).

Type of fluency	Source of ease	Illustrative effect
Perceptual	Font legibility, contrast, presentation duration	Hard-to-read instructions make a task seem more effortful (Song & Schwarz, 2008)
Linguistic	Pronounceability of names and words	Firms with easy-to-pronounce ticker names performed better shortly after listing (Alter & Oppenheimer, 2006)
Conceptual	Semantic priming, predictive contexts	Brands encountered in congruent contexts are evaluated more favourably
Retrieval	Ease of generating examples or arguments	Generating few reasons to choose a brand can improve evaluations more than generating many (Wanke et al., 1997)
Decision (preference)	Ease of forming a preference	Harder-to-read option descriptions increase choice deferral (Novemsky et al., 2007)
Memory (familiarity)	Prior exposure, repetition	Repeated names come to seem famous (Jacoby et al., 1989)

23.8.2 Fluency, liking, and familiarity: the warm glow

The link between fluency and affect runs in both directions. Fluency produces positive affect, but positive affect can also be mistaken for fluency and hence for familiarity. Monin (2003) called this the **warm glow heuristic**: attractive faces and positive words are more likely to be judged as previously seen, because the positive feeling they evoke is taken as a sign of familiarity. The phenomenon

is treated in detail in Warm glow heuristic (Section 29.2). Together, fluency and warm-glow effects explain why familiarity, liking, and perceived truth tend to travel together, and why marketers' efforts to make brands familiar can also make them liked.

23.8.3 Applications in business

BUSINESS CASE

Names, fonts, and arguments. Fluency research offers several well-documented design principles. First, **names matter**: shares of companies with easier-to-pronounce names and ticker symbols outperformed those with harder names in the days after listing, an effect that faded over time (Alter and Oppenheimer 2006). Second, **legibility matters** in both directions: instructions in a hard-to-read font make an exercise routine seem harder and reduce willingness to do it (Song and Schwarz 2008), and harder-to-process product descriptions increase choice deferral (Novemsky et al. 2007). Third, **asking for too much elaboration can backfire**: an advertisement that asked readers to generate many reasons to choose a BMW produced less favourable evaluations than one asking for a single reason, because generating many reasons felt difficult (Wänke, Bohner, and Jurkowitsch 1997). The common thread is that customers read their processing experience as information about the product.

TRY IT YOURSELF

Find two product pages for competing products in the same category. Rate each on perceptual fluency (layout, font, contrast) and linguistic fluency (product names, jargon). Predict which page produces higher perceived quality, and identify one change that would increase fluency without removing information.

23.9 Discrepancy-Attribution Hypothesis

THEORY CARD

Core claim. A feeling of familiarity arises not from fluent processing as such, but from processing that is *surprisingly* fluent relative to expectations. People attribute the unexpected fluency to prior experience.
Origins. Cognitive psychology of memory, Whittlesea (1993), Whittlesea and Williams (1998), elaborated in Whittlesea and Williams (2001a) and Whittlesea and Williams (2001b).
Level of analysis. Individual memory and judgment.
Key constructs. Processing fluency, expectation, discrepancy (surprise), attribution to the past, feeling of familiarity.
Boundary conditions. The discrepancy must be attributed to the past rather than to an obvious current cause. When fluency is expected (as with a close friend), it does not produce a feeling of familiarity.
Related theories. Processing fluency, Feelings-as-information, Warm glow heuristic (Section 29.2).

23.9.1 Core logic

Why do strangers sometimes feel familiar, while close friends do not produce a "feeling of familiarity" at all? Whittlesea and Williams (1998) argued that familiarity is an **inference**. People continuously and implicitly evaluate the quality of their own processing against expectations generated by the context. When processing is more fluent than expected, the discrepancy calls for an explanation, and one readily available explanation is that the stimulus has been encountered before. When processing is fluent but fluency is expected, as with a friend seen in a familiar place, no discrepancy arises and no special feeling of familiarity is experienced. The classic everyday example is meeting one's butcher on a bus: the face is processed more fluently than a stranger's face should be, but the unusual context prevents identification, leaving a nagging feeling of familiarity.

In laboratory tests, Whittlesea and Williams (1998) found that pronounceable, word-like nonwords were especially prone to false recognition, more so than real words, because they were processed more fluently than nonwords are expected to be. Whittlesea and Williams (2001a) established that the feeling of familiarity depends on this heuristic evaluation of processing, and Whittlesea and Williams (2001b) showed that expectation and surprise play a causal role: fluent resolution of an uncertain context increases feelings of familiarity. Figure 23.6 summarizes the mechanism.

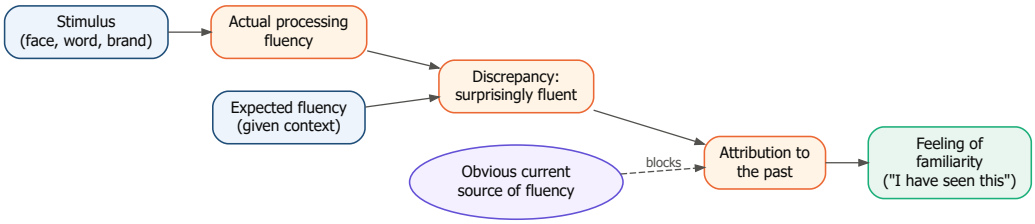


Figure 23.6. The discrepancy-attribution mechanism. Processing is continuously compared with context-based expectations. Surprisingly fluent processing is attributed to prior experience and felt as familiarity.

23.9.2 Implications

The discrepancy-attribution hypothesis sharpens the fluency account in a way that matters for practice. Fluency alone is not enough. What produces familiarity, and the warm feelings that tend to accompany it, is fluency that exceeds expectations. Three implications follow for marketing and communication:

- **Context matters.** A brand logo encountered in an unexpected context may feel more familiar (and be misremembered as previously seen) than the same logo in an expected context.
- **Lookalikes exploit discrepancy.** Copycat packaging that resembles a leading brand is processed more fluently than an unfamiliar brand should be, creating a sense of familiarity that consumers may misattribute to prior experience with the copycat itself.
- **Measurement artefacts.** Recognition-based advertising metrics (e.g., "Have you seen this ad?") are vulnerable to false familiarity for ads that are simply easy to process, so recognition scores should be benchmarked against foils of equal fluency.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* False recognition of a new brand name will be higher when the name follows the phonological patterns of the category than when it does not. *H2.* The famil-

iarity advantage of a copycat package will be reduced when consumers are given an explicit reason for its ease of processing (e.g., told that it resembles a known brand).
Designs. Recognition memory paradigms with lures that vary in fluency. Manipulations of contextual expectation (predictive vs. non-predictive contexts). Attribution manipulations.
Measures. Hit and false-alarm rates, signal-detection parameters, subjective familiarity ratings, response times as a fluency index.

23.10 Choice Architecture and Defaults

THEORY CARD

Core claim. Because decision makers are boundedly rational, loss averse, and influenced by how options are presented, the design of the choice environment (defaults, ordering, framing, number of options, feedback) systematically shapes choices without restricting options or changing economic incentives.
Origins. Behavioural economics and decision science, R. H. Thaler and Sunstein (2008), defaults in Madrian and Shea (2001) and E. J. Johnson and Goldstein (2003), tools in E. J. Johnson et al. (2012).
Level of analysis. Individual choice, organizational and public policy design.
Key constructs. Choice architect, nudge, default, active choice, choice overload, decision aids, libertarian paternalism.
Boundary conditions. Defaults have less effect when preferences are strong and well-formed, when stakes are salient, and when the default is perceived as a manipulative or self-serving recommendation.
Related theories. Bounded rationality, Prospect theory (Chapter 24), Dual-process theory, Social norms (Section 22.5).

23.10.1 Defaults

The most powerful tool of choice architecture is the **default**: the option that takes effect if the decision maker does nothing. E. J. Johnson and Goldstein (2003) compared organ-donor consent across European countries and found that effective consent rates were dramatically higher in countries with an opt-out ("presumed consent") default than in countries with an opt-in default. In an online experiment, the share of participants agreeing to be donors was roughly twice as high under an opt-out default as under an opt-in default, with a forced-choice condition producing rates close to the opt-out condition.

Three mechanisms explain default effects (Dinner et al. 2011, Jachimowicz et al. 2019):

1. **Effort (ease).** Accepting the default requires no action. Changing it requires time and attention.
2. **Endorsement.** Decision makers infer that the default reflects the recommendation of the choice architect.
3. **Endowment (reference dependence).** The default becomes the reference point, so switching away from it is coded as giving something up.

A meta-analysis of default studies found that defaults have a substantial average effect on choices, with considerable variation across domains, and that effects tend to be larger in consumer domains

and when defaults are perceived as endorsements (Jachimowicz et al. 2019).

BUSINESS CASE

Automatic enrollment in retirement plans. Madrian and Shea (2001) studied a large US firm that changed its 401(k) plan so that new employees were enrolled automatically unless they opted out. Participation among new hires rose sharply, but many automatically enrolled employees also stayed at the default contribution rate and in the default investment fund (a conservative money market fund), choices that few employees had made under the old opt-in regime. The study revealed both the power of defaults and their double edge: a default chosen for administrative convenience can lock employees into undersaving. The **Save More Tomorrow** programme (R. H. Thaler and Benartzi 2004) addressed this by asking employees to precommit to allocating a portion of *future* salary increases to savings. Because contributions rose only when pay rose, take-home pay never fell, sidestepping loss aversion. Participants' saving rates rose several-fold over the following years. Automatic enrollment and automatic escalation have since become widespread in employer-sponsored retirement plans.

23.10.2 The toolkit of the choice architect

R. H. Thaler and Sunstein (2008) organize choice architecture around six principles, summarized in Table 23.7 together with examples from business. E. J. Johnson et al. (2012) extend the list with tools for structuring choice sets (e.g., the number of alternatives, the use of decision aids) and for describing options (e.g., translating attributes into more meaningful units).

Table 23.7. Principles of choice architecture (Thaler & Sunstein, 2008), with underlying theories and business examples.

Principle	What it means	Underlying theory	Business example
Defaults	Pre-select the option that is best for most people	Status quo bias, endowment, endorsement	Auto-enrollment in savings plans, green-energy tariffs as default
Expect error	Design systems that anticipate and forgive mistakes	Bounded rationality	Undo buttons, reminders before a free trial converts
Give feedback	Tell people how they are doing	Learning under bounded rationality	Real-time energy or spending dashboards
Understand mappings	Help people translate choices into outcomes	Evaluability, attribute substitution	Total annual cost of ownership shown next to price
Structure complex choices	Simplify and organize large choice sets	Choice overload, satisficing	Filters, curated bundles, recommended plans
Incentives	Make costs and benefits salient	Prospect theory, mental accounting	Showing savings per month rather than per year

23.10.3 Choice overload

A widely cited experiment by S. S. Iyengar and Lepper (2000) set up a tasting booth for specialty jams in an upscale grocery store, displaying either 24 or 6 jams. The extensive display attracted more passers-by, but a far smaller share of those who stopped went on to buy (roughly 3 percent versus 30 percent). Parallel laboratory studies with chocolates and essay topics found that people offered fewer options were more satisfied and more motivated. The finding gave rise to the idea of

choice overload: beyond some point, additional options can reduce the likelihood of choosing and satisfaction with the choice.

DEBATE AND EVIDENCE

Is choice overload real? A meta-analysis of 50 experiments by Scheibehenne, Greifeneder, and Todd (2010) found a mean effect of choice-set size on choice outcomes that was close to zero, with large heterogeneity across studies: some found overload, others found that more choice helped. The authors concluded that choice overload, if it exists, depends on conditions that had not been well specified. A subsequent meta-analysis by Chernev, Böckenholt, and Goodman (2015) identified such conditions, finding that larger assortments are more likely to produce overload when the choice set is complex, the decision task is difficult, consumers are uncertain about their preferences, and they aim to minimize effort. The lesson for managers is not "offer fewer options" but "make large assortments easy to navigate", for example through categorization, recommendations, and good defaults.

DEBATE AND EVIDENCE

How large are nudge effects? A meta-analysis of more than 200 studies reported that choice architecture interventions have a small-to-medium average effect on behaviour, with interventions that change the decision structure (such as defaults) more effective than those that only change how information is described (Mertens et al. 2022). Re-analysing the same data with corrections for publication bias, Maier et al. (2022) found no evidence for an overall nudge effect, although they did not rule out effects in particular domains or for particular types of nudge. The exchange illustrates a general lesson: effects observed in published studies, especially small ones with flexible outcome measures, overstate what organizations should expect when they scale an intervention.

COMMON PITFALL

Nudging for the architect rather than the chooser. The ethical justification of choice architecture, "libertarian paternalism", requires that defaults and frames be designed to help people achieve their *own* goals while preserving freedom of choice (R. H. Thaler and Sunstein 2008). Using the same tools to make cancellation difficult, to enrol customers in add-ons they would not choose, or to obscure total prices exploits bounded rationality for the firm's benefit. Such practices invite regulatory scrutiny and erode trust. They also violate the reference-based fairness norms customers apply to firms (Kahneman, Knetsch, and Thaler 1986).

23.10.4 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Pre-selecting a premium service tier as the default will increase its uptake, and the effect will be larger among new customers than among experienced customers (preference strength as moderator).
- *H2.* Disclosing that a default was chosen because it is the most popular option will increase its uptake relative to an undisclosed default (endorsement mechanism).

- *H3.* Large assortments will reduce purchase likelihood only when products are difficult to compare on a common attribute.

Designs. Randomized field experiments that vary defaults or the number of options. Natural experiments from policy changes (e.g., auto-enrollment). Regression discontinuity around eligibility dates.

Measures. Opt-out rates, time spent on the choice, downstream outcomes (savings balances, retention, satisfaction), heterogeneity by preference strength.

23.11 Chapter Summary

KEY TAKEAWAYS

- **Bounded rationality** replaces optimization with satisficing: decision makers search until an option meets an aspiration level, which can be adaptive when search is costly (Simon 1955).
- **Heuristics** substitute easy questions for hard ones. They produce systematic biases (Tversky and Kahneman 1974) but can also be ecologically rational (Gigerenzer and Goldstein 1996). The two views differ mainly in their normative benchmark.
- **Prospect theory** explains choices through reference dependence, loss aversion, diminishing sensitivity, and probability weighting (Kahneman and Tversky 1979, Tversky and Kahneman 1992). Framing changes choices by moving the reference point.
- The **endowment effect** reflects both loss aversion and an aversion to bad deals relative to reference prices (Kahneman, Knetsch, and Thaler 1990, Weaver and Frederick 2012).
- **Mental accounting** shows that money is not fungible: labels, budgets, and the timing of payments shape spending (R. Thaler 1985, R. H. Thaler 1999).
- **Dual-process theory** distinguishes autonomous Type 1 processing from working-memory-dependent Type 2 processing (J. St. B. T. Evans and Stanovich 2013).
- **Feelings-as-information**, **processing fluency**, and **discrepancy-attribution** explain how moods and the ease of processing are read as evidence about the world (Schwarz and Clore 1983, Reber, Schwarz, and Winkielman 2004, Whittlesea and Williams 1998).
- **Choice architecture** applies these insights. Defaults are its most powerful tool, but published effects of nudges and choice overload are more heterogeneous and smaller than early studies suggested (E. J. Johnson and Goldstein 2003, Scheibehenne, Greifeneder, and Todd 2010, Maier et al. 2022).

Table 23.8 presents the comparison of the chapter's theories.

Table 23.8. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Typical business use
Bounded rationality	Satisficing against aspiration levels	Search stops at the first acceptable option	Search budgets, qualification thresholds
Heuristics and biases	Attribute substitution	Systematic deviations from probability norms	De-biasing forecasts and evaluations
Fast-and-frugal heuristics	Ecological fit of simple rules	Less-is-more effects	Simple decision rules in stable environments

Table 23.8. Comparison of the chapter's theories. *(continued)*

Theory	Key mechanism	Signature prediction	Typical business use
Prospect theory	Reference dependence, loss aversion, probability weighting	Risk aversion for gains, risk seeking for losses	Pricing frames, loss-framed messaging, insurance
Endowment effect	Ownership shifts reference point, aversion to bad deals	WTA exceeds WTP	Free trials, touch and try-on, return policies
Mental accounting	Non-fungible accounts, hedonic editing	Labelled money is spent differently	Gift cards, rebates, bundled pricing
Dual-process theory	Type 1 defaults, Type 2 intervention	Load and time pressure increase intuitive responses	Friction and reflection cues in interfaces
Feelings-as-information	Feelings used as evidence unless discounted	Mood effects vanish under misattribution	Atmospherics, timing of evaluations
Processing fluency	Ease of processing is hedonically marked	Fluent stimuli are liked and believed	Naming, design, readability
Discrepancy-attribution	Surprising fluency attributed to the past	False familiarity for unexpectedly fluent stimuli	Brand recognition metrics, lookalike packaging
Choice architecture	Defaults, structure, framing	Default options are disproportionately chosen	Enrollment, subscriptions, product configuration

23.12 Discussion Questions

1. Simon argued that rationality should be judged relative to the structure of the environment. Choose a managerial decision (e.g., hiring, supplier selection) and describe an environment in which a simple satisficing rule would outperform an elaborate scoring model, and one in which it would not.
2. Gigerenzer and Kahneman disagree about whether heuristics are "biased". What evidence would you need to decide which perspective better describes how loan officers evaluate small-business applications?
3. Prospect theory requires a reference point. For a subscription business that is raising prices, list three plausible reference points customers might use and design an experiment to identify which one drives churn.
4. Compare the loss-aversion and reference-price accounts of the endowment effect. What field data from a resale platform could distinguish them?
5. A retailer wants to increase spending by issuing store credit instead of cash refunds. Using mental accounting, predict the effect on spending and identify the ethical issues the retailer should consider.
6. Nudge meta-analyses disagree about average effects once publication bias is considered. How should an organization decide whether to scale a default change that worked in a pilot?

23.13 Further Reading

- Simon (1955): the founding statement of bounded rationality, still remarkably current in its critique of global rationality.

- Tversky and Kahneman (1974): the short *Science* article that launched the heuristics-and-biases programme.
- Gigerenzer and Gaissmaier (2011): a thorough review of the fast-and-frugal alternative and the concept of ecological rationality.
- Kahneman and Tversky (1979) and Tversky and Kahneman (1992): the original and cumulative versions of prospect theory. Read together to see how the theory was formalized.
- R. H. Thaler (1999): a lucid synthesis of mental accounting by its originator.
- J. St. B. T. Evans and Stanovich (2013): the most careful defence of dual-process theory, with replies to the main criticisms.
- Schwarz (2012): an authoritative overview of feelings-as-information theory by its co-founder.
- R. H. Thaler and Sunstein (2008): the book that turned choice architecture into a policy movement. Pair it with Maier et al. (2022) for a sober assessment of effect sizes.

Chapter 24

Prospect Theory and Reference-Dependent Preferences

On the professional golf tour, every hole comes with a number printed on the scorecard: par. Par is not a payoff. Golfers are paid for the total number of strokes over a tournament, so a stroke saved on a birdie putt is worth exactly as much as a stroke saved on a par putt. Yet when Pope and Schweitzer (2011) analysed more than 2.5 million putts measured by laser, they found that the best players in the world, Tiger Woods included, sink putts for par more often than putts for birdie from the same distance. A missed par putt is a *loss* relative to the number on the card. A missed birdie putt is merely a forgone *gain*. The same asymmetry appears far from the golf course. Boston condominium owners who would have to sell below their purchase price set higher asking prices and wait longer to sell than owners sitting on a gain (Genesove and Mayer 2001). Marathon finishing times bunch just before round numbers such as four hours (E. J. Allen et al. 2017). Upset losses by a local football team, but not expected losses, are followed by a rise in domestic violence (Card and Dahl 2011). In each case, behaviour depends less on final outcomes than on how outcomes compare with a **reference point**.

Prospect theory is the framework that turned observations like these into a formal, testable model of choice. Its 1979 statement by Daniel Kahneman and Amos Tversky (Kahneman and Tversky 1979) is among the most cited papers in the social sciences, and its central ideas, reference dependence, loss aversion, diminishing sensitivity, and probability weighting, now shape research in economics, finance, marketing, management, psychology, political science, law, and medicine (N. C. Barberis 2013, DellaVigna 2009). This chapter treats the theory in depth. We begin with the benchmark it was built to replace, **expected utility theory**, and the paradoxes that unsettled it. We then develop the original **prospect theory** model, its **cumulative** successor, and the most debated of its components, **loss aversion**. We follow the theory's modern extension to **expectation-based reference points** and three of its most important applications, **status quo bias**, **myopic loss aversion**, and the **disposition effect**, before surveying applications across business and policy and the evidence from large-scale replications. Chapter 23 (Section 23.3) gives a short overview for readers who need only the essentials (Table 24.1).

Table 24.1. Chapter at a glance: prospect theory and its family of models.

Theory	Core question	Level of analysis	Key source
Expected utility theory	How should (and do) people choose among risky prospects?	Individual	Bernoulli (1738), von Neumann & Morgenstern (1944)
Prospect theory	How do people actually evaluate risky prospects?	Individual / market	Kahneman & Tversky (1979)
Cumulative prospect theory	How can probability weighting be made consistent with dominance?	Individual	Tversky & Kahneman (1992)
Loss aversion	Why do losses weigh more than equal-sized gains?	Individual / dyad	Tversky & Kahneman (1991)
Reference-dependent preferences	What is the reference point, and where does it come from?	Individual / market	Kőszegi & Rabin (2006)
Status quo bias	Why do people stick with the current option?	Individual / organization	Samuelson & Zeckhauser (1988)
Myopic loss aversion	Why do investors demand so much to hold stocks?	Individual / market	Benartzi & Thaler (1995)
Disposition effect	Why do investors sell winners and hold losers?	Individual / market	Shefrin & Statman (1985)

Figure 24.1 maps the intellectual lineage that organizes the chapter. Each arrow marks a response to an anomaly or a gap in the model that preceded it.

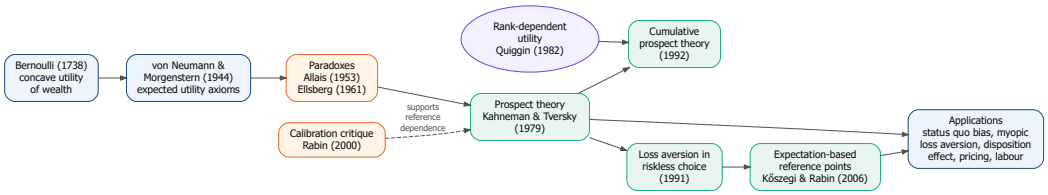


Figure 24.1. The intellectual lineage of prospect theory. Expected utility theory provided the normative benchmark, paradoxes and field anomalies motivated prospect theory, and later work generalized its probability weighting (cumulative prospect theory) and its reference point (expectation-based models), and carried it into finance, labour, marketing, and policy.

24.1 Expected Utility Theory

THEORY CARD

Core claim. A decision maker whose preferences over risky prospects satisfy a small set of axioms (completeness, transitivity, continuity, and independence) behaves *as if* she maximizes the probability-weighted average of the utility of final outcomes. Risk aversion is explained entirely by the curvature of the utility function over wealth.

Origins. Mathematics and economics. Bernoulli (1954) (the 1954 English translation of a 1738 paper) proposed concave utility of wealth. Neumann and Morgenstern (1944) axiomatized expected utility for objective probabilities, and Savage (1954) extended it to subjective probabilities.

Level of analysis. Individual choice under risk, used as a building block in models of markets,

contracts, insurance, and finance.
Key constructs. Lottery (prospect), utility of final wealth, expected utility, risk aversion, certainty equivalent, independence axiom.
Boundary conditions. Normatively compelling for many decision makers. Descriptively, it fails systematically when outcomes are certain versus nearly certain, when probabilities are small, when stakes are modest, and when probabilities are ambiguous.
Related theories. Bounded rationality (Section 23.1), Heuristics and biases (Section 23.2), Prospect theory, Portfolio theory (Section 16.1).

24.1.1 Origins and intellectual history

The modern theory of risky choice begins with a puzzle. In the **St. Petersburg game**, a fair coin is tossed until it lands heads, and the player receives a prize that doubles with every toss. The expected monetary value of the game is infinite, yet few people would pay more than a modest sum to play it. Daniel Bernoulli's resolution, published in 1738 and translated into English in 1954, was that people do not value money linearly: the utility of an additional ducat declines as wealth grows (Bernoulli 1954). With a concave (in his example logarithmic) utility of wealth, the expected *utility* of the St. Petersburg game is finite and small. Two ideas from this paper became the core of economic thinking about risk for two centuries: outcomes are evaluated as **final states of wealth**, and **risk aversion** reflects **diminishing marginal utility**.

Neumann and Morgenstern (1944) gave the idea an axiomatic foundation. In the axiomatic treatment, added as an appendix to the second edition of *Theory of Games and Economic Behavior*, they showed that if preferences over lotteries satisfy a handful of axioms, there exists a utility function, unique up to a positive linear transformation, such that one lottery is preferred to another if and only if it has higher expected utility. Savage (1954) then derived subjective probabilities and utilities jointly from preferences over acts, so that expected utility could be applied even when probabilities are not given objectively.

DEFINITION

Definition 24.1. Expected utility. A prospect (lottery) L yields outcome x_i with probability p_i, for $i = 1, \dots, n$. For a decision maker with initial wealth W and utility function $u(\cdot)$,

$$EU(L) = \sum_{i=1}^{n} p_i \, u(W + x_i).$$

The decision maker is **risk averse** if she prefers the expected value of a lottery for sure to the lottery itself, which holds for every lottery if and only if u is concave. The **certainty equivalent** CE solves $u(W + CE) = EU(L)$, and the **risk premium** is $E[x] - CE$.

Table 24.2 summarises the axioms. The **independence axiom** carries most of the theory's empirical content: it implies that preferences are *linear in probabilities*, so that mixing two lotteries with a common third lottery cannot reverse the preference between them.

Table 24.2. The von Neumann-Morgenstern axioms of expected utility theory.

Axiom	Statement	Intuition	Best-known challenge
Completeness	For any two lotteries A and B, either A is weakly preferred to B or B to A	The decision maker can always rank options	Indecision and incomplete preferences when options are hard to compare
Transitivity	If A is weakly preferred to B and B to C, then A is weakly preferred to C	No preference cycles, so no money pump	Preference reversals between choice and pricing tasks
Continuity	If A is preferred to B and B to C, some mixture of A and C is indifferent to B	No outcome is infinitely good or bad	Lexicographic concerns (for example, any chance of death)
Independence	A is preferred to B if and only if a mixture of A with C is preferred to the same mixture of B with C	Common components should not affect choice	Allais paradox: common consequence and common ratio effects

24.1.2 The paradoxes

Allais. Maurice Allais constructed choice problems in which many people, including some committed proponents of expected utility, answered in ways that violated the independence axiom (Allais 1953). Table 24.3 presents the version that textbooks usually give, with prizes in dollars.

Table 24.3. The Allais paradox in its common textbook form. Choosing A in Problem 1 and D in Problem 2 violates the independence axiom.

Problem	Option	Outcomes	Typical choice
1	A	\$1 million for sure	Most people choose A
1	B	89% chance of \$1 million, 10% chance of \$5 million, 1% chance of nothing	
2	C	11% chance of \$1 million, 89% chance of nothing	
2	D	10% chance of \$5 million, 90% chance of nothing	Most people choose D

Under expected utility, choosing A over B means $0.11\,u(1\text{M}) > 0.10\,u(5\text{M}) + 0.01\,u(0)$. Choosing D over C means the reverse inequality. Both problems share a common consequence (an 89% chance of \$1 million in Problem 1, an 89% chance of nothing in Problem 2), and independence requires that this common component be ignored. People do not ignore it. Certainty is special: removing the sure thing in Problem 1 removes a reason to choose A that has no counterpart in Problem 2. Kahneman and Tversky would later call this the **certainty effect**.

Ellsberg. Ellsberg (1961) identified a different kind of violation. Imagine an urn with 90 balls: 30 are red, and the remaining 60 are black or yellow in an unknown proportion. Most people prefer to bet on red rather than on black, but prefer to bet on "black or yellow" rather than on "red or yellow". No assignment of subjective probabilities to black and yellow can rationalize both choices. People are averse to **ambiguity**, uncertainty about the probabilities themselves, and this violates Savage's sure-thing principle. Ambiguity aversion is a distinct phenomenon from the probability distortions discussed below, but it showed early that the subjective expected utility model was not a complete description of how people treat uncertainty.

Rabin's calibration theorem. The most damaging critique arrived much later and from inside economics. Rabin (2000) proved that expected utility theory cannot explain the moderate risk aversion people display over modest stakes without implying absurd risk aversion over large stakes. His lead-

ing example: a person who, at every level of wealth, turns down a 50-50 gamble to lose $100 or gain $110 must also turn down a 50-50 gamble to lose $1,000 or gain *any* amount of money. The reason is that an expected-utility maximizer who rejects small favourable gambles must have marginal utility that falls very quickly over a narrow range of wealth, and compounding that rate of decline over larger ranges makes large gains nearly worthless. Rabin's conclusion was that modest-scale risk aversion is better explained by **loss aversion** relative to a reference point than by the curvature of the utility of wealth (Rabin 2000, Rabin and Thaler 2001).

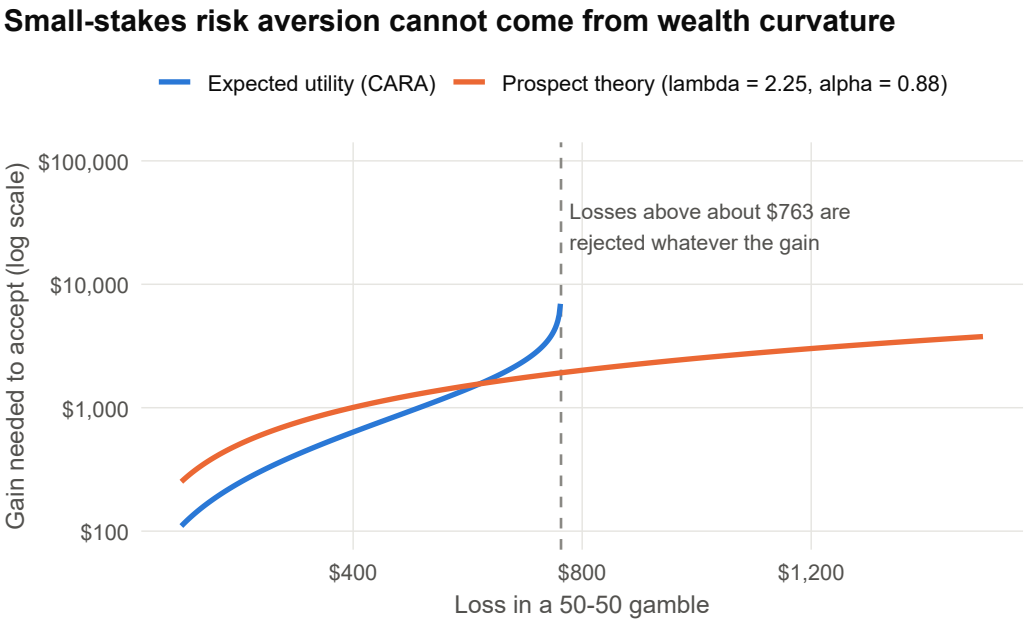


Figure 24.2. Calculated illustration of Rabin's calibration argument. An expected-utility maximizer with constant absolute risk aversion who rejects a 50-50 lose 100 or gain 110 gamble at every wealth level will reject any 50-50 gamble whose loss exceeds a fixed threshold, however large the gain. A prospect-theory decision maker with the Tversky and Kahneman (1992) value function requires a gain that grows only in proportion to the loss.

Figure 24.2 makes the logic concrete for one special case. Constant absolute risk aversion is the utility function for which rejecting the $100/$110 gamble at *every* wealth level is exactly consistent. A decision maker with that utility rejects any 50-50 gamble that risks more than about $763, however large the possible gain. Rabin's theorem shows that a similar, slightly weaker conclusion holds for *any* concave utility function. A loss-averse decision maker whose value function is defined over gains and losses has no such problem: the gain she requires grows in proportion to the loss.

DEBATE AND EVIDENCE

Is expected utility refuted, or misapplied? Defenders of expected utility make three replies. First, the theory is *normative*: it describes how a consistent decision maker *should* choose, and many people who violate the Allais pattern revise their choices when the independence logic is explained. Second, some anomalies can be absorbed by richer expected-utility models, for example by allowing utility over consumption streams, background risk, or narrow bracketing of decisions. Third, a family of **non-expected utility** theories relaxes independence

while keeping much of the structure, including regret theory (Loomes and Sugden 1982), disappointment aversion (Gul 1991), and rank-dependent utility (Quiggin 1982), reviewed by Starmer (2000). Rabin's calibration result sharpened the debate because it does not depend on laboratory paradoxes: it shows that *any* expected-utility account of everyday small-stakes risk aversion has implausible implications. The weight of opinion in behavioural economics is that expected utility remains the right normative benchmark but that descriptive models need a reference point.

COMMON PITFALL

Treating expected utility as a straw man. Expected utility is not "the assumption that people are rational calculators". It is a precise, testable model with clear axioms, and its violations are informative precisely because they are specific. When writing about prospect theory, state which axiom or which implication of expected utility a finding contradicts (independence, final-wealth evaluation, or linear probability weighting). Findings that contradict none of them, for example risk aversion over large stakes, are not evidence against expected utility.

24.2 Prospect Theory: The 1979 Model

THEORY CARD

Core claim. People evaluate risky prospects in two phases. In an *editing* phase they code outcomes as gains or losses relative to a reference point and simplify the prospects. In an *evaluation* phase they combine a value function that is concave for gains, convex for losses, and steeper for losses with decision weights that overweight small probabilities and underweight moderate and large ones.
Origins. Psychology and economics. Kahneman and Tversky (1979), building on a decade of collaboration on judgment under uncertainty (Tversky and Kahneman 1974). Framing extensions in Tversky and Kahneman (1981), Kahneman and Tversky (1984), and Tversky and Kahneman (1986).
Level of analysis. Individual choice. Aggregate implications for asset prices, insurance, labour supply, pricing, and policy.
Key constructs. Reference point, gains and losses, value function, diminishing sensitivity, loss aversion, decision weights, certainty effect, reflection effect, isolation effect, editing operations.
Boundary conditions. Formulated for choices among described prospects with few outcomes and known probabilities. The reference point must be identifiable. The editing phase is informal, so predictions depend on how a problem is presented.
Related theories. Expected utility theory, Cumulative prospect theory, Framing theory (Section 26.6), Heuristics and biases (Section 23.2), Mental accounting (Section 23.5).

24.2.1 Origins and intellectual history

By the mid-1970s Kahneman and Tversky had spent several years documenting the heuristics that people use to judge probabilities (Tversky and Kahneman 1974). Prospect theory turned from judgment to choice. Its method was disarmingly simple. The authors posed hypothetical choices

between pairs of gambles to student and faculty respondents and reported the share choosing each option. The problems were designed so that expected utility made sharp predictions, typically that the preference in one problem must match the preference in a closely related problem. The data repeatedly showed that it did not. The paper, published in *Econometrica* in 1979, combined these demonstrations with a formal model that could explain them, which is a large part of why it changed economics as well as psychology (N. C. Barberis 2013).

24.2.2 The empirical demonstrations

Table 24.4 reproduces several of the original choice problems. Payoffs were in Israeli pounds, and the percentages are the shares of respondents choosing each option as reported by Kahneman and Tversky (1979). Read each pair of rows together.

Table 24.4. Selected choice problems from Kahneman and Tversky (1979). Amounts are in Israeli pounds. Percentages are the shares of respondents choosing each option, as reported in the original article.

Problem	Option A	Option B	Choices (A / B)	Pattern illustrated
1	2,500 with p = .33, 2,400 with p = .66, 0 with p = .01	2,400 for sure	18% / 82%	Certainty effect (common consequence)
2	2,500 with p = .33, 0 with p = .67	2,400 with p = .34, 0 with p = .66	83% / 17%	Certainty effect (common consequence)
3	4,000 with p = .80	3,000 for sure	20% / 80%	Certainty effect (common ratio)
4	4,000 with p = .20	3,000 with p = .25	65% / 35%	Certainty effect (common ratio)
7	6,000 with p = .45	3,000 with p = .90	14% / 86%	Preference for the more probable gain
8	6,000 with p = .001	3,000 with p = .002	73% / 27%	Possibility effect: small probabilities
3’	Lose 4,000 with p = .80	Lose 3,000 for sure	92% / 8%	Reflection effect: risk seeking for losses
11	Given 1,000, then win 1,000 with p = .50	Given 1,000, then win 500 for sure	16% / 84%	Isolation effect: same final states as 12
12	Given 2,000, then lose 1,000 with p = .50	Given 2,000, then lose 500 for sure	69% / 31%	Isolation effect: same final states as 11

Four patterns stand out.

1. **The certainty effect.** Problems 1 and 2 reproduce the Allais pattern with modest stakes. Problems 3 and 4 show a related violation: scaling both winning probabilities by the same factor (from .80 and 1.00 to .20 and .25) reverses the majority preference. Under expected utility, preferences should be unaffected by such a common ratio. Outcomes that are certain receive extra weight relative to outcomes that are merely probable.
2. **The possibility effect.** Problems 7 and 8 have the same ratio of probabilities and prizes. When winning is probable, most people prefer the more likely, smaller prize. When winning is merely possible, most prefer the larger prize. Very small probabilities are treated differently from moderate ones.
3. **The reflection effect.** When the signs of the outcomes in Problem 3 are reversed (Problem 3’), the preference reverses as well. A large majority prefer a sure gain of 3,000 to an 80% chance of 4,000, but prefer an 80% chance of losing 4,000 to a sure loss of 3,000. Risk aversion

in the domain of gains is accompanied by **risk seeking in the domain of losses**.

4. **The isolation effect.** Problems 11 and 12 offer identical distributions of final wealth: in both, the choice is between 1,500 for sure and an even chance of 1,000 or 2,000. Yet respondents chose the sure thing when the problem was framed as gains and the gamble when it was framed as losses. People evaluate *changes*, not final states, and they evaluate the components of a problem in isolation rather than integrating them with what they already have.

The last result is the most radical. It is a direct contradiction of the assumption, shared by Bernoulli and by von Neumann and Morgenstern, that the carriers of utility are final states of wealth. It also made **framing** a central topic: if the same decision can be described as a choice between gains or between losses, and the description changes the choice, then description invariance fails (Tversky and Kahneman 1981, 1986).

24.2.3 Core logic: editing and evaluation

The model has two phases (Figure 24.3). In the **editing** phase, the decision maker organizes and reformulates the options using the operations listed in Table 24.5. In the **evaluation** phase, each edited prospect is assigned a value and the prospect with the highest value is chosen. For a simple prospect that pays x with probability p and y with probability q, with at most two non-zero outcomes,

$$V = \pi(p)\, v(x) + \pi(q)\, v(y),$$

where $v(\cdot)$ is the **value function** and $\pi(\cdot)$ is the **decision weight** attached to each probability. The original paper treated strictly positive and strictly negative prospects separately, segregating the riskless component of the prospect before applying decision weights.

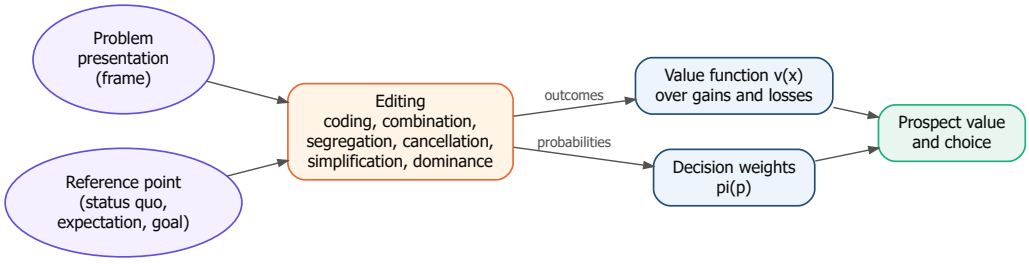


Figure 24.3. The two phases of the 1979 model. Editing codes outcomes relative to a reference point and simplifies the prospects. Evaluation applies the value function to outcomes and the weighting function to probabilities. How a problem is presented (the frame) enters through editing.

Table 24.5. Editing operations in the original prospect theory (Kahneman & Tversky, 1979).

Operation	What it does	Example
Coding	Outcomes are coded as gains or losses relative to a reference point	A salary of 90,000 is a loss for someone who expected 100,000
Combination	Probabilities attached to identical outcomes are added	Two chances of 200, each with p = .25, become one chance of 200 with p = .50
Segregation	A riskless component is separated from the risky component	300 with p = .80 or else 200 becomes a sure 200 plus a .80 chance of 100 more

Table 24.5. Editing operations in the original prospect theory (Kahneman & Tversky, 1979). *(continued)*

Operation	What it does	Example
Cancellation	Components shared by the options are discarded	The common first stage of a two-stage gamble is ignored
Simplification	Probabilities or outcomes are rounded, and extremely unlikely outcomes are dropped	A .49 chance of 101 is treated as an even chance of 100
Detection of dominance	Dominated alternatives are identified and rejected without evaluation	A gamble that is worse in every state is eliminated

24.2.4 The value function

The value function captures three psychological principles.

- **Reference dependence.** The carriers of value are gains and losses relative to a reference point, not final states of wealth. The reference point is usually the status quo but can be an expectation, an aspiration, or a social norm.
- **Diminishing sensitivity.** The difference between gains of 100 and 200 feels larger than the difference between gains of 1,100 and 1,200, and the same holds for losses. The function is therefore concave for gains and convex for losses, which produces risk aversion for gains and risk seeking for losses when probabilities are moderate.
- **Loss aversion.** The function is steeper for losses than for gains. Most people reject a 50-50 gamble to win or lose the same amount, and the gain needed to make such a gamble acceptable is typically about twice the loss.

A convenient parametric form, later estimated by Tversky and Kahneman (1992), is the power function

$$v(x) = \begin{cases} x^{\alpha} & \text{if } x \geq 0 \\ -\lambda(-x)^{\beta} & \text{if } x < 0, \end{cases}$$

with $0 < \alpha, \beta \leq 1$ governing curvature and $\lambda \geq 1$ the coefficient of loss aversion. Figure 24.4 shows how the shape responds to each parameter. Lowering α and β strengthens diminishing sensitivity, so the function flattens away from the reference point. Raising λ rotates the loss limb downward and strengthens the kink at zero. The kink matters as much as the curvature: it generates **first-order risk aversion**, aversion to small fair gambles that does not vanish as stakes shrink, which expected utility with a smooth utility function cannot produce.

24.2.5 Decision weights

In the 1979 model, decision weights $\pi(p)$ are not probabilities. They measure the impact of an event on the desirability of a prospect. The paper derived several properties from the choice data: π is increasing, with $\pi(0) = 0$ and $\pi(1) = 1$. Very small probabilities are **overweighted**, $\pi(p) > p$, which explains the possibility effect and the joint appeal of lotteries and insurance. Weights display **subcertainty**, $\pi(p) + \pi(1 - p) < 1$, which explains the certainty effect. Weights are **subproportional**: for a fixed ratio of probabilities, the ratio of the corresponding weights is closer to one when probabilities are low than when they are high, which explains the common ratio pattern in Problems 3 and 4 and in Problems 7 and 8. Finally, the function is poorly behaved near the endpoints, where small probabilities may be either ignored in editing or greatly overweighted.

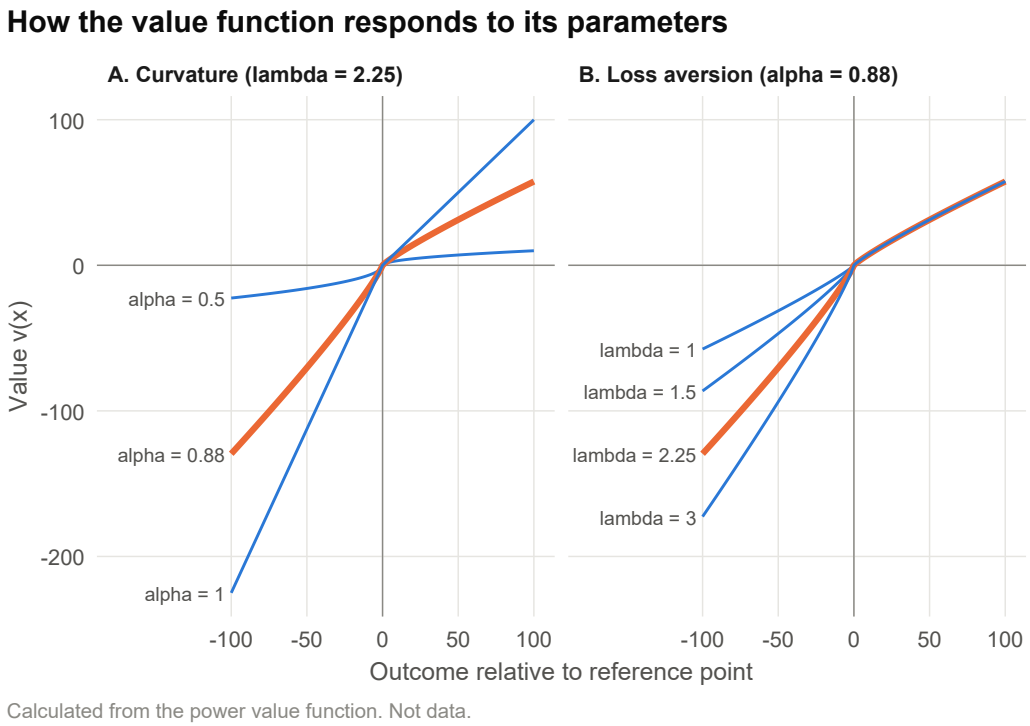


Figure 24.4. Parameter sensitivity of the prospect theory value function. Left: curvature (alpha = beta) varied with lambda fixed at 2.25. Right: loss aversion (lambda) varied with curvature fixed at 0.88. The highlighted curves use the Tversky and Kahneman (1992) estimates.

24.2.6 Propositions

The 1979 model can be summarized in six propositions.

- **P1 (Reference dependence).** Choices depend on outcomes coded as gains and losses relative to a reference point, not on final wealth.
- **P2 (Loss aversion).** Losses have a greater impact on choice than gains of equal size.
- **P3 (Diminishing sensitivity).** The marginal impact of a gain or loss declines with its distance from the reference point, producing the reflection effect.
- **P4 (Probability weighting).** Small probabilities are overweighted and moderate to large probabilities underweighted, with a discontinuity at certainty.
- **P5 (Isolation and framing).** Components that are shared across options are cancelled, and descriptions that shift the reference point change choices even when final outcomes are identical.
- **P6 (Path dependence).** Prior gains and losses that have not been fully integrated into the reference point change subsequent risk taking.

P6 is not a formal part of the 1979 model, but it follows from the logic of reference dependence and has strong empirical support. R. H. Thaler and Johnson (1990) documented a **house money effect**, greater risk taking after a prior gain, and a **break-even effect**, greater willingness to accept gambles that offer a chance to recover a prior loss. In the high-stakes television game show *Deal or No Deal*, contestants became less risk averse both after unfavourable outcomes had shattered their expectations and after favourable outcomes had exceeded them, a pattern that is hard to reconcile

with expected utility of final wealth (Post et al. 2008).

BUSINESS CASE

Timid choices in organizations. Kahneman and Lovallo (1993) argued that managers evaluate risky projects one at a time, in isolation, and therefore apply loss aversion to each project separately. A firm that would happily accept a portfolio of twenty independent projects, each with a positive expected value and a modest chance of loss, may reject every one of them when they are presented sequentially to different managers. The prescription follows from the isolation effect: aggregate decisions into portfolios, evaluate managers on the portfolio rather than on single outcomes, and make the reference point explicit (for example, the hurdle rate rather than zero). **Illustrative scenario:** a pharmaceutical R&D committee that reviews compounds one at a time, with a separate "kill or continue" vote each quarter, will terminate more promising but risky compounds than a committee that reviews the pipeline as a whole against an explicit expected-value threshold.

TRY IT YOURSELF

Answer Problems 3 and 3' in Table 24.4 before reading the original percentages. Then ask three friends. Do your answers show the reflection effect? Now rewrite Problem 12 so that it is framed as a gain from a lower starting endowment, and check whether your own preference changes.

24.3 Cumulative Prospect Theory

THEORY CARD

Core claim. Decision weights apply to *cumulative* probabilities of outcomes ranked by their attractiveness, separately for gains and losses. This rank-dependent formulation preserves the insights of prospect theory while satisfying first-order stochastic dominance and extending the model to prospects with many outcomes and to uncertainty.
Origins. Decision theory and economics. Tversky and Kahneman (1992), building on rank-dependent utility (Quiggin 1982). Axiomatic weighting functions in Prelec (1998), curvature and elevation in Gonzalez and Wu (1999), and a comprehensive treatment in Wakker (2010).
Level of analysis. Individual choice under risk and uncertainty.
Key constructs. Rank-dependent decision weights, cumulative weighting function, inverse-S shape, curvature (likelihood insensitivity), elevation (optimism), fourfold pattern of risk attitudes, sign dependence.
Boundary conditions. Best supported for described lotteries with known probabilities. When people learn probabilities from experience, rare events are often *underweighted* (Hertwig et al. 2004). Parameter estimates vary across people, tasks, and elicitation methods.
Related theories. Prospect theory (1979), Expected utility theory, Loss aversion.

24.3.1 Why the original model needed revision

The 1979 model had two technical weaknesses. First, applying a nonlinear weighting function to the probabilities of individual outcomes can violate **first-order stochastic dominance**: a decision

maker might prefer a prospect that is worse in every state of the world. Kahneman and Tversky handled this with an editing rule that eliminates dominated options, but the fix was ad hoc. Second, the model was formulated for prospects with at most two non-zero outcomes, which made it hard to apply to stock returns, insurance contracts, or any continuous distribution.

The solution came from **rank-dependent utility** (Quiggin 1982). Instead of transforming the probability of each outcome, rank the outcomes from best to worst and transform the *cumulative* probability of doing at least as well as a given outcome. The decision weight of an outcome is then the marginal contribution of that outcome to the transformed cumulative probability. Tversky and Kahneman (1992) combined this idea with sign dependence: gains and losses are ranked and weighted separately, each with its own weighting function.

24.3.2 The model

Consider a prospect with outcomes ordered $x_{-m} \leq \ldots \leq x_{-1} < 0 \leq x_1 \leq \ldots \leq x_n$ and probabilities $p_{-m}, \ldots, p_n$. Cumulative prospect theory evaluates it as

$$V = \sum_{i=1}^{n} \pi_i^{+} \, v(x_i) + \sum_{j=1}^{m} \pi_{-j}^{-} \, v(x_{-j}),$$

with decision weights for gains defined from the best outcome downward,

$$\pi_i^{+} = w^{+}(p_i + \cdots + p_n) - w^{+}(p_{i+1} + \cdots + p_n)\,,$$

and for losses from the worst outcome upward,

$$\pi_{-j}^{-} = w^{-}\big(p_{-j} + \cdots + p_{-m}\big) - w^{-}\big(p_{-j+1} + \cdots + p_{-m}\big)\,.$$

Because the weights are differences of a monotone function of cumulative probabilities, a prospect that shifts probability toward better outcomes can never be valued less, so dominance is preserved automatically. The inverse-S shape of w means that the *extreme* outcomes, the best gain and the worst loss, receive the most overweighting, while intermediate outcomes are underweighted.

Tversky and Kahneman (1992) proposed the one-parameter weighting function

$$w(p) = \frac{p^{\gamma}}{\left(p^{\gamma} + (1-p)^{\gamma}\right)^{1/\gamma}},$$

and estimated median parameters from certainty equivalents for a set of gain, loss, and mixed prospects: $\alpha = \beta = 0.88$ for the value function, $\lambda = 2.25$ for loss aversion, $\gamma = 0.61$ for the weighting of gains, and $\delta = 0.69$ for the weighting of losses. These values remain the default in applied work, although later studies find substantial heterogeneity across individuals and methods (Abdellaoui, Bleichrodt, and Paraschiv 2007, Wakker 2010).

Figure 24.5 plots the Tversky-Kahneman function for several values of γ alongside the one-parameter function axiomatized by Prelec (1998), $w(p) = \exp\{-(-\ln p)^{a}\}$, whose inverse-S shape always crosses the diagonal at $p = 1/e \approx 0.37$. Smaller parameter values mean greater

curvature: the function is flatter in the middle, so the decision maker is less sensitive to changes in intermediate probabilities and more sensitive to the move from impossibility to possibility and from probability to certainty. Gonzalez and Wu (1999) showed that two psychologically distinct properties are at work. **Curvature** reflects *discriminability*, how sensitive a person is to changes in probability. **Elevation** reflects *attractiveness*, how much a person likes betting on the chance domain overall. A two-parameter linear-in-log-odds form separates them.

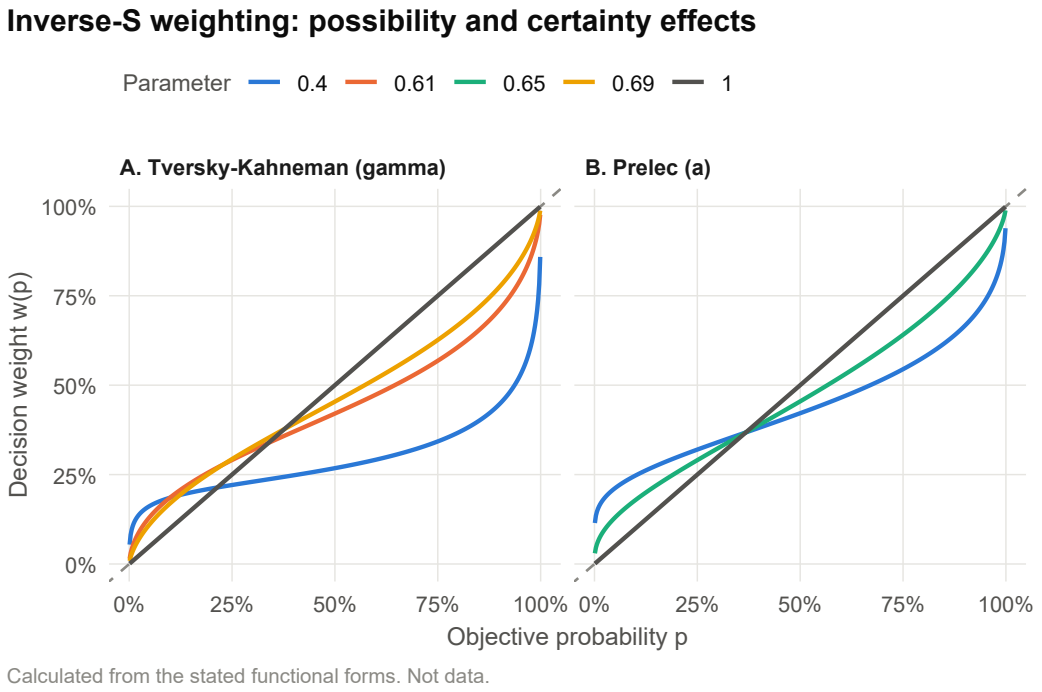


Figure 24.5. Probability weighting functions. Left: the Tversky and Kahneman (1992) form for several values of gamma (0.61 is their median estimate for gains, 0.69 for losses, and 1 is linear weighting). Right: the one-parameter Prelec (1998) form, which crosses the diagonal at 1/e for every parameter value.

Table 24.6 collects the functional forms used most often in empirical work. The choice matters less than is often assumed for qualitative predictions, but it matters a great deal for parameter estimates, which are comparable only within a functional form.

Table 24.6. Common functional forms for the value and probability weighting functions.

Component	Functional form	Parameters	Notes
Value function	Power: x^alpha for gains, -lambda(-x)^beta for losses	alpha, beta (curvature), lambda (loss aversion)	Most common. Loss aversion is only well defined at a unit scale when alpha differs from beta
Value function	Exponential: (1 - exp(-a x)) / a for gains, mirrored for losses	a, b, lambda	Constant absolute sensitivity. Avoids the scaling problem of the power form
Value function	Piecewise linear: x for gains, lambda x for losses	lambda	Isolates loss aversion. Used in many riskless-choice and field studies
Weighting function	Tversky-Kahneman (1992)	gamma (gains), delta (losses)	Inverse-S. Not monotone for very small parameter values

Table 24.6. Common functional forms for the value and probability weighting functions. *(continued)*

Component	Functional form	Parameters	Notes
Weighting function	Prelec (1998) one-parameter: exp(-(-ln p)^a)	a	Axiomatically derived. Fixed point at 1/e
Weighting function	Prelec two-parameter: exp(-b(-ln p)^a)	a (curvature), b (elevation)	Separates curvature from elevation
Weighting function	Linear in log odds: d p^g / (d p^g + (1 - p)^g)	g (curvature), d (elevation)	Gonzalez and Wu (1999). Popular in individual-level estimation

24.3.3 How rank-dependent weights work

Figure 24.6 illustrates the decision weights for a four-outcome prospect: win 100 with probability .05, win 40 with probability .30, win 10 with probability .40, and lose 30 with probability .25. The best outcome and the only loss are overweighted relative to their probabilities. The intermediate gains are underweighted. The weights on gains and losses need not sum to one, which is one way prospect theory departs from any model that treats decision weights as beliefs.

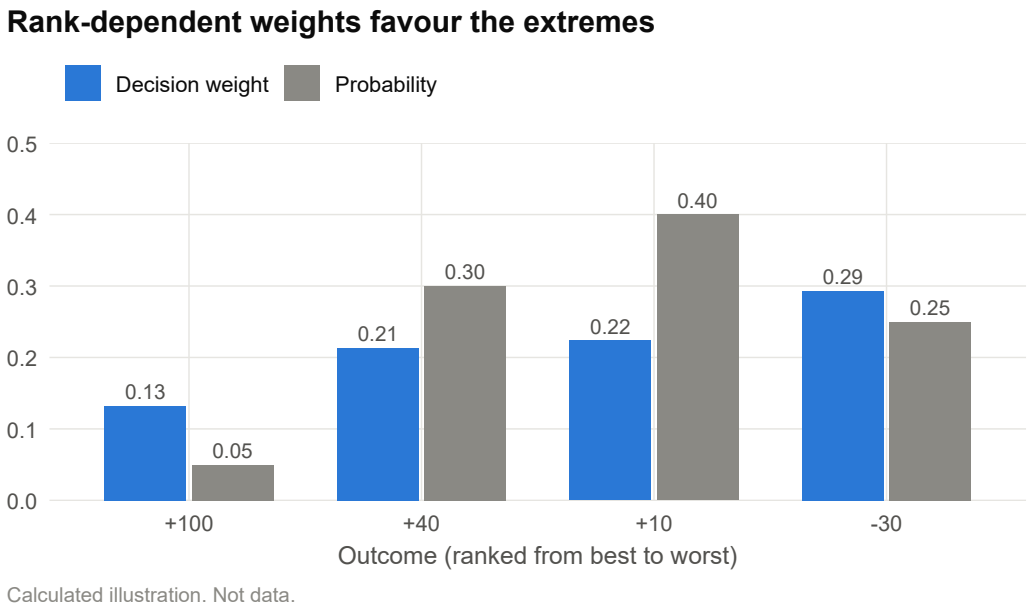


Figure 24.6. Cumulative prospect theory decision weights for an illustrative four-outcome prospect, computed with the Tversky and Kahneman (1992) weighting functions (gamma = 0.61 for gains, delta = 0.69 for losses). Extreme outcomes are overweighted and intermediate outcomes underweighted.

24.3.4 A reusable implementation

The following code implements the cumulative prospect theory value and certainty equivalent of any discrete prospect. It is shown so that readers can reuse it for teaching, for simulation, or as a starting point for estimation. The default parameters are the median estimates of Tversky and Kahneman (1992).

```
# Value function: power form with loss aversion
pt_value <- function(x, alpha = 0.88, beta = 0.88, lambda = 2.25) {
  ifelse(x >= 0, abs(x)^alpha, -lambda * abs(x)^beta)
}

# Tversky-Kahneman (1992) probability weighting function
pt_weight <- function(p, g) p^g / (p^g + (1 - p)^g)^(1 / g)

# Cumulative prospect theory value of a discrete prospect (outcomes x, probabilities p)
cpt <- function(x, p, alpha = 0.88, beta = 0.88, lambda = 2.25,
                gamma = 0.61, delta = 0.69) {
  stopifnot(length(x) == length(p), abs(sum(p) - 1) < 1e-8)
  w <- numeric(length(x))
  gains <- which(x > 0)
  losses <- which(x < 0)
  if (length(gains)) {                          # rank gains from best to worst
    o <- gains[order(x[gains], decreasing = TRUE)]
    cp <- cumsum(p[o])
    w[o] <- pt_weight(cp, gamma) - pt_weight(c(0, head(cp, -1)), gamma)
  }
  if (length(losses)) {                         # rank losses from worst to best
    o <- losses[order(x[losses])]
    cp <- cumsum(p[o])
    w[o] <- pt_weight(cp, delta) - pt_weight(c(0, head(cp, -1)), delta)
  }
  V <- sum(w * pt_value(x, alpha, beta, lambda))
  ce <- if (V >= 0) V^(1 / alpha) else -(-V / lambda)^(1 / beta)
  list(value = V, certainty_equivalent = ce, weights = w)
}

# Example: the four-outcome prospect used in the figure above
ex <- cpt(x = c(100, 40, 10, -30), p = c(0.05, 0.30, 0.40, 0.25))
round(ex$certainty_equivalent, 2)
```

```
#> [1] 1.67
```

The certainty equivalent of the example prospect is 1.7, compared with an expected value of 13.5. The gap reflects loss aversion: the single loss of 30 is multiplied by $\lambda = 2.25$ and, as the worst outcome, is also overweighted, so it offsets most of the value of the three gains. The same function reproduces the **fourfold pattern of risk attitudes**, the theory's signature prediction for simple prospects (Table 24.7).

```
fourfold <- expand.grid(p = c(0.05, 0.95), x = c(100, -100))
fourfold$ev <- fourfold$p * fourfold$x
fourfold$ce <- mapply(function(p, x) cpt(c(x, 0), c(p, 1 - p))$certainty_equivalent,
                      fourfold$p, fourfold$x)
fourfold$attitude <- ifelse(fourfold$ce > fourfold$ev, "Risk seeking", "Risk averse")
fourfold |>
```

```
  transmute(Prospect = sprintf("%s %d with p = %.2f", ifelse(x > 0, "Win", "Lose"), abs(x), p)
            `Expected value` = round(ev, 1),
            `Certainty equivalent` = round(ce, 1),
            Prediction = attitude) |>
  book_table(caption = "The fourfold pattern computed with cumulative prospect theory and the
```

Table 24.7. The fourfold pattern computed with cumulative prospect theory and the Tversky and Kahneman (1992) median parameters.

Prospect	Expected value	Certainty equivalent	Prediction
Win 100 with p = 0.05	5	10.0	Risk seeking
Win 100 with p = 0.95	95	76.9	Risk averse
Lose 100 with p = 0.05	-5	-8.3	Risk averse
Lose 100 with p = 0.95	-95	-83.1	Risk seeking

Figure 24.7 extends the calculation across the full range of probabilities. Each cell shows the risk premium implied by cumulative prospect theory, expressed as a percentage of the expected value and signed so that positive numbers indicate risk seeking. The four quadrants of the pattern are visible at a glance: risk seeking for low-probability gains (lotteries), risk aversion for high-probability gains (settling a strong lawsuit), risk aversion for low-probability losses (insurance), and risk seeking for high-probability losses (gambling to avoid a sure loss). The same person may therefore buy both insurance and lottery tickets, a long-standing puzzle for expected utility with a single concave utility function.

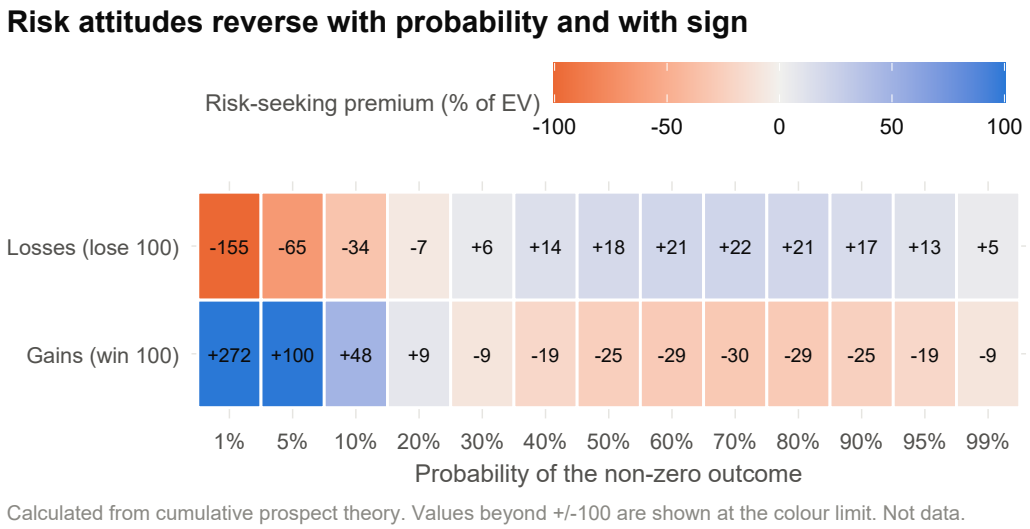


Figure 24.7. The fourfold pattern of risk attitudes. Cells show the risk-seeking premium implied by cumulative prospect theory for a prospect that pays (or costs) 100 with probability p and nothing otherwise, as a percentage of its expected value. Positive values (blue) indicate risk seeking, negative values (orange) risk aversion. Computed from the Tversky and Kahneman (1992) median parameters.

Table 24.8 presents the original and cumulative prospect theory compared.

Table 24.8. Original and cumulative prospect theory compared.

Feature	Prospect theory (1979)	Cumulative prospect theory (1992)
Weights applied to	Probabilities of individual outcomes	Cumulative probabilities of ranked outcomes
Stochastic dominance	Can be violated, handled by an editing rule	Always satisfied
Number of outcomes	At most two non-zero outcomes	Any number, including continuous distributions
Gains and losses	Treated separately for strictly positive or negative prospects	Separate weighting functions, w+ and w-
Uncertainty (unknown probabilities)	Not covered	Covered through capacities (Choquet integrals)
Editing phase	Central and informal	Largely dropped from the formal model
Parameter estimates	None reported	alpha = beta = 0.88, lambda = 2.25, gamma = 0.61, delta = 0.69

DEBATE AND EVIDENCE

Description versus experience. The overweighting of small probabilities is the most robust feature of the weighting function when people choose among *described* lotteries. Hertwig et al. (2004) asked participants to learn the payoff distributions by sampling from them, as people often do in daily life. In this "decisions from experience" paradigm, rare events were *underweighted*, in part because small samples often contain no rare event at all. The finding does not falsify cumulative prospect theory, which was formulated for described probabilities, but it limits its domain: predictions about insurance, warnings, or rare disasters depend on whether people know the probabilities or have learned them from their own history. Salience theory offers yet another account, in which the most striking payoffs, rather than the least likely ones, attract excess weight (Bordalo, Gennaioli, and Shleifer 2012).

RESEARCH TOOLKIT

Estimating prospect theory parameters.

- **Elicitation.** Certainty equivalents for simple gain and loss prospects identify curvature and weighting. Acceptance of mixed 50-50 gambles identifies loss aversion. Parameter-free methods such as the trade-off method avoid assumptions about functional form (Abdellaoui, Bleichrodt, and Paraschiv 2007, Wakker 2010).
- **Estimation.** Fit the model by maximum likelihood with a stochastic choice rule (for example, a logit on the difference in prospect values), or by hierarchical Bayesian methods that pool information across participants.
- **Identification.** Curvature and probability weighting trade off against each other when only a few probabilities are used. Vary probabilities and outcomes independently, and include mixed prospects to pin down λ.
- **Sample hypothesis.** *H1.* Participants will pay more to reduce a loss probability from .01 to 0 than from .11 to .10 (certainty effect for losses).

24.4 Loss Aversion

THEORY CARD

Core claim. Losses relative to a reference point have a greater impact on choice and on experience than gains of the same size. In riskless choice, this asymmetry produces the endowment effect (Section 23.4) and status quo bias. In risky choice, it produces aversion to mixed gambles with positive expected value.
Origins. Psychology and economics. Introduced as a property of the value function in Kahneman and Tversky (1979), formalized for riskless choice in Tversky and Kahneman (1991), and documented in exchange experiments by Kahneman, Knetsch, and Thaler (1990) and Kahneman, Knetsch, and Thaler (1991).
Level of analysis. Individual valuation and choice. Dyadic exchange (buyers and sellers, negotiators). Aggregate effects on trading volume, prices, and policy outcomes.
Key constructs. Loss aversion coefficient (λ), willingness to accept versus willingness to pay, reference point, mixed gambles, endowment effect, status quo bias.
Boundary conditions. Weak or absent for money spent as intended and for goods held for exchange (Novemsky and Kahneman 2005), among experienced traders (List 2003), and for very small amounts (Harinck et al. 2007). Its measured size depends on the range of gains and losses in the choice set (Walasek and Stewart 2015).
Related theories. Endowment effect (Section 23.4), Status quo bias, Regulatory focus theory (Section 20.1), Reference-dependent preferences.

24.4.1 Origins and core logic

Loss aversion began as one property of the value function, but it soon became the most widely used idea in behavioural economics. Tversky and Kahneman (1991) showed that it explains choice even when no risk is involved. In their **reference-dependent model of riskless choice**, preferences over bundles of goods depend on a reference bundle, and each attribute is evaluated with a value function that is steeper for losses. Two implications follow. First, indifference curves are *kinked* at the reference point, so the rate at which a person will trade one good for another depends on which good she currently holds. Second, the disadvantages of a change loom larger than its advantages, so options that represent a departure from the reference point are penalized.

The best-known demonstration is the mug experiment (Kahneman, Knetsch, and Thaler 1990), discussed in Chapter 23 (Section 23.4): owners of a randomly allocated coffee mug demanded roughly twice as much to sell it as non-owners were willing to pay. Kahneman, Knetsch, and Thaler (1991) connected the endowment effect, status quo bias, and loss aversion as manifestations of a single principle. Rabin's calibration theorem added a second, independent argument for loss aversion: modest-scale risk aversion in mixed gambles is difficult to explain in any other way (Rabin 2000).

24.4.2 Measuring loss aversion

Loss aversion is easy to describe but surprisingly difficult to measure. The coefficient λ has no single definition. With the power value function, λ is the ratio of the slopes of the loss and gain limbs only at an outcome of one unit, unless $\alpha = \beta$, so estimates depend on the currency and the functional form. Estimates from mixed gambles also depend on probability weighting: if the decision weights on a 50% chance of gaining and a 50% chance of losing differ, part of what looks like loss aversion

may be sign dependent weighting. Riskless measures, such as the ratio of willingness to accept to willingness to pay, capture loss aversion along with transaction costs, strategic misstatement, and reference-price effects. Neuroimaging evidence suggests a neural correlate: in mixed gambles, activity in valuation regions increased with potential gains and decreased more steeply with potential losses, and the asymmetry predicted individual differences in behavioural loss aversion (Tom et al. 2007).

The best current summary of the evidence is a large meta-analysis by A. L. Brown et al. (2024), who collected 607 empirical estimates of the loss aversion coefficient from 150 articles in economics, psychology, neuroscience, and other fields published between 1992 and 2017. Their estimate of the mean coefficient is 1.955, with a 95% probability that the true value lies between 1.820 and 2.102. Few observable features of study design were substantially associated with differences in mean estimates. Loss aversion is therefore real on average and somewhat smaller than the Tversky-Kahneman estimate of 2.25. Figure 24.8 translates these numbers into a behavioural quantity: the gain a decision maker requires to accept a 50-50 gamble with a loss of 100.

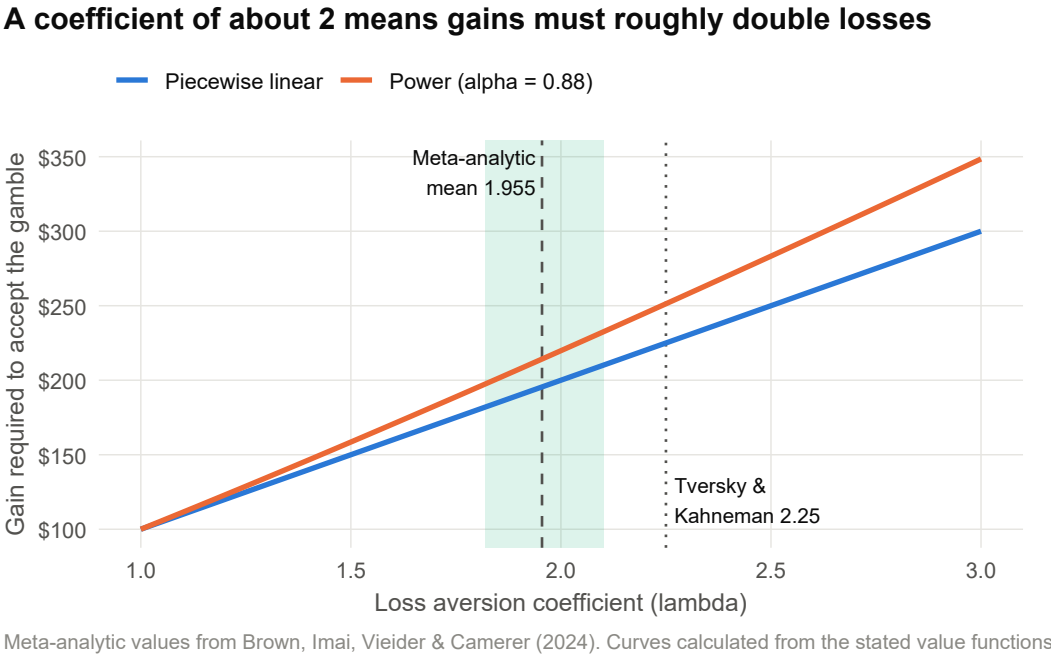


Figure 24.8. What a loss aversion coefficient means in practice: the minimum gain needed to accept a 50-50 gamble that risks a loss of 100, as a function of lambda, under a piecewise-linear value function and under the power value function with alpha = beta = 0.88 (no probability weighting). The shaded band is the 95% interval for the mean coefficient from the meta-analysis of Brown, Imai, Vieider and Camerer (2024), and the dashed line their mean estimate of 1.955. The dotted line marks the Tversky and Kahneman (1992) estimate of 2.25.

24.4.3 Boundaries and moderators

Loss aversion is not a constant of nature. Table 24.9 summarises the principal boundary conditions documented in the literature.

Table 24.9. Boundary conditions and moderators of loss aversion.

Moderator	Finding	Key source
Exchange goods and intended spending	Money spent as intended and goods held for resale are not coded as losses when given up	Novemsky & Kahneman (2005)
Market experience	Experienced traders of sports memorabilia show little or no endowment effect	List (2003)
Stake size	For very small amounts of money, gains can loom larger than losses	Harinck et al. (2007)
Range of outcomes in the task	Loss aversion disappears when gain and loss ranges are symmetric and reverses when losses span a wider range	Walasek & Stewart (2015)
Task format	Small, repeated, or experienced mixed gambles often show little loss aversion	Ert & Erev (2013)
Individual differences	Riskless and risky measures of loss aversion are related at the individual level and vary with personal characteristics	Gächter, Johnson & Herrmann (2022)
Decision maker and decision characteristics	Loss aversion is robust in large samples but systematically moderated	Mrkva et al. (2020)

Ert and Erev (2013) documented how features of the task, such as small stakes and repeated, experienced gambles, reduce measured loss aversion in risky choice. Gächter, Johnson, and Herrmann (2022) measured loss aversion in both riskless and risky tasks for the same individuals and found the two measures to be related and to vary with personal characteristics, which suggests that loss aversion is in part a stable individual difference. Two of the boundaries deserve elaboration. Novemsky and Kahneman (2005) argued that loss aversion applies to goods that are given up as *consumption* goods, not to goods given up as part of an intended exchange. A shopper who hands over $5 for a sandwich she meant to buy does not experience the $5 as a loss in the same way she would if it were stolen. This boundary explains why ordinary market transactions do not grind to a halt. Walasek and Stewart (2015) tested an account derived from **decision by sampling**, the idea that the subjective value of an amount depends on its rank among amounts recently encountered (Stewart, Chater, and Brown 2006). In many experiments that estimate loss aversion, the range of possible gains is about twice the range of possible losses, so a given loss ranks higher among losses than the same-sized gain ranks among gains. When Walasek and Stewart equalized the ranges, loss aversion disappeared, and when losses spanned the wider range, the asymmetry reversed. At least part of what is measured as loss aversion in such designs is a property of the choice context rather than of the person.

Figure 24.9 organizes the main explanations. The hedonic account is the original: losses simply hurt more than equivalent gains please. An attentional account holds that losses increase on-task attention and arousal, which can affect performance and choice even when they do not receive greater decision weight (Yechiam and Hochman 2013). The sampling account locates the asymmetry in the distribution of outcomes people compare against (Stewart, Chater, and Brown 2006, Walasek and Stewart 2015). The inertia account holds that many apparent demonstrations of loss aversion reflect a reluctance to act (Gal 2006).

DEBATE AND EVIDENCE

The loss of loss aversion? Gal and Rucker (2018b) argued that the evidence does not support loss aversion as a *general* principle. Building on Gal (2006), they contended that the endow-

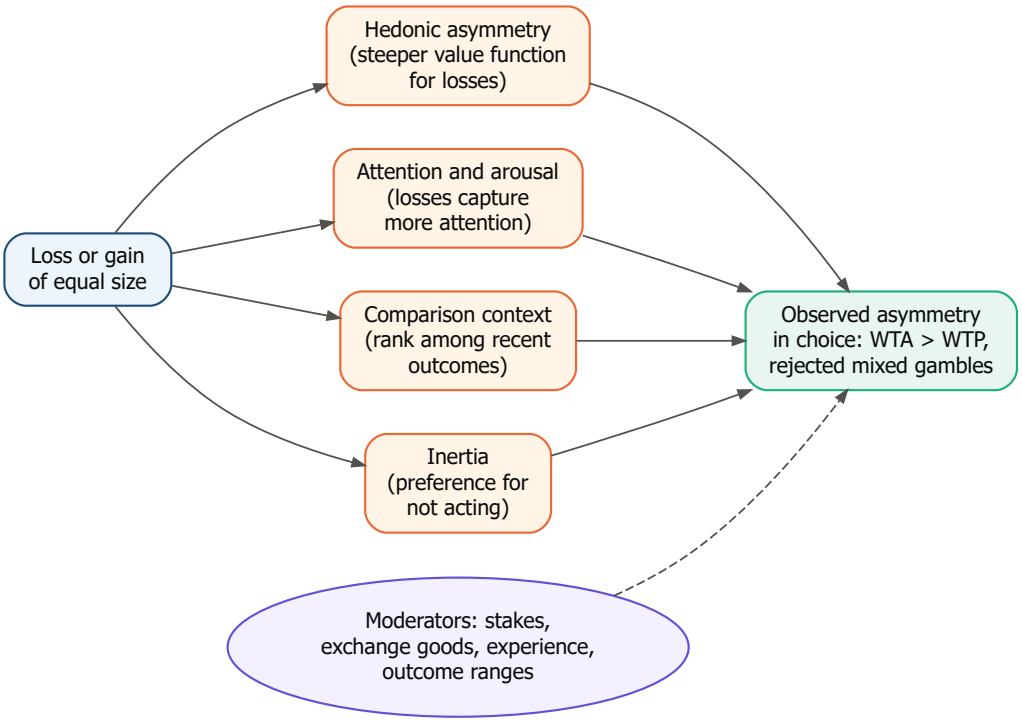


Figure 24.9. Competing accounts of the asymmetry between losses and gains. The classic account locates it in the value function itself. Alternatives attribute observed asymmetries to attention, to the context of comparison, or to a general preference for inaction. Moderators determine when the asymmetry is observed.

ment effect and status quo bias can be explained by psychological inertia, that losses do not have greater impact than gains in many studies once the status quo is not the default, and that the field had accepted the principle with too little scrutiny. The *Journal of Consumer Psychology* published commentaries alongside the critique. Simonson and Kivetz (2018) agreed that loss aversion is contingent but argued that its contingency does not make it less real or less useful. Higgins and Liberman (2018) argued that the debate turns on how reference points are construed and on the decision maker's motivational state. Gal and Rucker (2018a) replied by calling for a more contrarian science. In a later response, Mrkva et al. (2020) reported large-sample evidence that loss aversion is robust but systematically moderated by characteristics of the decision maker and the decision, so that the productive question is *when* and *for whom* losses loom larger. The meta-analysis of A. L. Brown et al. (2024) supports this view: the average coefficient is reliably above one, but the dispersion across studies is large. The debate has sharpened the theory rather than ended it: researchers are now expected to specify the reference point, rule out inertia, and consider outcome-range artefacts before invoking loss aversion.

24.4.4 Applications in business

BUSINESS CASE

Sales and production bonuses framed as losses. Hossain and List (2012) ran a natural field experiment in a Chinese high-tech manufacturing facility. Workers received conditional bonuses framed either as gains (a bonus paid if productivity met a target) or as losses (a provisional bonus that would be taken back if the target was missed). Both framings raised productivity. Teams responded more strongly to the loss frame, and the framing effect alone raised total team productivity by about 1%, with no sign of fading over the study period. In education, Fryer et al. (2022) paid teachers an up-front bonus that had to be returned if their students did not improve sufficiently. Pooling two waves, the loss-framed incentive improved mathematics achievement by an estimated 0.124 standard deviations, with large effects in the first wave and none in the second, while a gain-framed bonus produced a smaller effect of about 0.051 standard deviations that was not statistically significant. For managers designing sales compensation, the lesson is that *pre-paid, claw-back* incentives can motivate more than equivalent end-of-period bonuses, but they raise fairness, retention, and legal concerns, and their effects may not persist.

COMMON PITFALL

Calling every negative frame "loss aversion". A message that mentions a loss, a price above some benchmark, or a bad outcome is not evidence of loss aversion unless (1) a reference point is specified, (2) gains and losses of comparable magnitude are compared, and (3) alternative explanations such as inertia, attention, or the range of outcomes are ruled out. Many published "loss aversion" effects are better described as negativity effects or framing effects.

24.5 Reference-Dependent Preferences

THEORY CARD

Core claim. People derive utility from outcomes themselves (consumption utility) and from comparing outcomes with a reference point (gain-loss utility). In the leading modern formulation, the reference point is the person's recent rational expectations about outcomes, so that surprises, not outcomes alone, generate feelings of gain and loss.
Origins. Economics. Kőszegi and Rabin (2006) and Kőszegi and Rabin (2007), building on Kahneman and Tversky (1979) and Tversky and Kahneman (1991). Reviewed in O'Donoghue and Sprenger (2018).
Level of analysis. Individual choice and effort. Market implications for pricing, labour supply, insurance, and job search.
Key constructs. Reference point (status quo, expectation, goal, norm), consumption utility, gain-loss utility, stochastic reference point, personal equilibrium, choice-acclimating equilibrium, first-order risk aversion.
Boundary conditions. Expectations must be formed before the outcome is realized and must be salient. Evidence for expectations as reference points is mixed in replication, and the reference point may adapt with a lag.
Related theories. Loss aversion, Goal-setting theory (Section 19.4), Social comparison the-

> ory (Section 22.1), Expectancy-disconfirmation (Section 27.1).

24.5.1 What is the reference point?

Prospect theory says that outcomes are coded relative to a reference point but says little about where the reference point comes from. Kahneman and Tversky usually took it to be the status quo while acknowledging that aspirations, expectations, and norms could play the role. This flexibility is the theory's greatest strength and its greatest weakness. Applied work has used at least the candidates listed in Table 24.10.

Table 24.10. Candidate reference points used in research on reference-dependent preferences.

Reference point	Definition	Illustrative evidence
Status quo	Current endowment or default option	Endowment effect, default effects (Kahneman et al., 1990, Madrian & Shea, 2001)
Purchase price	Price originally paid for an asset	Disposition effect, condominium sellers (Odean, 1998, Genesove & Mayer, 2001)
Expectations	Recent beliefs about likely outcomes	Effort provision, football and family violence (Abeler et al., 2011, Card & Dahl, 2011)
Goals	Self-set or assigned targets	Goals as reference points, marathon bunching (Heath et al., 1999, Allen et al., 2017)
Salient numbers	Par, round numbers, thresholds	Golf putting, four-hour marathons (Pope & Schweitzer, 2011, Allen et al., 2017)
Past prices	Last price paid or average past price	Reference price effects in brand choice (Hardie et al., 1993)
Recent income	Earnings in the recent past	Job search around benefit cuts (DellaVigna et al., 2017)
Fair or past wages	Wage demanded or previously received	Police performance after arbitration losses (Mas, 2006)

Heath, Larrick, and Wu (1999) made the link between goals and reference points explicit: a goal inherits the properties of the value function. Because of diminishing sensitivity, a unit of progress is worth more the closer one is to the goal, so people work harder as they approach it, and because of loss aversion, the shortfall from a missed goal is felt as a loss. This idea connects prospect theory to goal-setting theory (Section 19.4) and the goal-gradient hypothesis (Section 19.5).

24.5.2 The Kőszegi-Rabin model

Kőszegi and Rabin (2006) proposed a disciplined answer to the reference-point question. A person's utility from an outcome c given a reference outcome r is

$$u(c \mid r) = m(c) + \mu\big(m(c) - m(r)\big),$$

where $m(\cdot)$ is ordinary **consumption utility** and $\mu(\cdot)$ is **gain-loss utility**, typically piecewise linear with slope η for gains and $\eta\lambda$ for losses. The key move is to set the reference point equal to the person's **recent rational expectations** about outcomes. Because expectations can be uncertain, the reference point is itself a probability distribution F, and an outcome is compared with every outcome the person expected, weighted by its probability:

$$U(c \mid F) = \int u(c \mid r)\, dF(r).$$

Expectations must also be internally consistent. A **personal equilibrium** is a plan that the person would want to carry out given that she expects to carry it out. This makes the reference point endogenous to the economic environment. The model predicts that willingness to pay for a good rises with the expected probability of buying it and with expected prices, and that a worker is less likely to keep working when income so far is unexpectedly high but more likely to show up when expected income is high (Kőszegi and Rabin 2006). Kőszegi and Rabin (2007) applied the framework to risk. When a risk is a surprise, the model reproduces classical prospect theory. When a risk is anticipated and the person can plan for it, it implies **first-order risk aversion**, and a prior expectation of taking risk reduces aversion both to that risk and to additional risk.

Figure 24.10 illustrates the core intuition. A person who will earn either 0 or 100 evaluates an outcome of 50 differently depending on how likely she thought the high outcome was. When she expected 0, an outcome of 50 is a pure gain. When she expected 100, the same outcome is a pure loss. With uncertain expectations, it is a mixture of both.

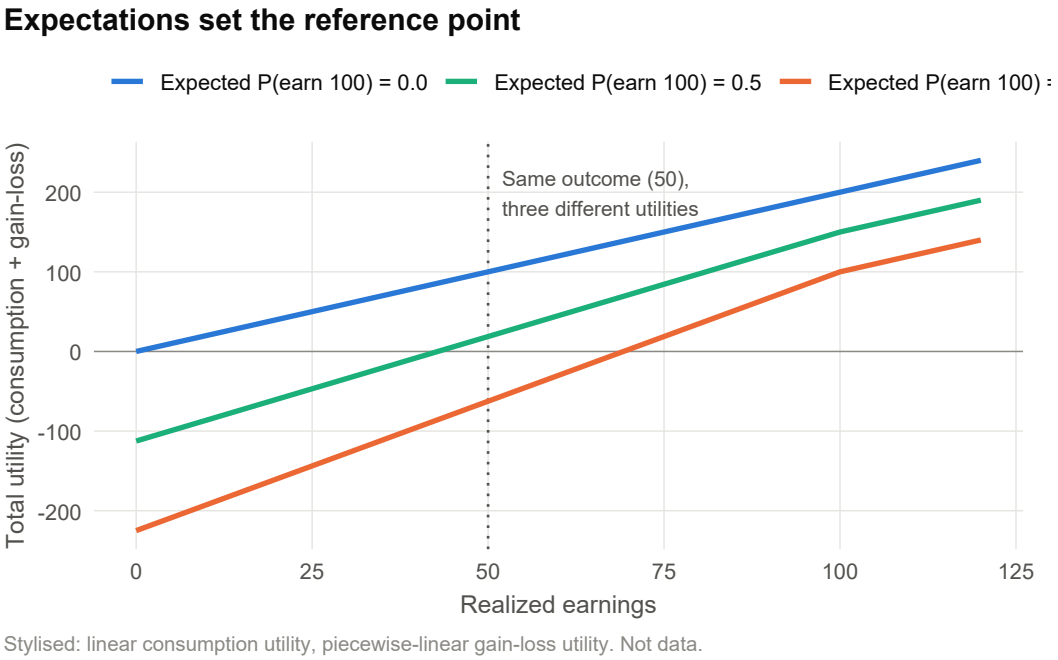


Figure 24.10. Stylised illustration of expectation-based reference points (Kőszegi & Rabin, 2006). A person expects to earn 100 with probability q and 0 otherwise. Total utility of a realized outcome equals consumption utility plus gain-loss utility relative to every outcome she expected, with eta = 1 and lambda = 2.25. The same outcome yields lower utility when higher earnings were expected.

24.5.3 Evidence from the laboratory

2 tested the expectations hypothesis directly. Participants performed a tedious counting task and decided when to stop. When they stopped, a coin flip determined whether they received their accumulated piece-rate earnings or a fixed payment announced in advance. The fixed payment manipulated expectations without changing the piece rate. When the fixed payment was high, participants

worked longer and earned more than when it was low, as models of expectation-based reference dependence predict. Ericson and Fuster (2011) found that manipulating the expected probability of keeping a good changed valuations and trading in the predicted direction. Gill and Prowse (2012) estimated a structural model of disappointment aversion in a real-effort competition and found that effort responded to a rival's effort in the way expectation-based reference points imply. Sprenger (2015) showed that risk attitudes depend on whether the reference point is a certain amount or a lottery, an "endowment effect for risk" consistent with stochastic reference points.

The laboratory evidence is not uniform. Heffetz and List (2014) found that manipulating the probability of keeping an object did not change the endowment effect in their experiments, which pointed to ownership rather than expectations as the operative reference point. Gneezy et al. (2017) extended the Abeler et al. design to a wider range of fixed payments and found that the predicted relationship between expectations and effort did not generalize. The present consensus is that expectations *can* serve as reference points, but which reference point operates, and how quickly it adapts, depends on the setting (O'Donoghue and Sprenger 2018).

24.5.4 Evidence from the field

Field studies have been central to the case for reference dependence because they involve real stakes, experienced decision makers, and outcomes that matter to them. Table 24.11 summarises influential studies.

Table 24.11. Selected field evidence on reference-dependent preferences.

Study	Setting	Reference point	Finding
Pope & Schweitzer (2011)	PGA Tour, over 2.5 million putts	Par on each hole	Golfers are more accurate putting for par than for birdie from the same distance
Allen et al. (2017)	About 9.8 million marathon finishing times	Round-number times (e.g., four hours)	Finishing times bunch just before round numbers, driven by effort near the finish
Card & Dahl (2011)	Police reports on National Football League Sundays	Pregame point spread (expected result)	Upset losses raise at-home male violence against partners by about 10%. Upset wins have little effect
Genesove & Mayer (2001)	Boston condominiums, 1990s	Original purchase price	Owners facing nominal losses set higher asking prices and sell more slowly
Mas (2006)	New Jersey police after final-offer arbitration	Wage demanded	Performance declined after unions lost arbitration
Fehr & Goette (2007)	Zurich bicycle messengers, temporary wage increase	Daily income target	Messengers worked more shifts but less effort per shift, a decline found among loss-averse messengers
DellaVigna et al. (2017)	Hungarian unemployment insurance reform	Recent income	Job search responds to benefit cuts as if the reference point adapts to recent income

Labour markets supply several of the entries. Mas (2006) found that the performance of New Jersey police declined after their unions lost final-offer arbitration, as if the wage they had demanded served as a reference point. Fehr and Goette (2007) offered Zurich bicycle messengers a temporary wage increase: the messengers worked more shifts but delivered less per shift, and the decline in effort per shift was concentrated among messengers who were loss averse in separate lottery choices,

consistent with a daily income target. DellaVigna et al. (2017) studied a reform of Hungarian unemployment insurance and found that job search responded to benefit cuts as if the reference point were recent income that adapts over time.

The football study is a particularly clean test of expectation-based reference points. Card and Dahl (2011) used betting markets to measure what fans expected before each game. Losses that were expected to be close had small and insignificant effects on domestic violence, and upset wins had little effect, but upset losses, defeats when the home team was predicted to win by four or more points, were followed by a 10% increase in at-home violence by men against their wives and girlfriends. The asymmetry between upset wins and upset losses mirrors loss aversion, and the dependence on the point spread mirrors expectation-based reference points.

DEBATE AND EVIDENCE

Do taxi drivers target a daily income? C. Camerer et al. (1997) studied New York City cab drivers, whose hourly earnings vary from day to day but are correlated within a day. A standard life-cycle model predicts that drivers work longer on high-wage days. Instead, estimated wage elasticities were significantly negative in two of three samples, averaging about minus one among inexperienced drivers. The authors interpreted this as evidence that drivers take decisions "one day at a time" and quit when they reach a loose daily income target, a reference point below which shortfalls feel like losses. Farber (2008) estimated a reference-dependent model directly and found that, although a daily reference income may exist, it varies so much from day to day and most shifts end so far from it that reference dependence has little predictive power. Using the complete record of New York taxi trips from 2009 to 2013, Farber (2015) found that drivers generally respond *positively* to both anticipated and unanticipated increases in earnings opportunities, that stopping depends much more on hours worked than on income earned, and that drivers who remain in the industry quickly learn to behave like optimizers. Two lines of work keep the reference-dependence account alive. V. P. Crawford and Meng (2011) showed that a model with expectation-based targets for *both* hours and income, in the spirit of Kőszegi and Rabin, reconciles much of the evidence. Thakral and Tô (2021) found that the likelihood of ending a shift responds more to earnings accumulated recently than to earlier earnings, consistent with a reference point that adapts during the day. The debate illustrates a general lesson: reference dependence is easy to assert and hard to identify, and its empirical success depends on modelling how the reference point is formed.

BUSINESS CASE

Pricing when customers expect prices. If the reference point is what customers expect to pay, firms face a demand curve with a kink at the expected price: a price slightly above expectations is coded as a loss and reduces demand sharply, while a price slightly below is only a modest gain. Heidhues and Kőszegi (2008) showed that, with loss-averse consumers and expectation-based reference points, competing firms may charge the same "focal" price even when their costs differ, and keep prices sticky as costs change. This logic explains why many retailers prefer stable everyday prices to frequent small changes, why price increases are often bundled with visible product changes, and why promotions that are too frequent can lower the reference price (Section 27.11) itself (Kalyanaram and Winer 1995). **Illustrative scenario:** a subscription software firm that raises its monthly price from $10 to $11 will see more churn among customers who expected $10 than a firm that launched at $11, even though

the final price is identical.

RESEARCH TOOLKIT

Studying reference points.

- **Sample hypotheses.** *H1.* Sales representatives will exert more effort in the final week of a quarter when they are just below their quota than when they are just above it. *H2.* Customers whose renewal price exceeds the price they paid last year will churn more than customers facing the same price without a prior lower price. *H3.* Upset losses in customer satisfaction surveys (scores below expectations set by marketing claims) will reduce repurchase more than upset gains increase it.
- **Designs.** Bunching estimators at salient thresholds (round numbers, quotas, tax kinks). Regression discontinuity around reference points. Experiments that manipulate expectations independently of outcomes, as in
    2. Structural models that let the reference point adapt.
- **Measures.** Expected outcomes (forecasts, betting lines, guidance), distance from goals, past prices and wages, and self-reported reference points.

COMMON PITFALL

Choosing the reference point after seeing the data. Because almost any number can be a reference point, a study that searches for one that "works" will usually find it. Pre-register the reference point, justify it from institutional features or measured expectations, and show that alternative candidates do not fit equally well.

24.6 Status Quo Bias

THEORY CARD

Core claim. People disproportionately choose the option that represents the current state of affairs or the default, even when switching is cheap and the alternatives are at least as good.
Origins. Economics and decision science. Named and documented by W. Samuelson and Zeckhauser (1988). Linked to loss aversion by Kahneman, Knetsch, and Thaler (1991) and Tversky and Kahneman (1991).
Level of analysis. Individual choice, with large aggregate effects in organizations and public policy (retirement saving, insurance, organ donation, subscriptions).
Key constructs. Status quo, default option, loss aversion, inertia, transaction and switching costs, regret, implied endorsement.
Boundary conditions. Weaker when the status quo is clearly inferior, when active choice is required, when decision makers are expert or highly motivated, and when switching is made salient and easy.
Related theories. Endowment effect (Section 23.4), Choice architecture (Section 23.10), Loss aversion, Escalation of commitment (Section 14.2).

24.6.1 Core logic and evidence

W. Samuelson and Zeckhauser (1988) presented participants with hypothetical decision problems, such as how to invest an inheritance, in two versions. In the neutral version, all options were presented on an equal footing. In the status quo version, one option was described as the current state of affairs. Options were chosen substantially more often when they were designated as the status quo. The authors then showed the same pattern in field data: in employee health plans and in retirement allocations, people rarely revisited choices they had made years earlier, even as their circumstances and the options changed.

Loss aversion offers a natural explanation (Kahneman, Knetsch, and Thaler 1991). If the status quo is the reference point, then every alternative involves giving up some features of the current option (losses) in exchange for others (gains). Because the losses loom larger, the current option wins unless the alternative is clearly better. But status quo bias is overdetermined. Table 24.12 lists the main explanations and the evidence that would distinguish among them, and Figure 24.11 shows how they combine.

Table 24.12. Explanations for status quo bias.

Explanation	Mechanism	Distinguishing prediction
Loss aversion	Departures from the reference point are coded as losses on some attributes	Bias larger when alternatives trade off attributes (some better, some worse)
Inertia and procrastination	Acting requires effort and attention that people postpone	Bias shrinks when active choice is required or switching is made effortless
Transaction and search costs	Switching has real costs	Bias scales with the cost of switching
Implied endorsement	Defaults are read as advice from the choice architect	Bias larger when the default setter is trusted and the decision maker is uncertain
Anticipated regret	Bad outcomes after switching feel worse than after staying	Bias larger when outcomes will be observed and attributed to the choice
Psychological inertia	A general preference for not acting without a compelling reason	Bias appears even when no losses are involved (Gal, 2006)

Insurance framing. E. J. Johnson et al. (1993) documented a natural experiment in automobile insurance. Around 1990, New Jersey and Pennsylvania both offered drivers a choice between a cheaper policy that limited the right to sue for pain and suffering and a more expensive policy that preserved the full right to sue. In New Jersey, the default was the limited policy. In Pennsylvania, the default was the full policy. In each state, most drivers ended up with their state's default, so that two economically similar populations held very different coverage. The same paper reported laboratory evidence that insurance choices violate basic laws of probability. In one study, people were willing to pay more for flight insurance covering death from terrorism than for insurance covering death from any cause, which necessarily includes terrorism.

Retirement saving. Madrian and Shea (2001) studied a large US firm that changed its 401(k) plan from opt-in to automatic enrollment for new employees. Participation rose sharply, from about 37% to 86% among employees with comparable tenure. Many automatically enrolled employees also kept both the default contribution rate (3% of pay) and the default investment (a money market fund), even though few employees hired before the change had chosen that combination. The authors attributed the effect to inertia and to employees' perception of the default as investment advice. The study helped launch the modern policy interest in defaults, discussed in Choice ar-

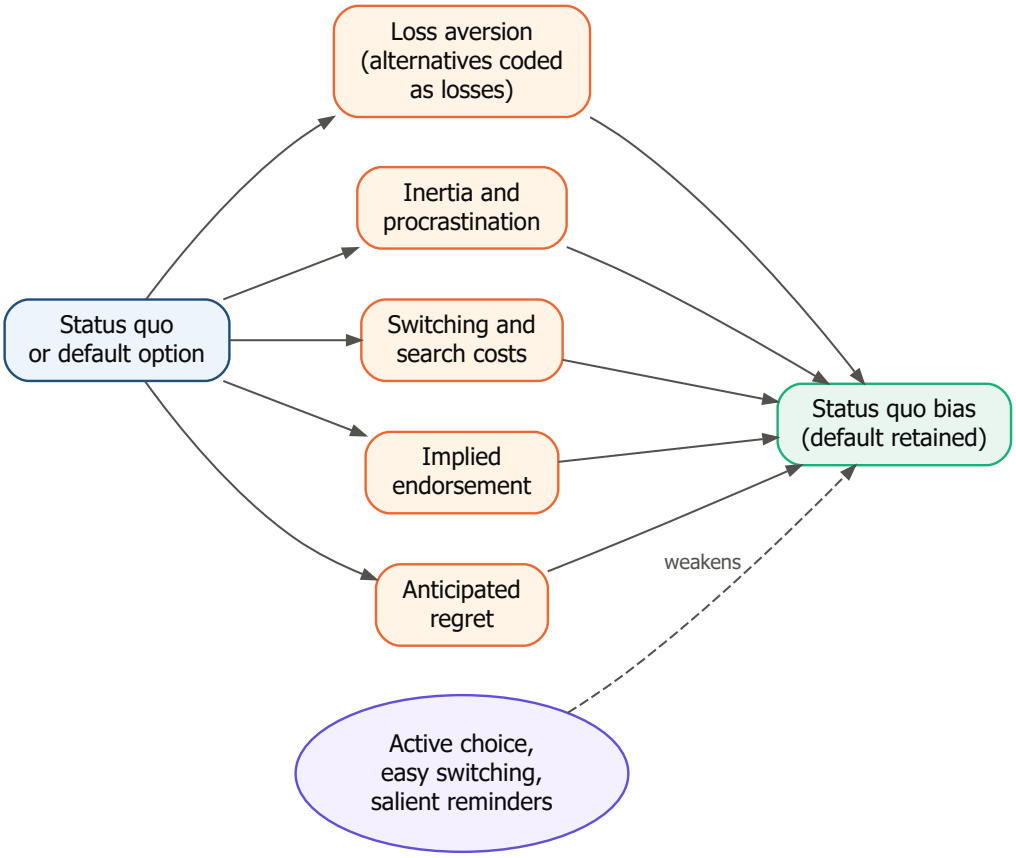


Figure 24.11. Multiple pathways to status quo bias. Loss aversion is one route among several, and choice architects can intervene at each one (active choice, simplified switching, careful selection of defaults).

chitecture (Section 23.10), and inspired *Save More Tomorrow*, which asks employees to commit in advance to raising their saving rate out of future pay increases, so that saving more never requires a cut in take-home pay that would be coded as a loss (R. H. Thaler and Benartzi 2004).

BUSINESS CASE

401(k) auto-enrollment. Automatic enrollment converts status quo bias from an obstacle into a lever. The design lessons from Madrian and Shea (2001) and R. H. Thaler and Benartzi (2004) are now standard practice. (1) Make participation the default. (2) Choose the default contribution rate and fund carefully, because many employees will keep them, so a low default rate can reduce saving for employees who would otherwise have chosen more. (3) Use automatic escalation tied to pay increases, so that higher contributions are never experienced as a loss of take-home pay. (4) Preserve easy opt-out, which keeps the policy libertarian and protects employees for whom the default is wrong. The same logic underlies auto-renewal of subscriptions, but there the firm, not the customer, benefits from inertia, which is why regulators increasingly require easy cancellation.

BUSINESS CASE

Free trials and the reference point. A free trial combines three prospect-theory mechanisms. During the trial the customer acquires a sense of ownership, so cancelling is coded as giving up something already possessed (Endowment effect (Section 23.4)). If the trial converts automatically into a paid subscription, continuing becomes the status quo. And the first payment, which arrives after the customer has adapted to having the service, may be evaluated against a reference point of zero cost. **Illustrative scenario:** a streaming service comparing an opt-in trial (customers must actively subscribe at the end) with an opt-out trial (customers are charged unless they cancel) should expect higher conversion under opt-out, but also more complaints, chargebacks, and regulatory risk, and lower retention among customers who converted only through inertia.

24.7 Myopic Loss Aversion

THEORY CARD

Core claim. Investors who are loss averse and who evaluate their portfolios frequently will find risky assets unattractive, because frequent evaluation reveals many short-run losses. The combination, *myopic loss aversion*, can explain why stocks must offer a large premium over bonds to be held.
Origins. Finance and behavioural economics. Benartzi and Thaler (1995), motivated by the equity premium puzzle of Mehra and Prescott (1985). Experimental tests in Gneezy and Potters (1997) and R. H. Thaler et al. (1997).
Level of analysis. Individual investors, with implications for asset prices and for the design of pension plans and investment platforms.
Key constructs. Loss aversion, evaluation period, narrow framing (choice bracketing), equity premium, feedback frequency.
Boundary conditions. Requires that gains and losses be evaluated over short horizons and in isolation from other wealth. Weaker when investors can commit to long horizons or receive aggregated feedback.
Related theories. Loss aversion, Mental accounting (Section 23.5), Behavioral finance (Section 16.8), Portfolio theory (Section 16.1).

24.7.1 The equity premium puzzle

Over long periods, US stocks have earned far higher average returns than short-term government bonds. Mehra and Prescott (1985) showed that standard consumption-based asset pricing models with plausible levels of risk aversion can account for only a small fraction of the historical premium: matching it requires a degree of risk aversion that seems implausibly high. This is the **equity premium puzzle**.

Benartzi and Thaler (1995) offered a behavioural explanation built from two ingredients. First, investors are loss averse. Second, even investors with long horizons evaluate their portfolios frequently. Stocks are volatile, so the more often their value is checked, the more often it shows a loss, and loss aversion makes each observed loss painful. Using simulations based on historical returns and the prospect theory parameters of Tversky and Kahneman (1992), they found that the size of

the equity premium is consistent with those parameters if investors evaluate their portfolios annually. The idea echoes a famous anecdote, discussed by Benartzi and Thaler (1999), in which a colleague of Paul Samuelson declined a single favourable bet but said he would accept a hundred of them.

Figure 24.12 reproduces the logic in a stylised simulation. It asks how large an equity premium is needed for a loss-averse investor to be indifferent between stocks and a riskless bond when gains and losses are evaluated over horizons from one month to ten years. Table 24.13 shows the reason: the probability of observing a loss falls steeply as the evaluation period lengthens.

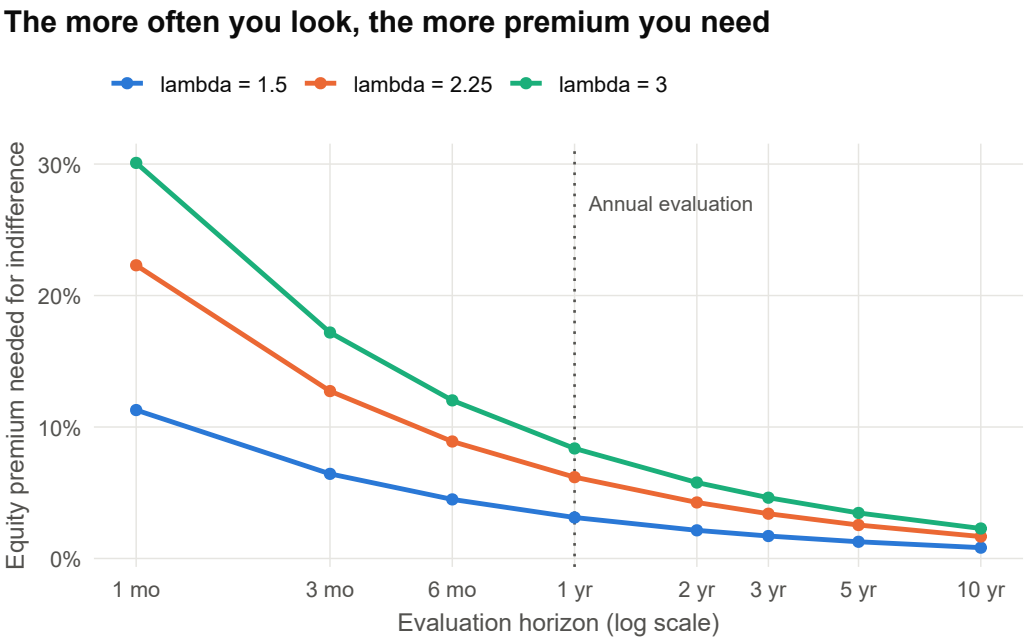


Figure 24.12. Stylised simulation of myopic loss aversion. For each evaluation horizon, the curve shows the annual equity premium at which a loss-averse investor with a piecewise-linear value function is indifferent between stocks and a riskless bond. Shorter horizons reveal more losses and require larger premiums. Stylised assumptions: lognormal stock returns with 20% annual volatility, a 1% riskless rate, no probability weighting. This is not the calibration of Benartzi and Thaler (1995).

Table 24.13. Stylised probability that a stock portfolio shows a nominal loss over different evaluation horizons (lognormal returns, 7% expected annual return, 20% volatility). Illustrative values, not historical estimates.

	Horizon	Probability of a loss
1 day	1 day	49%
1 month	1 month	47%
1 quarter	1 quarter	45%
1 year	1 year	40%
5 years	5 years	29%
10 years	10 years	21%
20 years	20 years	13%

24.7.2 Experimental evidence

The theory makes a sharp, testable prediction: holding the investment opportunity constant, people who receive feedback and make decisions more frequently should take *less* risk. Table 24.14 summarises the key tests.

Table 24.14. Experimental tests of myopic loss aversion.

Study	Design	Finding
Gneezy & Potters (1997)	Repeated bets on a favourable lottery, with outcomes evaluated each round or in blocks of rounds	The more frequently returns were evaluated, the less participants bet
Thaler, Tversky, Kahneman & Schwartz (1997)	Allocation between a risky and a safe fund, learned from experience, with feedback at different frequencies	Investors with the most frequent feedback took the least risk and earned the least. Eliminating losses increased risk taking
Haigh & List (2005)	Gneezy-Potters design with professional futures and options traders and students	Traders displayed myopic loss aversion to a greater extent than students
Benartzi & Thaler (1999)	Retirement allocation with one-year versus long-run return distributions	Showing the long-run distribution increased allocations to stocks

The professional trader result is especially important because it contradicts the common presumption that market experience eliminates behavioural anomalies. Haigh and List (2005) expected traders to show less myopic loss aversion than students. They found more. Asset pricing models have since incorporated these ideas more formally. N. Barberis, Huang, and Santos (2001) embedded loss aversion over financial wealth fluctuations, with a reference point that depends on prior gains and losses, into an equilibrium model that generates a high equity premium and volatile, predictable returns. N. Barberis, Huang, and Thaler (2006) argued that **narrow framing**, evaluating a gamble in isolation from other risks, is needed to explain why people reject small independent gambles and why many households do not hold stocks at all.

BUSINESS CASE

Investor feedback and platform design. Myopic loss aversion implies that the *frequency and format* of portfolio feedback is a design variable with consequences for investor welfare. Retirement plan statements that display long-horizon return distributions, apps that default to showing annual rather than daily performance, and communications that aggregate accounts rather than reporting each fund separately all lengthen the effective evaluation period. **Illustrative scenario:** a robo-advisor testing two dashboards, one showing daily changes in red and green and one showing progress toward a retirement goal updated quarterly, would expect the second to produce fewer panic sales after market declines and higher equity allocations, consistent with Gneezy and Potters (1997) and R. H. Thaler et al. (1997). The ethical question is whether hiding short-run volatility serves investors or simply keeps assets under management.

TRY IT YOURSELF

Use the simulation code behind Figure 24.12 (change `sigma`, `rf`, or the `lambda` values) to find the evaluation horizon at which a loss-averse investor with $\lambda = 2$ would hold stocks at an annual premium of 4%. How does the answer change if volatility is 15% rather than 20%?

24.8 Disposition Effect

THEORY CARD

Core claim. Investors are more likely to sell assets that have risen in value since purchase than assets that have fallen, holding losers too long and selling winners too early.
Origins. Finance. Named by Shefrin and Statman (1985), documented in brokerage data by Odean (1998), in housing by Genesove and Mayer (2001), and in the laboratory by M. Weber and Camerer (1998).
Level of analysis. Individual investors and property owners, with implications for trading volume, price momentum, and real estate markets.
Key constructs. Purchase price as reference point, realized versus paper gains and losses, proportion of gains realized (PGR), proportion of losses realized (PLR), realization utility, regret.
Boundary conditions. Weaker among more experienced and sophisticated investors, and reversed in contexts with strong tax incentives to realize losses. Its link to prospect theory's value function is debated.
Related theories. Loss aversion, Mental accounting (Section 23.5), Escalation of commitment (Section 14.2), Behavioral finance (Section 16.8).

24.8.1 Origins and core logic

Shefrin and Statman (1985) coined the term **disposition effect** and proposed an explanation combining four elements: prospect theory (with the purchase price as reference point), mental accounting (each stock held in its own account), regret aversion (realizing a loss confirms a mistake), and self-control (investors may need rules to overcome the reluctance to sell losers). The prospect theory logic is that a stock trading above its purchase price is in the concave, risk-averse region of the value function, which encourages locking in the gain, while a stock trading below its purchase price is in the convex, risk-seeking region, which encourages holding on in the hope of breaking even.

24.8.2 Evidence

Odean (1998) developed the standard empirical test using the trading records of 10,000 accounts at a large discount brokerage. On each day an investor sold any stock, every stock in the portfolio was classified as a realized gain, a paper (unrealized) gain, a realized loss, or a paper loss. The **proportion of gains realized**, PGR, is the number of realized gains divided by realized plus paper gains. The **proportion of losses realized**, PLR, is defined analogously. Odean found that investors realized gains at a much higher rate than losses, that the pattern was not explained by portfolio rebalancing or by the higher trading costs of low-priced stocks, and that it was not justified by subsequent performance: the winners investors sold went on to outperform the losers they kept. The code below implements the PGR and PLR calculation for any position-level dataset and applies it to simulated data.

```
# PGR and PLR from position-level data (Odean, 1998).
# d: one row per position held on a day on which the investor sold something,
#    with logical columns `gain` (price above purchase price) and `sold`.
pgr_plr <- function(d) {
  rg <- sum(d$gain & d$sold)
```

```
  pg <- sum(d$gain & !d$sold)
  rl <- sum(!d$gain & d$sold)
  pl <- sum(!d$gain & !d$sold)
  pgr <- rg / (rg + pg)
  plr <- rl / (rl + pl)
  se <- sqrt(pgr * (1 - pgr) / (rg + pg) + plr * (1 - plr) / (rl + pl))
  data.frame(PGR = pgr, PLR = plr, Difference = pgr - plr, `t statistic` = (pgr - plr) / se,
             check.names = FALSE)
}

# Simulated (illustrative) investors: selling propensity rises with the size of
# the unrealized return in either direction and jumps just above zero.
set.seed(1998)
sim <- data.frame(investor = rep(1:400, each = 8 * 30),
                  day = rep(rep(1:30, each = 8), 400),
                  ret = rnorm(400 * 8 * 30, 0.02, 0.25))
sim$gain <- sim$ret > 0
sim$sold <- runif(nrow(sim)) < plogis(-2.8 + 0.6 * sim$gain + 2.5 * abs(sim$ret))
sale_days <- unique(sim[sim$sold, c("investor", "day")])
sim_sd <- merge(sim, sale_days)                    # keep only days with a sale
round(pgr_plr(sim_sd), 3)
```

```
#>     PGR   PLR Difference t statistic
#> 1 0.234 0.143      0.091       29.86
```

Figure 24.13 shows the stylised disposition effect in simulated trading data.

The effect is not confined to stock markets. Genesove and Mayer (2001) found that Boston condominium owners facing nominal losses in the 1990s set asking prices higher by 25-35% of the difference between the property's expected selling price and their original purchase price, achieved higher selling prices by 3-18% of that difference, and sold much more slowly than other sellers. The list price effects were about twice as large for owner-occupants as for investors but held for both. In a controlled experimental market where the expected returns of assets were known, M. Weber and Camerer (1998) also found that participants sold more winners than losers.

DEBATE AND EVIDENCE

Is the disposition effect really prospect theory? The textbook explanation has been challenged on theoretical and empirical grounds. N. Barberis and Xiong (2009) modelled a prospect-theory investor carefully and found that when preferences are defined over annual gains and losses, the model often predicts the *opposite* of the disposition effect. A loss-averse investor buys a stock only if its expected return is high. After a gain she moves away from the kink in the value function and is willing to hold more, while after a loss she moves toward the kink, where loss aversion bites hardest, and reduces her position. The disposition effect emerges more reliably if utility is derived from *realizing* gains and losses, a preference that N. Barberis and Xiong (2009) call **realization utility**. Kaustia (2010) showed that the propensity to sell jumps at a zero return but is approximately constant over a wide range of losses and constant or increasing over gains, a pattern hard to reconcile with the curvature of the value

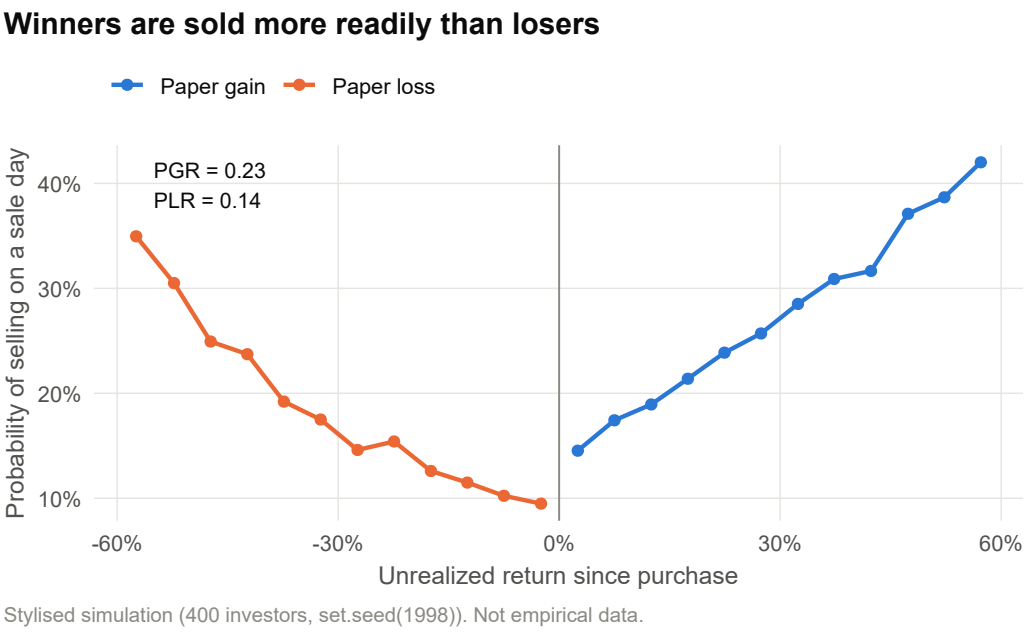


Figure 24.13. Stylised disposition effect in simulated trading data. The probability of selling a position on a sale day is plotted against its unrealized return since purchase. Selling rises with the size of the return in both directions and jumps at zero, so the proportion of gains realized (PGR) exceeds the proportion of losses realized (PLR). Illustrative simulation, not empirical data.

function. Ben-David and Hirshleifer (2012) found that selling is V-shaped in past returns: investors are more likely to sell large winners *and* large losers than small ones, which suggests belief-based trading rather than a preference for realizing gains. Imas (2016) added a twist: after a *realized* loss people take less risk, but after a *paper* loss they take more, so the act of realization itself changes the reference point. The consensus is that the disposition effect is a robust empirical regularity whose cause is more likely realization utility, regret, or beliefs than the S-shaped value function alone.

Table 24.15 presents the explanations for the disposition effect.

Table 24.15. Explanations for the disposition effect.

Explanation	Core idea	Key evidence or source
Prospect theory value function	Concavity above and convexity below the purchase price	Shefrin & Statman (1985), but see Barberis & Xiong (2009)
Realization utility	Investors derive utility from the act of realizing a gain or loss	Barberis & Xiong (2009)
Regret and self-justification	Selling a loser confirms a mistake	Shefrin & Statman (1985)
Beliefs in mean reversion	Investors expect losers to rebound and winners to fall back	Odean (1998) rules out several rational motives, Ben-David & Hirshleifer (2012)
Break-even seeking	Paper losses invite risk taking to get back to the reference point	Thaler & Johnson (1990), Imas (2016)
Tax considerations	Realizing losses reduces tax liability	Predicts the opposite pattern, so it cannot explain the effect

BUSINESS CASE

Brokerage and wealth-management practice. The disposition effect has practical consequences for retail investors, who sell winners that keep winning and hold losers that keep losing, and for the firms that serve them. Practices that weaken the effect include automatic rebalancing rules, pre-set stop-loss and take-profit rules decided before positions move, year-round tax-loss harvesting that reframes a realized loss as a tax gain, and statements that report portfolio-level rather than position-level returns, which makes the purchase price of each holding less salient. In real estate, agents can use the evidence of Genesove and Mayer (2001) to anticipate that sellers facing nominal losses will over-price and stay on the market longer, and price-setting conversations can shift the reference point from the purchase price to recent comparable sales.

24.9 Further Applications

Few theories have travelled as far as prospect theory. Table 24.16 maps its components to application domains, and the subsections that follow discuss the most important ones for business and policy. Each application relies on a different part of the theory, which is worth keeping in mind: evidence for one component (for example, probability weighting in lottery markets) is not evidence for another (for example, loss aversion in labour supply).

Table 24.16. Prospect theory applications by domain.

Domain	Component of the theory	Signature prediction	Illustrative evidence
Pricing	Reference dependence, loss aversion	Surcharges and price increases hurt more than equivalent discounts and decreases help	Thaler (1980), Hardie et al. (1993)
Promotions and messaging	Framing, loss aversion	Loss-framed messages can motivate more than gain-framed ones	Ganzach & Karsahi (1995)
Insurance	Probability weighting, loss aversion	Over-insurance of modest risks, preference for low deductibles	Johnson et al. (1993), Sydnor (2010)
Lotteries and skewed assets	Overweighting of small probabilities	Demand for and overpricing of lottery-like payoffs	Barberis & Huang (2008)
Asset pricing	Myopic loss aversion, narrow framing	High equity premium, non-participation	Benartzi & Thaler (1995), Barberis et al. (2001)
Trading	Purchase price as reference point	Disposition effect	Odean (1998), Genesove & Mayer (2001)
Labour and effort	Expectations and goals as reference points	Bunching at targets, effort responds to expectations	Abeler et al. (2011), Allen et al. (2017)
Tax compliance	Loss aversion around a zero balance due	More avoidance when a payment is owed than when a refund is due	Engström et al. (2015), Rees-Jones (2018)
Health and energy	Goal framing, loss-framed incentives	Deposit contracts and loss-framed incentives can outperform equivalent rewards	Patel et al. (2016), Halpern et al. (2015)
Negotiation and politics	Reference dependence, reflection effect	Loss frames make negotiators tougher and leaders more risk seeking	Neale & Bazerman (1985), Levy (1992)

24.9.1 Pricing and reference prices

The idea that customers code prices relative to a reference point is the most direct business application of the theory. R. Thaler (1980) noted that the credit card industry preferred that any price difference between card and cash payments be described as a "cash discount" rather than a "credit card surcharge", anticipating that a surcharge would be coded as a loss. Reference prices also govern perceptions of fairness: customers accept price increases that pass on higher costs but regard increases that exploit a surge in demand as unfair, because the previous price acts as a reference transaction (Kahneman, Knetsch, and Thaler 1986). In scanner-panel data, Hardie, Johnson, and Fader (1993) modelled brand choice with reference-dependent utility and found evidence of loss aversion with respect to the reference brand's price and quality: consumers were more responsive to being worse off than to being better off relative to the brand they had last bought. Kalyanaram and Winer (1995) summarized the empirical generalizations of reference price research, beginning with the finding that reference prices significantly affect demand.

BUSINESS CASE

Framing price differences. The same price structure can be presented as a surcharge (the base price is the low price, and some customers pay extra) or as a discount (the base price is the high price, and some customers pay less). Prospect theory predicts that customers who end up paying the higher price will feel worse under the surcharge frame, and customers who pay the lower price will feel only mildly better under the discount frame. Airlines, utilities, and payment providers face this choice every time they differentiate prices by channel, time, or payment method. Combined with mental accounting, the theory also predicts that many small fees, each coded as a separate loss, will feel worse than one all-inclusive price (Mental accounting (Section 23.5)). The managerial implication is not always to choose the discount frame: a surcharge may be the more effective way to steer customers toward a cheaper channel precisely because it is more aversive, but it carries reputational costs that the discount frame avoids.

24.9.2 Promotions and marketing messages

Loss framing is a standard tool in promotional copy ("Don't miss out", "You are losing $X a month"). The evidence is supportive but more nuanced than practitioners assume. In a field experiment with credit card holders, Ganzach and Karsahi (1995) found that a message emphasizing what customers lose by not using their card produced more subsequent card use than a message emphasizing what they gain by using it. Goal-framing effects of this kind are, however, distinct from risky-choice framing and attribute framing, and they vary in size and even direction across settings (Levin, Schneider, and Gaeth 1998). Firms should treat loss framing as a hypothesis to test in controlled experiments.

24.9.3 Insurance, lotteries, and skewed assets

Insurance and lotteries are where probability weighting does most of its work. Overweighting of small probabilities makes people willing to pay well above expected value both to eliminate a small chance of loss (insurance) and to buy a small chance of a large gain (lotteries), the two "low probability" cells of the fourfold pattern. Reviewing three decades of applications in economics, N. C. Barberis (2013) highlighted insurance and finance as domains in which the theory helps to explain facts that are otherwise hard to account for. Sydnor (2010) studied deductible choices in home insurance and found that most customers chose low deductibles despite premiums far above the

expected value of the additional coverage, a level of risk aversion over modest stakes that a standard expected-utility model can fit only with implausible parameters. In financial markets, N. Barberis and Huang (2008) showed that investors with cumulative prospect theory preferences overvalue securities with positively skewed, lottery-like returns, which can make such securities earn low average returns, and N. Barberis, Mukherjee, and Wang (2016) found that stocks whose past return distributions are more attractive under prospect theory tend to have lower subsequent returns.

BUSINESS CASE

Insurance deductibles and extended warranties. Low deductibles, extended warranties, and add-on protection plans sell well even when their price is many times the expected cost of the risk they cover (Sydnor 2010). Prospect theory offers two reasons. Small probabilities of loss are overweighted, and paying a deductible after a claim is coded as an additional, salient loss on top of the damage itself. Insurers can use this insight responsibly, for example by offering a low-deductible option alongside clear information on expected claim costs, or exploitatively, by pricing low-deductible options far above cost. Regulators and consumer advocates increasingly treat the second as a consumer-protection issue.

24.9.4 Labour and tax compliance

Reference points shape how people respond to taxes. In Sweden, taxpayers receive a preliminary statement showing whether they owe money or are due a refund. Using data on 3.6 million Swedish taxpayers and a regression kink and discontinuity design, Engström et al. (2015) found strong causal effects of a preliminary deficit on the probability of claiming deductions, and concluded that calibrating withholding so that most taxpayers receive refunds would increase compliance. In US tax returns, Rees-Jones (2018) estimated that taxpayers facing a payment on tax day reduce their tax liability by $34 more than taxpayers owed a refund. In both studies, the reference point is a zero balance due, and owing money is coded as a loss. Labour-market applications, including effort provision, daily labour supply, and job search, are discussed in Reference-dependent preferences.

24.9.5 Sports

Sport provides unusually clean tests because outcomes are precisely measured and reference points are institutionally defined: par in golf (Pope and Schweitzer 2011), round-number finishing times in marathons (E. J. Allen et al. 2017), and betting lines in football (Card and Dahl 2011). Sports settings also show that reference dependence survives expertise, high stakes, and competition, which answers the common objection that behavioural anomalies disappear among experienced professionals operating in markets.

24.9.6 Public policy and health messaging

Rothman and Salovey (1997) proposed an influential refinement of goal framing for health communication. Behaviours that *prevent* illness (sunscreen, exercise) are perceived as low risk, so gain-framed messages should be more persuasive. Behaviours that *detect* illness (screening) carry the risk of an unwelcome discovery, so loss-framed messages should be more persuasive, by analogy with the risk seeking that prospect theory predicts in the loss domain. Early evidence was encouraging: women who read pamphlets emphasizing the negative consequences of *not* performing breast self-examination showed more favourable attitudes and behaviour than those who read pamphlets

emphasizing the benefits of performing it (Meyerowitz and Chaiken 1987). Meta-analytic evidence is more sobering: across disease-prevention studies, the average difference between gain-framed and loss-framed messages is small (O'Keefe and Jensen 2007).

Loss aversion has had more consistent success in the design of *incentives* than of messages. In a randomized trial with overweight and obese adults, Patel et al. (2016) compared financial incentives for meeting a daily step goal that were framed as gains, as lotteries, or as losses (money allocated in advance and taken away for each day the goal was missed). Only the loss-framed incentive increased goal attainment significantly relative to control. In a large trial of smoking-cessation programmes, Halpern et al. (2015) found that deposit contracts, in which participants put their own money at risk, were accepted by fewer people than reward programmes, but were more effective among those who accepted them, illustrating the trade-off between the motivational power of loss-based incentives and their unattractiveness at enrolment. Volpp et al. (2008) found that both deposit contracts and lottery-based incentives produced more weight loss than a control condition during a weight-loss intervention.

BUSINESS CASE

Loss-framed energy and health messages. In a classic quasi-experiment, home energy auditors were trained to communicate their findings vividly and to frame the potential savings as money the household was *losing* through leaks and poor insulation. Homeowners who received these audits were more likely to apply for retrofit programmes than homeowners who received standard audits (Gonzales, Aronson, and Costanzo 1988). Utilities and health insurers have since used similar framing alongside other behavioural levers such as social comparisons (Social norms (Section 22.5)), and loss-framed incentive designs such as pre-funded wellness accounts draw directly on Patel et al. (2016). The lesson from the meta-analytic evidence (O'Keefe and Jensen 2007) is that loss frames are not a universal improvement: they work best when the behaviour involves risk or uncertainty and when the reference point is clear to the audience.

24.9.7 Negotiation and politics

In a negotiation, what one side gives up is coded as a loss and what it receives as a gain, so concessions loom larger to the conceder than to the recipient. Neale and Bazerman (1985) showed that framing changes bargaining: negotiators instructed to think about their outcomes as gains were more concessionary and completed more transactions than negotiators instructed to think about them as losses. In international relations, Levy (1992) argued that prospect theory explains why states and leaders accept large risks to avoid losses (for example, to recover lost territory or status) that they would not accept to make equivalent gains, while noting the analytical difficulty of identifying the reference point of a state.

RESEARCH TOOLKIT

Bringing prospect theory to a new domain.

- **Specify the component.** State whether the prediction rests on reference dependence, loss aversion, diminishing sensitivity, or probability weighting.
- **Identify the reference point ex ante.** Use institutional features (par, purchase price, zero balance due, quota) or measured expectations.
- **Find a discontinuity.** Reference points create kinks and notches. Bunching and regres-

sion discontinuity designs exploit them.
- **Rule out alternatives.** Inertia, attention, transaction costs, beliefs, and the range of outcomes can all mimic loss aversion.
- **Sample hypothesis.** *H1.* Customers who receive a loss-framed renewal reminder ("You will lose your saved settings") will renew at a higher rate than customers who receive an equivalent gain-framed reminder, and the effect will be larger among customers with higher usage (a stronger sense of ownership).

24.10 Replication and Robustness

Prospect theory is built on demonstrations that are cheap to run, which makes it unusually easy to test whether its foundations replicate. The answer is encouraging. Ruggeri et al. (2020) repeated the original 1979 choice problems with 4,098 participants in 19 countries and 13 languages, adjusting only the currency and amounts to local conditions and requiring every participant to answer every item. The results replicated for 94% of items, with some attenuation of effect sizes, and 12 of 13 theoretical contrasts replicated, with complete replication in some countries. The authors also documented meaningful heterogeneity between countries and within individuals. The framing effect in the "Asian disease" problem (Tversky and Kahneman 1981) was among the effects that replicated in the first Many Labs project (R. A. Klein et al. 2014).

The picture is more mixed for the theory's *extensions* and *field applications*. Table 24.17 summarises the state of the evidence as of this writing.

Table 24.17. What replicates and what does not: the state of the evidence on prospect theory.

Claim	Status	Evidence
Certainty, reflection, and isolation effects in described lotteries	Replicates	Ruggeri et al. (2020): 94% of items and 12 of 13 contrasts, with some attenuation
Risky-choice framing (Asian disease)	Replicates	Many Labs 1 (Klein et al., 2014)
Loss aversion on average	Replicates, smaller than 2.25	Meta-analytic mean 1.955 (Brown et al., 2024)
Loss aversion as a universal constant	Does not hold	Context and range effects, small stakes, exchange goods (Walasek & Stewart, 2015, Harinck et al., 2007, Novemsky & Kahneman, 2005)
Overweighting of rare events	Holds for description, often reverses for experience	Hertwig et al. (2004)
Expectations as reference points	Mixed	Abeler et al. (2011) versus Gneezy et al. (2017) and Heffetz & List (2014)
Daily income targeting by taxi drivers	Contested	Camerer et al. (1997) versus Farber (2008, 2015), Thakral & Tô (2021)
Myopic loss aversion	Replicates in experiments, including with professionals	Gneezy & Potters (1997), Haigh & List (2005)
Disposition effect as an empirical regularity	Robust	Odean (1998), Genesove & Mayer (2001), Weber & Camerer (1998)
Disposition effect caused by the S-shaped value function	Doubtful	Barberis & Xiong (2009), Kaustia (2010), Ben-David & Hirshleifer (2012)

Reference dependence in high-stakes field settings	Robust	Pope & Schweitzer (2011), Allen et al. (2017), Card & Dahl (2011)

Three lessons emerge. First, the *qualitative* patterns that motivated prospect theory are among the most reliable findings in the behavioural sciences. Second, *quantitative* parameters, especially the loss aversion coefficient, vary with people, tasks, and contexts, so a single "true" value should not be imported uncritically into applied models. Third, the theory's empirical success in a new domain depends on how well the reference point is specified, which is a modelling choice rather than a prediction of the theory itself (N. C. Barberis 2013, O'Donoghue and Sprenger 2018).

COMMON PITFALL

Importing parameter values without checking the context. Simulations and structural models often plug in $\alpha = 0.88$, $\lambda = 2.25$, and $\gamma = 0.61$ as if they were universal constants. They are median estimates from one study of described lotteries. When results depend on these values, report sensitivity analyses over plausible ranges, for example λ from about 1.5 to 2.5, and prefer estimates from populations and tasks similar to the application.

24.11 Chapter Summary

KEY TAKEAWAYS

- **Expected utility theory** remains the normative benchmark, but the Allais and Ellsberg paradoxes and Rabin's calibration theorem show that it cannot describe how people actually evaluate risk, especially over modest stakes (Allais 1953, Ellsberg 1961, Rabin 2000).
- **Prospect theory** replaces final wealth with gains and losses relative to a reference point, and combines a value function (diminishing sensitivity and loss aversion) with decision weights (overweighting of small probabilities) (Kahneman and Tversky 1979).
- **Cumulative prospect theory** applies weights to cumulative probabilities, preserving dominance and generating the fourfold pattern of risk attitudes. Its median parameters, $\alpha = \beta = 0.88$, $\lambda = 2.25$, $\gamma = 0.61$, and $\delta = 0.69$, are useful defaults, not constants (Tversky and Kahneman 1992).
- **Loss aversion** is real on average, with a meta-analytic mean coefficient near 2, but it is moderated by stakes, goods, experience, and the range of outcomes (A. L. Brown et al. 2024, Mrkva et al. 2020, Walasek and Stewart 2015).
- **Reference points** can be the status quo, expectations, goals, or salient numbers. Expectation-based models discipline the choice but have mixed experimental support (Kőszegi and Rabin 2006, Abeler et al. 2011, Gneezy et al. 2017).
- **Status quo bias**, **myopic loss aversion**, and the **disposition effect** are the theory's most consequential applications, with large effects on retirement saving, asset prices, and trading (Madrian and Shea 2001, Benartzi and Thaler 1995, Odean 1998).
- The original choice patterns **replicate** across 19 countries (Ruggeri et al. 2020). The theory's weakest point is not its foundations but the specification of the reference point in each new application.

Table 24.18 presents the comparison of expected utility, prospect theory, cumulative prospect the-

ory, and expectation-based reference-dependent preferences.

Table 24.18. Comparison of expected utility, prospect theory, cumulative prospect theory, and expectation-based reference-dependent preferences.

Feature	Expected utility	Prospect theory (1979)	Cumulative prospect theory (1992)	Expectation-based reference dependence (2006)
Carriers of value	Final wealth	Gains and losses	Gains and losses	Consumption plus gains and losses
Reference point	None	Usually the status quo, unspecified	Usually the status quo, unspecified	Recent rational expectations (stochastic)
Risk attitude from	Curvature of utility	Value function and decision weights	Value function and rank-dependent weights	Consumption utility and gain-loss utility
Probability treatment	Linear	Decision weights on individual probabilities	Weights on cumulative probabilities	Usually linear
Loss aversion	No	Yes	Yes (lambda)	Yes (lambda), relative to expectations
Stochastic dominance	Satisfied	Can be violated	Satisfied	Satisfied in standard applications
Signature prediction	Consistent preferences across common consequences	Certainty, reflection, isolation effects	Fourfold pattern of risk attitudes	Surprises, not outcomes, drive gain-loss feelings
Key source	von Neumann & Morgenstern (1944)	Kahneman & Tversky (1979)	Tversky & Kahneman (1992)	Kőszegi & Rabin (2006)

24.12 Discussion Questions

1. Rabin's calibration theorem is sometimes described as the strongest argument for prospect theory. Explain the theorem in your own words. Does it show that expected utility is wrong, or only that wealth is the wrong argument of the utility function?
2. Choose a pricing decision in a firm you know (a price increase, a new fee, a change in subscription terms). Identify at least three candidate reference points customers could use, and design a field experiment that would reveal which one drives their response.
3. The debate over loss aversion has moved from *whether* losses loom larger than gains to *when* and *for whom*. Propose a study that would distinguish loss aversion from inertia in a setting of your choice.
4. Using Figure 24.12 and the evidence on myopic loss aversion, argue for or against a regulation that would require retirement plan providers to show long-horizon return distributions by default. Whose interests would such a rule serve?
5. The disposition effect is robust, but its cause is debated. What data from a brokerage platform would allow you to distinguish realization utility, beliefs in mean reversion, and the S-shaped value function as explanations?
6. Loss-framed incentives (deposit contracts, claw-back bonuses) can be more effective than equivalent rewards but are less attractive at enrolment and can feel unfair. Under what conditions should an employer or health insurer use them?

24.13 Further Reading

- Kahneman and Tversky (1979): the original paper. Read the choice problems first, then the model, and notice how much of later research is foreshadowed in the discussion of editing.
- Tversky and Kahneman (1992): the cumulative version, with the parameter estimates that became the field's defaults.
- Wakker (2010): the definitive technical treatment of prospect theory for risk and ambiguity, with careful attention to measurement.
- N. C. Barberis (2013): an accessible assessment of thirty years of applications in economics, especially finance and insurance.
- O'Donoghue and Sprenger (2018): a rigorous handbook review of reference-dependent preferences, including the Kőszegi-Rabin model and the field evidence.
- A. L. Brown et al. (2024): the most comprehensive meta-analysis of loss aversion estimates. Pair it with Gal and Rucker (2018b) and Mrkva et al. (2020) for the debate.
- Ruggeri et al. (2020): a multinational replication of the original problems and a model of how to test a classic finding at scale.
- DellaVigna (2009): a survey of field evidence in psychology and economics, with an extensive section on reference dependence.

Chapter 25

Emotion and Well-Being

Why do customers rate a long service encounter favourably if it ends well, yet forget an otherwise excellent stay because of a rude check-out? Why does the thrill of a new car fade within months, while the memory of a family trip keeps paying dividends for years? Why are some online articles shared thousands of times while equally useful ones vanish, and why do angry investors take risks that fearful investors avoid? Each question concerns the role of **emotion**, the short-lived, object-directed feelings that colour judgment and behaviour, and **well-being**, the longer-run evaluation of how one's life is going.

This chapter introduces the theories that explain how emotions arise, what they do, and how they shape and are shaped by economic life. We begin with **broaden-and-build theory**, which proposes that positive emotions have a distinctive function. We then turn to **hedonic adaptation**, the tendency of emotional reactions to fade, and to the **affect circumplex**, which organizes feelings by valence and arousal. Arousal turns out to be central to **emotion and virality**. **Appraisal theories** explain why emotions of the same valence, such as fear and anger, have opposite effects on risk taking. We close with the research on **money and happiness** and the **peak-end rule**, which links emotion to memory and to the design of customer experiences (Table 25.1).

Table 25.1. Chapter at a glance: theories of emotion and well-being.

Theory	Core question	Level of analysis	Key source
Broaden-and-build theory	What are positive emotions for?	Individual / team	Fredrickson (1998, 2001)
Hedonic adaptation	Why do emotional reactions to circumstances fade?	Individual	Brickman & Campbell (1971), Frederick & Loewenstein (1999)
Affect circumplex	How are feelings structured?	Individual	Russell (1980)
Emotion and virality	Which emotions make content spread?	Individual / network	Berger (2011), Berger & Milkman (2012)
Appraisal theory	Why do specific emotions arise, and how do they shape judgment?	Individual	Smith & Ellsworth (1985), Lerner & Keltner (2001)
Money and happiness	How do income and spending relate to well-being?	Individual / society	Dunn et al. (2008), Killingsworth et al. (2023)
Peak-end rule	How are experiences remembered and evaluated?	Individual	Kahneman et al. (1993)

Two distinctions recur throughout the chapter. First, **affect** is the broad term for valenced feel-

ing states. **Moods** are diffuse and lack a clear object, whereas **emotions** are brief and directed at something. Second, **subjective well-being** has an *affective* component (the balance of positive and negative feelings in daily life) and a *cognitive* component (life satisfaction, an overall evaluation) (Diener et al. 1999). Findings about one component do not automatically transfer to the other, a point that becomes important when we examine income and happiness.

25.1 Broaden-and-Build Theory

THEORY CARD

Core claim. Positive emotions *broaden* people's momentary thought-action repertoires, and over time this broadening *builds* durable personal resources (physical, intellectual, social, and psychological) that outlast the emotion itself.
Origins. Positive and emotion psychology, Fredrickson (1998), formalized in Fredrickson (2001).
Level of analysis. Individual, extended to dyads and teams.
Key constructs. Positive emotions (joy, interest, contentment, love, amusement), broadened attention and cognition, durable resources, undoing effect, resilience, upward spirals.
Boundary conditions. Broadening is most consistent for low-approach positive states (amusement, contentment). High-approach positive states such as desire can narrow attention (Gable and Harmon-Jones 2008).
Related theories. Feelings-as-information (Section 23.7), Appraisal theory, Self-determination theory (Section 19.6), Hedonic adaptation.

25.1.1 Origins and intellectual history

Twentieth-century emotion research focused largely on negative emotions, whose functions seemed clear. Fear prepares a person to flee and anger to attack. Such emotions are associated with **specific action tendencies** that narrow attention and mobilize the body for a particular response. Positive emotions fit this model poorly. Joy and contentment do not prompt any obvious specific action, and their physiological signatures are weak. Fredrickson (1998) proposed that positive emotions have a different kind of function. Rather than narrowing the range of actions, they *broaden* it, prompting people to play, explore, savour, and integrate. Because these activities build skills, knowledge, and relationships, positive emotions have long-term adaptive value even though their immediate effects are diffuse.

25.1.2 Core logic and propositions

Fredrickson (2001) states the theory in terms of a set of hypotheses:

- **P1 (Broaden).** Positive emotions broaden the scope of attention and cognition and widen the array of thoughts and actions that come to mind.
- **P2 (Build).** Through repeated broadening, positive emotions build enduring personal resources: physical (health), intellectual (knowledge, problem-solving skills), social (relationships, support), and psychological (resilience, optimism, sense of purpose).
- **P3 (Undo).** Positive emotions speed recovery from the cardiovascular after-effects of negative emotions (the **undoing hypothesis**) (Fredrickson et al. 2000).

- **P4 (Resilience).** Resilient people use positive emotions to recover from negative experiences (Tugade and Fredrickson 2004).
- **P5 (Upward spirals).** Resources built through positive emotion increase the likelihood of future positive emotion, creating upward spirals of well-being (Figure 25.1).

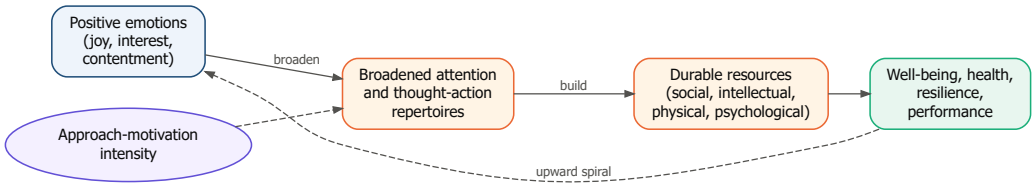


Figure 25.1. The broaden-and-build mechanism (after Fredrickson, 2001). Positive emotions broaden momentary thought-action repertoires. Repeated broadening builds durable resources, which in turn increase future positive emotion (an upward spiral). Approach motivation moderates the broadening effect.

25.1.3 Evidence and extensions

The *broaden* hypothesis has been tested by inducing emotions and measuring the scope of attention and thought. In an early demonstration, people placed in a mild positive mood (for example, by watching a comedy film or receiving a small gift) were more likely to solve problems that require creative, remote associations (Isen, Daubman, and Nowicki 1987). Fredrickson and Branigan (2005) showed participants film clips that elicited amusement, contentment, neutrality, anger, or anxiety. Positive emotions broadened the scope of visual attention, leading participants to match figures on global rather than local features, and increased the number of things participants reported they would like to do in that moment.

Evidence for the *build* hypothesis is necessarily longitudinal. In a field experiment with working adults, Fredrickson et al. (2008) randomly assigned some participants to learn loving-kindness meditation. Meditation increased daily positive emotions, which over several weeks were associated with gains in personal resources such as mindfulness, purpose in life, and social support. These resources in turn predicted greater life satisfaction and fewer depressive symptoms. On the *undoing* hypothesis, participants who watched films eliciting contentment or amusement after an anxiety-inducing task showed faster cardiovascular recovery than those who watched neutral or sad films (Fredrickson et al. 2000).

DEBATE AND EVIDENCE

The rise and fall of the "positivity ratio." In a widely publicized article, Fredrickson and Losada (2005) claimed that individuals and teams flourish when their ratio of positive to negative emotions exceeds a critical value of about 2.9, derived from a nonlinear dynamic model borrowed from fluid dynamics. N. J. L. Brown, Sokal, and Friedman (2013) showed that the mathematical modelling was unfounded: the differential equations were applied without justification, and the claimed critical ratio had no valid derivation. The journal subsequently published a correction withdrawing the modelling element and the specific ratio. In a response, Fredrickson (2013) acknowledged the problems with the mathematics while maintaining that the broader claim (that higher ratios of positive to negative emotion are associated with better outcomes, within bounds) remains supported by other evidence. The episode is a cautionary tale about precise-sounding quantitative claims in applied psychology, and it does *not* invali-

date the experimental evidence for broadening.

A second, more theoretical debate concerns **motivational intensity**. Gable and Harmon-Jones (2008) found that approach-motivated positive affect, such as the desire evoked by pictures of appetizing desserts, *narrowed* rather than broadened attention. They argue that the scope of attention depends on motivational intensity rather than valence: low-intensity positive states (contentment, amusement) broaden, while high-intensity positive states (desire, enthusiasm for a goal) focus attention on the goal.

25.1.4 Applications in business

Organizational research had documented links between positive emotion and work outcomes even before broaden-and-build theory was formalized. Employees who expressed more positive emotion at work received more favourable supervisor evaluations, higher pay, and more social support some time later (Staw, Sutton, and Pelled 1994). A meta-analysis of cross-sectional, longitudinal, and experimental studies concluded that frequent positive affect is not only a consequence of success but also a cause of it, in work, relationships, and health (Lyubomirsky, King, and Diener 2005). Laboratory experiments that induced happiness with a short comedy clip found higher productivity on a paid task (Oswald, Proto, and Sgroi 2015).

BUSINESS CASE

Positive leadership and employee well-being. Emotions spread within teams. In a laboratory study of groups working on a managerial exercise, a trained confederate who displayed pleasant emotion led group members to experience more positive mood, which improved cooperation and reduced conflict (Barsade 2002). Leaders' moods have a similar effect: groups whose leaders were in a positive mood showed more positive group affective tone and better coordination, while groups whose leaders were in a negative mood exerted more effort (Sy, Côté, and Saavedra 2005). At the business-unit level, a meta-analysis of Gallup data found that employee satisfaction and engagement were related to customer satisfaction, productivity, profitability, and lower turnover (Harter, Schmidt, and Hayes 2002). Investors appear to underweight these benefits: a portfolio of firms listed among the "100 Best Companies to Work For in America" earned abnormal long-run returns, suggesting that the market did not fully value employee satisfaction (Edmans 2011). Broaden-and-build theory supplies a mechanism: positive emotions broaden cognition and build social resources, both of which matter in knowledge work. The theory also implies a caution: "mandated positivity" that suppresses legitimate negative emotion (for example, about safety or ethics) may undermine the problem detection that negative emotions support.

25.1.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Team members' daily positive affect will predict the number of novel ideas they contribute in subsequent meetings, mediated by breadth of attention.
- *H2.* The relationship between positive affect and creativity will be weaker for high-approach positive states (e.g., excitement about a bonus) than for low-approach states

(e.g., contentment).
- *H3.* A workplace intervention that increases daily positive emotion will increase perceived social support six months later (build hypothesis).

Designs. Emotion inductions with global-local attention tasks. Experience-sampling studies with lagged models. Randomized field interventions (e.g., gratitude or meditation programmes) with longitudinal follow-up.
Measures. Positive and Negative Affect Schedule [PANAS, Watson, Clark, and Tellegen (1988)], daily emotion checklists, global-local processing tasks, resource measures (social support, resilience), performance and turnover data.

COMMON PITFALL

Treating positivity as a target ratio. After the positivity-ratio controversy, researchers and practitioners should avoid prescribing a specific balance of positive to negative emotion. The theory concerns the *functions* of positive emotions, not a numerical threshold for flourishing.

25.2 Hedonic Adaptation

THEORY CARD

Core claim. Emotional reactions to stable or repeated circumstances diminish over time, so changes in circumstances, whether good or bad, have smaller and shorter-lived effects on well-being than people expect.
Origins. Adaptation-level theory in perception, applied to happiness by Brickman and Campbell (1971) ("hedonic relativism", later the "hedonic treadmill"), classic evidence in Brickman, Coates, and Janoff-Bulman (1978), reviewed by Frederick and Loewenstein (1999), revised by Diener, Lucas, and Scollon (2006) and R. E. Lucas (2007).
Level of analysis. Individual.
Key constructs. Adaptation level, set point, hedonic treadmill, desensitization, affective forecasting, incomplete adaptation.
Boundary conditions. Adaptation is faster for material circumstances and slower or incomplete for some events (e.g., disability, unemployment, widowhood). It varies across people and across components of well-being (R. E. Lucas 2007, Luhmann et al. 2012).
Related theories. Broaden-and-build theory, Prospect theory (Chapter 24), Peak-end rule, Social comparison theory (Section 22.1).

25.2.1 Origins and intellectual history

Adaptation-level theory in perception holds that judgments of stimuli (how heavy, how bright, how warm) depend on their relation to a level set by recent experience. Brickman and Campbell (1971) applied this idea to happiness. If people evaluate their circumstances relative to an adaptation level that rises with improvement, then gains in income or comfort produce only temporary increases in happiness. They called this predicament **hedonic relativism**. The image of a **hedonic treadmill**, on which people must keep running simply to stay in place, became its popular name.

The most famous early evidence came from Brickman, Coates, and Janoff-Bulman (1978), who compared lottery winners, people paralysed in accidents, and controls. Lottery winners were not reli-

ably happier than controls and reported taking less pleasure in ordinary activities, consistent with a contrast effect against the peak experience of winning. Accident victims reported lower current happiness than controls, but not as low as observers would expect. Although the samples were small, the study became the emblem of adaptation.

Frederick and Loewenstein (1999) defined **hedonic adaptation** as any process that reduces the affective intensity of a constant or repeated stimulus, and distinguished two forms: **shifting adaptation levels**, in which the neutral reference point moves toward the current stimulus, and **desensitization**, in which sensitivity to changes declines. The distinction matters because the former implies that people return to neutral, whereas the latter implies that they merely feel changes less intensely.

25.2.2 From the treadmill to set-point revisions

The strong form of the treadmill (that everyone returns to hedonic neutrality after any event) has not survived empirical scrutiny. Drawing on large panel studies, Diener, Lucas, and Scollon (2006) proposed five revisions, summarized in Table 25.2.

Table 25.2. Five revisions to the hedonic treadmill model proposed by Diener, Lucas, and Scollon (2006).

Revision	Original treadmill assumption	Revised view	Implication
1. Set points are not neutral	People adapt back to hedonic neutrality	Most people's baseline is mildly positive	Average well-being is above the scale midpoint
2. Set points differ across people	Everyone has the same set point	Set points vary, partly because of temperament	Personality predicts baseline happiness
3. Multiple set points	Happiness is a single quantity	Components (positive affect, negative affect, life satisfaction) can move in different directions	Measure components separately
4. Set points can change	Long-run happiness cannot be changed	Well-being can shift durably under some conditions	Policy and interventions can matter
5. Individual differences in adaptation	Everyone adapts to the same degree	People differ in whether and how fast they adapt	Average trajectories mask heterogeneity

Panel data from Germany and the United Kingdom were especially influential. R. E. Lucas et al. (2003) found that, on average, people received a boost in life satisfaction around marriage and then returned toward baseline, while widowhood produced a sharp drop followed by slow and, for many, incomplete recovery. R. E. Lucas (2007) summarized evidence that some events, including disability, unemployment, and divorce, are associated with lasting changes in life satisfaction, and that individual differences in reaction and adaptation are large. A meta-analysis by Luhmann et al. (2012) found that the course of adaptation differs across life events and across affective and cognitive components of well-being. Figure 25.2 illustrates these patterns with stylised data.

25.2.3 Can happiness be increased? The "happiness pie" and its critique

If people adapt to circumstances, can deliberate effort make anyone happier in the long run? Lyubomirsky, Sheldon, and Schkade (2005) proposed an influential "architecture of sustainable change", often summarized as a pie chart: roughly 50 percent of the differences in chronic happiness reflect a genetically determined set point, about 10 percent reflect life circumstances (income, health, marital status), and about 40 percent reflect intentional activities, the things people choose

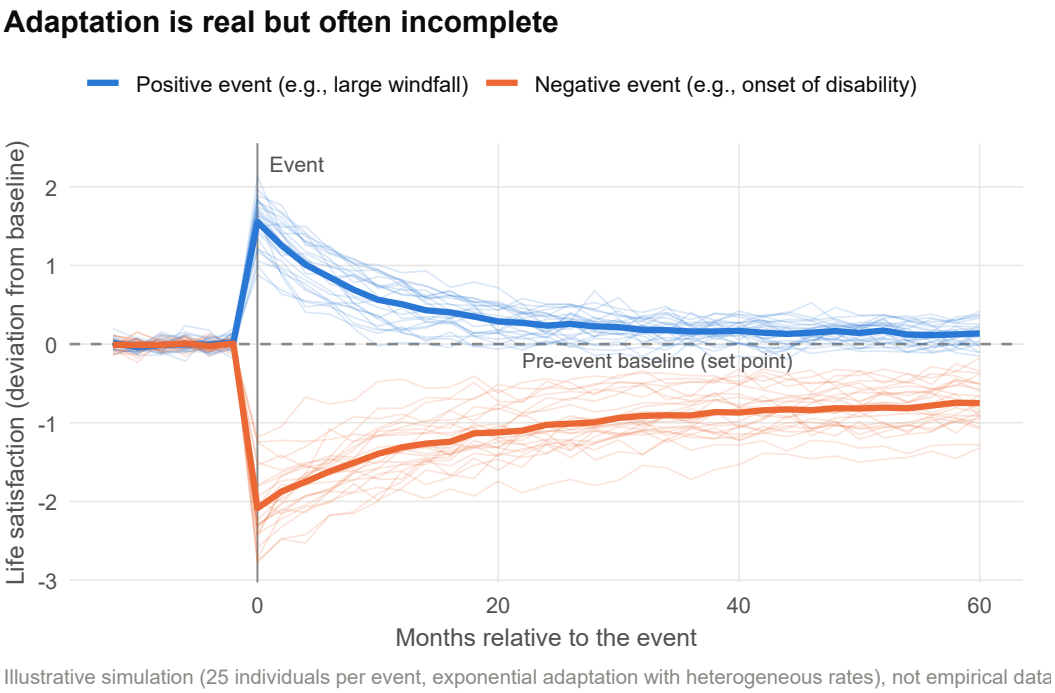


Figure 25.2. Illustrative adaptation trajectories after a positive and a negative life event. Thin lines show simulated individuals with heterogeneous reactions and adaptation rates. Thick lines show the average. Adaptation to the positive event is nearly complete, whereas adaptation to the negative event is partial, consistent with the revised set-point view (Diener et al., 2006, Lucas, 2007).

to think and do. Because intentional activities can be varied, they argued, they are less susceptible to adaptation than circumstances and offer the most promising route to sustainable increases in happiness.

DEBATE AND EVIDENCE

Is the 40 percent figure meaningful? N. J. L. Brown and Rohrer (2020) offer a detailed critique of the pie-chart model. First, the 50 percent figure is a heritability estimate, which describes the share of *variance in a population* associated with genetic differences. It does not imply that half of any individual's happiness is fixed. Second, the 10 percent figure for circumstances was derived from correlations between selected life circumstances and well-being and is likely to underestimate their influence. Third, the 40 percent for intentional activities was not measured. It is a residual obtained by subtracting the other two figures, so it also absorbs measurement error and everything else not accounted for. Finally, variance components of this kind are not additive when genes and environments interact. The critique does not show that intentional activities are ineffective. It shows that the pie chart cannot quantify how much they matter. Experimental evidence on specific interventions is the appropriate test.

25.2.4 Affective forecasting and adaptation

People not only adapt. They also fail to anticipate that they will adapt. In studies of **affective forecasting**, people overestimate how long they will be unhappy after negative events such as a roman-

tic breakup or a failed job application, a **durability bias** that arises partly because they neglect the psychological processes, such as rationalization, that speed recovery (**immune neglect**) (Gilbert et al. 1998, T. D. Wilson and Gilbert 2005). Forecasting errors matter for decisions: people may overpay to avoid outcomes they would have adapted to and underinvest in experiences whose benefits persist.

The **hedonic adaptation prevention** model proposes two levers that slow adaptation to positive changes: **variety** in the experiences a change brings and **appreciation** of the change, which counteracts rising aspirations (Sheldon and Lyubomirsky 2012).

25.2.5 Applications in business

BUSINESS CASE

Hedonic adaptation in subscriptions and retention. Subscription businesses face hedonic adaptation directly: the novelty that drives sign-up fades, and customers reassess whether the service is worth the price. Research suggests two counterintuitive levers. First, **interruptions** can restore enjoyment. Breaking up a pleasant experience such as a massage made it more enjoyable, whereas breaking up an unpleasant one made it more irritating (L. D. Nelson and Meyvis 2008). Even television programmes were enjoyed more when interrupted by commercials (L. D. Nelson, Meyvis, and Galak 2009). Second, the hedonic adaptation prevention model implies that **variety** and **appreciation** sustain satisfaction (Sheldon and Lyubomirsky 2012). Streaming services that rotate their catalogues, meal-kit services that change menus weekly, and loyalty programmes that remind members of accumulated benefits all work against adaptation. **Illustrative scenario:** a fitness app whose engagement declines after the first two months might test (a) rotating programme "seasons" to add variety, (b) a monthly summary of progress to cultivate appreciation, and (c) planned breaks, measuring retention against a control group.

COMMON PITFALL

Inferring adaptation from cross-sectional data. Comparing people who experienced an event years ago with those who did not cannot establish adaptation, because people who experience events differ systematically from those who do not. Adaptation is a within-person process and should be studied with panel data that include pre-event baselines, or with experiments.

25.3 The Affect Circumplex

THEORY CARD

Core claim. Affective states are not independent categories but points in a two-dimensional space defined by **valence** (pleasure-displeasure) and **arousal** (activation-deactivation). Emotion terms fall in a circular order around this space.
Origins. Psychology of emotion, Russell (1980), developed into the theory of **core affect** in Russell (2003).
Level of analysis. Individual feeling states.

Key constructs. Valence, arousal, core affect, circumplex ordering, positive and negative activation.
Boundary conditions. Captures the felt quality of affect well but not the appraisals and action tendencies that distinguish emotions with similar valence and arousal, such as fear and anger.
Related theories. Emotion and virality, Appraisal theory, Feelings-as-information (Section 23.7).

25.3.1 Core logic

Russell (1980) asked people to sort and rate 28 affect words and analysed the results with several scaling methods. The words consistently fell in a circle around two bipolar dimensions: pleasure versus displeasure (valence) and arousal versus sleepiness (arousal). *Excited* and *delighted* lie in the pleasant, high-arousal quadrant. *Content* and *relaxed* in the pleasant, low-arousal quadrant. *Tense* and *angry* in the unpleasant, high-arousal quadrant, and *sad* and *bored* in the unpleasant, low-arousal quadrant (Figure 25.3).

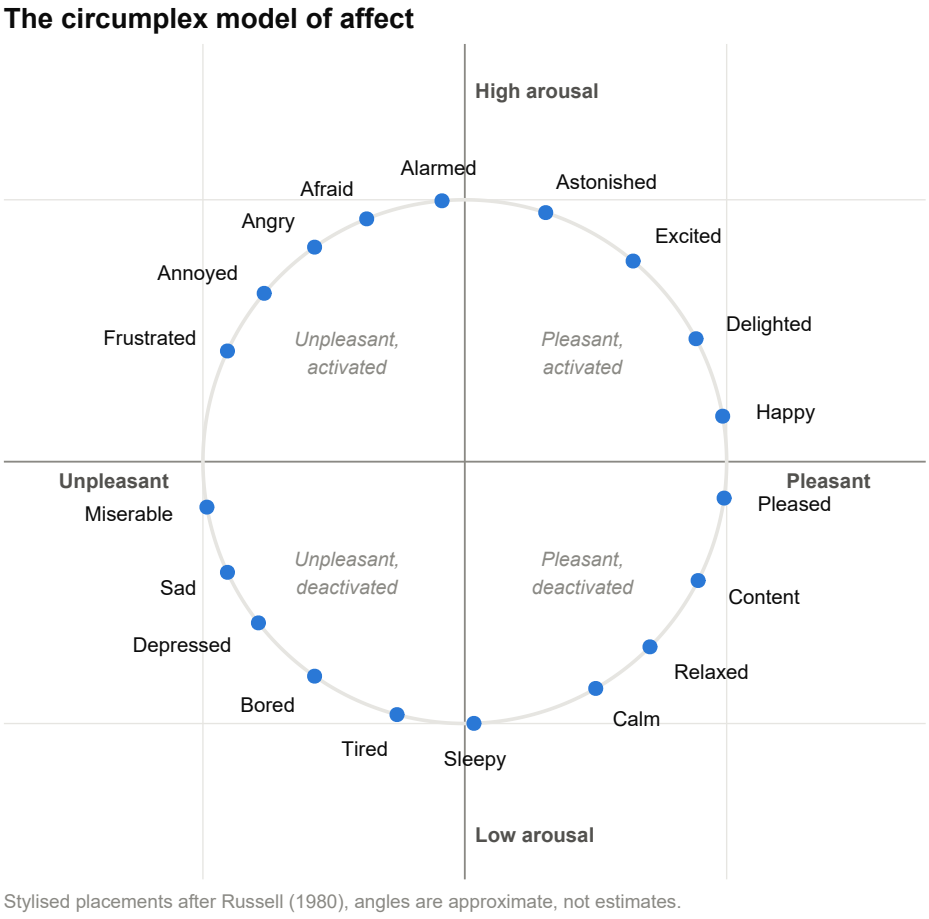


Figure 25.3. The affect circumplex. Emotion terms arranged by valence (horizontal) and arousal (vertical). Positions are approximate and stylised, following the ordering reported by Russell (1980).

Russell (2003) later developed the circumplex into a broader theory. **Core affect** is a continuously present neurophysiological state, experienced as a blend of pleasure and arousal, that is not neces-

sarily directed at anything. Full-blown emotional episodes (an instance of fear, say) are constructed when core affect is attributed to an object and combined with appraisals, behaviour, and categorization. A related tradition represents affect with two unipolar dimensions, **positive affect** and **negative affect** (Watson and Tellegen 1985). Barrett and Russell (1999) argued that the two representations are rotations of the same space: positive and negative activation lie at roughly 45 degrees to the valence and arousal axes.

DEBATE AND EVIDENCE

Dimensions or discrete emotions? Basic emotion theorists hold that a small set of emotions (such as fear, anger, disgust, sadness, happiness, and surprise) are discrete biological programs with distinctive signals and physiology (Ekman 1992). Dimensional theorists counter that such categories are constructed from more basic ingredients and that physiological differences between categories are less consistent than the basic-emotion view implies (Russell 2003). For applied researchers, the choice has practical consequences. Dimensional measures (valence, arousal) are parsimonious and work well for predicting outcomes driven by activation, such as sharing. Discrete measures are needed when emotions of similar valence and arousal have different effects, as fear and anger do on risk taking.

RESEARCH TOOLKIT

Measuring affect in business research. Self-report scales such as the PANAS (Watson, Clark, and Tellegen 1988) and valence-arousal grids. Experience sampling and the Day Reconstruction Method for affect in daily life (Kahneman et al. 2004). Automated facial-expression coding of video (Teixeira, Wedel, and Pieters 2012). Text-based sentiment and emotion dictionaries for reviews and social media, and physiological indicators (skin conductance for arousal). Validate any automated measure against self-report in the target context before relying on it.

25.4 Emotion and Virality

THEORY CARD

Core claim. Content that evokes high-arousal emotions, whether positive (awe, amusement, excitement) or negative (anger, anxiety), is more likely to be shared than content that evokes low-arousal emotions (contentment, sadness), because physiological arousal mobilizes action.
Origins. Consumer psychology and marketing, Berger (2011), Berger and Milkman (2012), popularized in Berger (2013).
Level of analysis. Individual sharing decisions aggregated into diffusion through networks.
Key constructs. Arousal (activation), valence, social transmission, virality, practical value.
Boundary conditions. Arousal increases transmission of whatever content is at hand, including content unrelated to its source. Positive content is, other things equal, more viral than negative content.
Related theories. Affect circumplex, Social transmission (Section 26.3), Identity signaling (Section 8.5).

25.4.1 Core logic and evidence

Why should emotion matter for sharing? A natural hypothesis is valence: people share good news to make themselves look good or to make others feel good. Berger (2011) proposed that **arousal**, a state of mobilization and readiness for action, is the more fundamental driver. In one experiment, participants who watched film clips evoking high-arousal emotions (amusement or anxiety) were more willing to share an article with others than those who watched clips evoking low-arousal emotions (contentment or sadness). In a second experiment, participants who jogged in place for a minute, a purely physiological manipulation of arousal, were more likely to share a neutral article than those who sat still.

Berger and Milkman (2012) tested the idea in the field. They analysed thousands of *New York Times* articles and asked which made the paper's most-emailed list, controlling for placement, timing, and author. Positive content was more viral than negative content, but the effect of specific emotions depended on arousal. Articles that evoked awe were more likely to be shared, as were articles evoking anger or anxiety, despite their negative valence. Articles evoking sadness, a low-arousal emotion, were *less* likely to be shared. Table 25.3, adapted from Berger (2013, 109), organizes emotions by valence and arousal and shows the predicted effect on sharing.

Table 25.3. Emotions by valence and arousal and their predicted effect on sharing (adapted from Berger, 2013, p. 109. Evidence from Berger, 2011, and Berger & Milkman, 2012).

Valence	High arousal (activating)	Low arousal (deactivating)
Positive	Awe, excitement, amusement (humour)	Contentment
Negative	Anger, anxiety	Sadness
Predicted effect on sharing	Increases sharing	Reduces sharing or has little effect

Figure 25.4 shows the mechanism of emotion-driven virality.

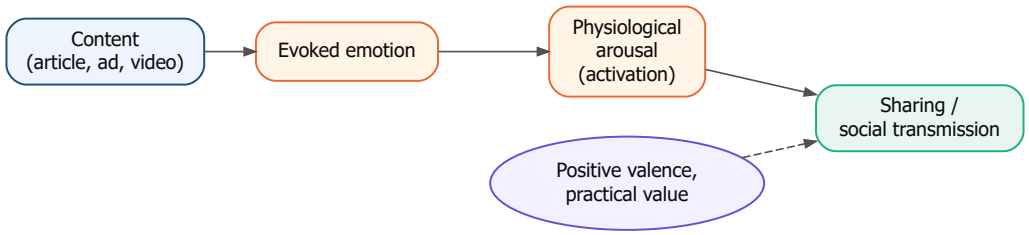


Figure 25.4. Mechanism of emotion-driven virality. Content evokes emotion. High-arousal emotions activate the reader, increasing the likelihood of sharing. Positive valence and practical value add to sharing independently of arousal.

25.4.2 Extensions

Emotion also shapes attention to advertising. Using automated facial-expression tracking, Teixeira, Wedel, and Pieters (2012) found that joy and surprise evoked by internet video advertisements increased viewers' attention and reduced the likelihood that they would stop watching. Across more than 1,000 television commercials, Pham, Geuens, and De Pelsmacker (2013) found that the feelings an ad evoked were systematically related to viewers' evaluations of the advertised brand.

BUSINESS CASE

Emotional content marketing. The arousal account offers a simple diagnostic for content strategy: will this piece *activate* its audience? Content that merely makes people feel content or sad may be liked but not shared. Content that evokes awe, amusement, excitement, or (with care) indignation is more likely to travel (Berger and Milkman 2012, Berger 2013). Three cautions apply. First, arousal should be connected to the brand. Content that goes viral without being linked to the product builds awareness of the content, not the brand. Second, negative high-arousal emotions such as anger spread readily, but associating a brand with outrage carries reputational risk. Third, practical value is an independent driver of sharing, so useful content can succeed without strong emotion.

DEBATE AND EVIDENCE

The ethics of engineering emotion. In an experiment involving roughly 690,000 Facebook users, Kramer, Guillory, and Hancock (2014) reduced the amount of positive or negative content in users' News Feeds and found small but statistically significant shifts in the emotional tone of users' own posts. The study demonstrated emotional contagion at scale but provoked controversy over informed consent in platform experiments. Combined with evidence that high-arousal negative emotions such as anger spread readily, it raises a broader concern: if engagement metrics reward arousal, platforms and marketers may be tempted to optimize content for outrage and anxiety. Researchers and managers should weigh short-run engagement against well-being and trust.

COMMON PITFALL

Equating "emotional" with "viral". The key variable is arousal, not emotional intensity in general and not valence alone. A heart-wrenching but saddening campaign may be memorable and still not be shared. Measure arousal directly or, at least, classify emotions by their position on the circumplex.

25.5 Appraisal Theory

THEORY CARD

Core claim. Emotions are elicited by people's evaluations (**appraisals**) of what events mean for their well-being. Specific emotions correspond to specific patterns of appraisal, and each emotion in turn carries an **appraisal tendency** that colours subsequent judgments, even of unrelated targets.

Origins. Emotion psychology, C. A. Smith and Ellsworth (1985), Lazarus (1991), reviewed in Moors et al. (2013), applied to judgment and decision making by Lerner and Keltner (2000) and Lerner and Keltner (2001).

Level of analysis. Individual.

Key constructs. Appraisal dimensions (pleasantness, certainty, control, responsibility, anticipated effort, attention), core relational themes, appraisal tendencies, incidental vs. integral emotion.

Boundary conditions. Appraisal tendencies carry over to unrelated judgments mainly when people are unaware of the emotion's source and when the judgment is related to the emotion's core appraisals.
Related theories. Affect circumplex, Feelings-as-information (Section 23.7), Prospect theory (Chapter 24), Endowment effect (Section 23.4).

25.5.1 Origins and core logic

The affect circumplex places fear and anger close together: both are unpleasant and high in arousal. Yet they feel different and lead to different behaviour. Appraisal theories explain why. C. A. Smith and Ellsworth (1985) asked people to recall experiences of various emotions and rate them on many candidate appraisal dimensions. Six dimensions, namely pleasantness, anticipated effort, certainty, attentional activity, responsibility (self vs. other), and situational control, were sufficient to distinguish fifteen emotions. Anger, for example, involves the appraisal that someone else is responsible for an unpleasant event and a sense of certainty about what happened. Fear involves uncertainty and the sense that the situation, rather than any person, is in control.

Lazarus (1991) integrated appraisal into a broader cognitive-motivational- relational theory. Each emotion has a **core relational theme**, a summary of the person-environment relationship that produces it. Anger stems from a demeaning offence against oneself or one's own. Anxiety from facing an uncertain, existential threat. Sadness from an irrevocable loss. Happiness from making reasonable progress toward a goal. Table 25.4 summarizes appraisal patterns for several emotions that matter in business settings.

Table 25.4. Appraisal patterns and decision effects of selected emotions (synthesized from Smith & Ellsworth, 1985, Lazarus, 1991, Lerner & Keltner, 2000, 2001, Lerner et al., 2004).

Emotion	Core relational theme (paraphrased from Lazarus, 1991)	Key appraisals	Typical influence on judgment and choice
Anger	A demeaning offence against me or mine	Other-responsible, high certainty, high individual control	Optimistic risk estimates, risk seeking, heuristic processing
Fear / anxiety	Facing uncertain threat	Low certainty, situational control	Pessimistic risk estimates, risk aversion
Sadness	Experiencing an irrevocable loss	Loss, situational control	Desire to change circumstances, higher buying and lower selling prices
Disgust	Being too close to something offensive	Contamination, urge to expel	Desire to get rid of things, lower buying and selling prices
Happiness	Making reasonable progress toward a goal	Pleasant, high certainty, low effort	Optimistic risk estimates, heuristic processing
Pride	Enhancement of identity by taking credit for an achievement	Self-responsible, pleasant	Persistence, self-reward

25.5.2 The appraisal-tendency framework

Lerner and Keltner (2000) proposed that each emotion activates a cognitive predisposition to appraise future events in line with the appraisals that produced the emotion. Because fear is associated with uncertainty and lack of control, fear should increase perceived risk. Because anger is

associated with certainty and individual control, anger should reduce it. Lerner and Keltner (2001) confirmed these predictions: dispositionally fearful people made pessimistic risk estimates and risk-averse choices, whereas angry people made optimistic risk estimates and risk-seeking choices, as did happy people. In a national field experiment conducted shortly after the September 11, 2001 attacks, participants induced to feel anger estimated lower risks of future terrorism than participants induced to feel fear (Lerner et al. 2003). Certainty appraisals also shape processing style: emotions associated with certainty (anger, happiness) promote heuristic processing, whereas emotions associated with uncertainty promote systematic processing (Tiedens and Linton 2001).

Appraisal tendencies extend to economic valuation. Lerner, Small, and Loewenstein (2004) found that incidental disgust reduced both buying and selling prices, eliminating the endowment effect, whereas incidental sadness *reversed* it, lowering selling prices and raising buying prices, consistent with sadness motivating a change of circumstances. The finding is a striking example of how emotion moderates a phenomenon usually explained by loss aversion (Section 24.4) (see Endowment effect (Section 23.4)). Figure 25.5 summarizes the framework. Lerner et al. (2015) review the broader literature on emotion and decision making, and S. Han, Lerner, and Keltner (2007) apply it to consumer behaviour.

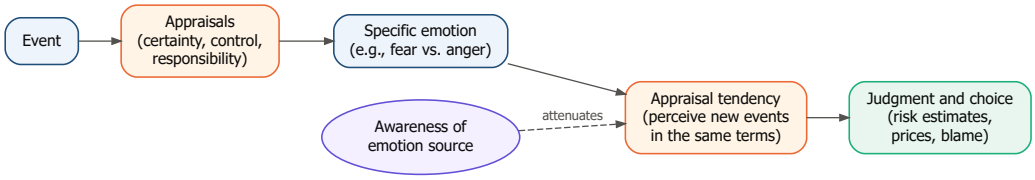


Figure 25.5. The appraisal-tendency framework (after Lerner & Keltner, 2000). Appraisals of an event produce a specific emotion. The emotion carries an appraisal tendency that shapes judgments and choices, including judgments unrelated to the emotion's source.

25.5.3 Applications in business

BUSINESS CASE

Service recovery and crisis communication. Appraisal theory implies that the *same* service failure can produce different emotions, and therefore different behaviour, depending on how customers appraise responsibility and control. A delayed flight attributed to the airline's poor planning (other-responsible, controllable) is likely to elicit anger, which motivates complaints, blame, and switching. A delay attributed to weather (situational control) is more likely to elicit frustration or anxiety, which motivates information seeking. Communications that accurately clarify cause and control, and that restore a sense of control (rebooking options, real-time updates), address the appraisals that drive the emotion. In crisis communication, the appraisal-tendency framework suggests that messages that evoke anger at a culprit may reduce perceived risk among the public, whereas messages that evoke fear increase it, a consideration for public-health and safety communicators (Lerner et al. 2003).

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Customers who attribute a service failure to the firm (vs. to external circumstances)

will report more anger and be more likely to post negative reviews, and anger will mediate the effect.
- *H2.* Incidental fear (vs. anger) induced before a pricing decision will increase willingness to pay for insurance-like product add-ons.
- *H3.* The effect of incidental emotion on investment risk taking will disappear when participants are made aware of the emotion's source.

Designs. Emotion inductions (recall tasks, film clips) followed by ostensibly unrelated judgments. Appraisal manipulations (e.g., attribution of responsibility). Field data on emotional events (e.g., sports results, news shocks).
Measures. Discrete emotion self-reports, appraisal ratings, risk estimates, willingness to pay and accept, complaint and switching behaviour.

25.6 Money and Happiness

Does money buy happiness? The answer depends on which component of well-being is measured and on how money is spent. Table 25.5 summarizes the key findings.

Table 25.5. Selected findings on money and happiness.

Question	Key study	Finding	Business implication
Does how you spend matter?	Dunn, Aknin & Norton (2008)	Spending on others predicted and (experimentally) caused greater happiness than spending on oneself	Gift, donation, and team-reward programmes
Is prosocial spending universal?	Aknin et al. (2013)	Association between prosocial spending and well-being observed across many countries	Prosocial rewards travel across cultures
Experiences or things?	Van Boven & Gilovich (2003)	Experiential purchases tend to make people happier than material purchases	Experiential offerings and bundles
Why experiences?	Carter & Gilovich (2010), Kumar et al. (2014)	Experiences are less subject to comparison and more pleasurable to anticipate	Market anticipation, reduce comparison shopping
Time or things?	Whillans et al. (2017)	Spending to buy time is associated with greater life satisfaction	Time-saving services as well-being products
Does income raise emotional well-being?	Kahneman & Deaton (2010)	Life evaluation rose with log income. Emotional well-being levelled off at higher incomes	Distinguish evaluation from experience
A plateau?	Killingsworth (2021)	Experienced well-being rose with log income with no plateau	Measurement choices matter
Resolution	Killingsworth, Kahneman & Mellers (2023)	A plateau appears only among the least happy minority. For most, happiness keeps rising with income	Heterogeneity across the happiness distribution

25.6.1 How money is spent

Dunn, Aknin, and Norton (2008) combined three studies. In a national survey, spending on others (gifts and donations) predicted greater happiness, whereas personal spending did not. In a field study, employees who devoted more of a profit-sharing bonus to others reported greater hap-

piness weeks later. In an experiment, participants who were randomly assigned to spend a small windfall on others by the end of the day were happier than those assigned to spend it on themselves, regardless of whether the windfall was $5 or $20. A cross-cultural follow-up found the association between prosocial spending and well-being in many countries (Aknin et al. 2013).

Van Boven and Gilovich (2003) found that **experiential purchases** (a trip, a concert, a meal) tend to make people happier than **material purchases** (clothing, jewellery, gadgets). Later work identified reasons: experiences are less easily compared with forgone alternatives (Carter and Gilovich 2010), are more pleasurable to anticipate (Kumar, Killingsworth, and Gilovich 2014), are more central to identity, and are more often shared with others (Gilovich, Kumar, and Jampol 2015). Spending money to buy time, for example on services that eliminate disliked chores, is also associated with greater well-being (Whillans et al. 2017).

25.6.2 How much money: an adversarial collaboration

Kahneman and Deaton (2010) analysed more than 450,000 responses to a daily survey of US residents and distinguished **life evaluation** (where respondents placed their life on a ladder from worst to best possible) from **emotional well-being** (reports of positive and negative emotions experienced the previous day). Life evaluation rose steadily with the logarithm of income. Emotional well-being also rose with income, but beyond an annual household income of about $75,000 there was no further progress. The $75,000 "plateau" became one of the most widely cited numbers in happiness research.

Killingsworth (2021) challenged the plateau. Using more than a million experience-sampling reports from tens of thousands of employed US adults, collected through a smartphone app, he found that experienced well-being rose linearly with log income with no sign of levelling off. The two findings appeared to conflict.

Rather than trading critiques, Kahneman and Killingsworth undertook an **adversarial collaboration** with Barbara Mellers as arbiter (Killingsworth, Kahneman, and Mellers 2023). They hypothesized that both findings could be true if the relationship between income and happiness differs across the happiness distribution. Re-analysing the experience-sampling data, they found that for the least happy minority of people (roughly the unhappiest fifth), happiness rose with income up to about $100,000 and then plateaued. For the majority, happiness continued to rise with income, and among the happiest it rose even faster above that level. They concluded that the earlier plateau reflected the fact that the 2010 measures of emotional well-being, which asked yes/no questions about the previous day, were most sensitive to differences in unhappiness. The episode is an exemplar of how scientific disagreements can be resolved constructively.

BUSINESS CASE

Experiential versus material offerings. Firms increasingly package goods as experiences: car makers offer driving events, retailers host workshops, and banks reward customers with access to concerts rather than merchandise. The research above offers a rationale and some design principles. Experiences are more satisfying when they are anticipated (pre-sell and build anticipation), shared (design for groups), and hard to compare directly with alternatives (create distinctive, non-commoditized experiences). Prosocial spending research suggests that reward programmes that let customers or employees give to others, such as donation options for loyalty points or team celebrations instead of individual merchandise, may

produce more happiness per dollar (Dunn, Aknin, and Norton 2008). These are hypotheses to test, not guarantees: most of the evidence concerns individuals' own spending rather than firm-designed programmes.

25.7 Peak-End Rule

THEORY CARD

Core claim. Retrospective evaluations of an experience are dominated by the affect at its most intense moment (the **peak**) and at its **end**, while its duration is largely neglected.
Origins. Psychology and behavioural economics, Kahneman et al. (1993), Fredrickson and Kahneman (1993), Redelmeier and Kahneman (1996), framed within the distinction between experienced and remembered utility by Kahneman, Wakker, and Sarin (1997).
Level of analysis. Individual memory and evaluation of episodes.
Key constructs. Experienced utility, remembered utility, peak, end, duration neglect.
Boundary conditions. Applies to evaluations of bounded episodes with a clear beginning and end. Duration matters more when it is made salient or is part of the evaluation criterion.
Related theories. Hedonic adaptation, Mental accounting (Section 23.5), Prospect theory (Chapter 24).

25.7.1 Core logic and evidence

Kahneman, Wakker, and Sarin (1997) distinguish **experienced utility**, the moment-by-moment hedonic quality of an experience, from **remembered utility**, the retrospective evaluation of the whole episode. Normatively, the total experienced utility of an episode should be the integral of momentary experience over time, so that adding unpleasant time can only make the episode worse. Remembered utility follows different rules.

In the experiment by Kahneman et al. (1993), participants immersed one hand in painfully cold water in two trials. In the short trial, the hand was immersed for 60 seconds. In the long trial, the hand was immersed for the same 60 seconds and then for an additional 30 seconds during which the water was warmed slightly, by about one degree Celsius, which was still unpleasant but less so. The long trial contained strictly more discomfort. Yet when asked which trial they would choose to repeat, most participants chose the long one. Their memories were shaped by the less painful ending, not by the extra duration. Studies of film clips found the same **duration neglect** in retrospective evaluations of pleasant and unpleasant episodes (Fredrickson and Kahneman 1993).

The rule matters outside the laboratory. Patients' retrospective evaluations of painful colonoscopies were predicted by peak and end pain, not by the procedure's duration (Redelmeier and Kahneman 1996). In a randomized trial, extending a colonoscopy with a short, less uncomfortable period at the end led patients to remember the procedure as less unpleasant, and they were more likely to return for follow-up procedures (Redelmeier, Katz, and Kahneman 2003). Figure 25.6 shows stylised profiles of the cold-water design.



Figure 25.6. Stylised discomfort profiles modelled on the design of Kahneman et al. (1993). The long trial adds 30 seconds of milder discomfort. Total discomfort is higher, but the peak-end average is lower, so the long trial is remembered as less unpleasant.

25.7.2 Applications in business

BUSINESS CASE

Designing customer journeys around peaks and ends. The peak-end rule implies that customers' memories of a service, which drive word of mouth and repurchase, are shaped disproportionately by its best or worst moment and its final moments. Chase and Dasu (2001) drew on this research to propose service design principles: finish strong rather than front-loading the best elements. Get unpleasant elements out of the way early. Segment pleasant experiences and combine unpleasant ones (echoing the hedonic editing principles of Mental accounting (Section 23.5)), and give customers a sense of choice and control. Practical applications include a final personal touch at hotel check-out, a smooth and friendly ending to a support call even after a long wait, and dessert or a small gift at the end of a restaurant meal. Because memories, not moment-by-moment experience, drive future choices, firms should measure both: real-time experience metrics to reduce actual discomfort, and retrospective satisfaction to capture what customers will remember.

COMMON PITFALL

Using the peak-end rule to justify a worse experience. The colonoscopy trial added a milder period that improved memory without causing additional harm. Extending an unpleasant service to "end on a high note" may improve retrospective ratings while increasing the total discomfort customers actually experience. Remembered and experienced utility can diverge, and ethical design should not sacrifice the latter to game the former (Kahneman, Wakker, and

Sarin 1997).

TRY IT YOURSELF

Map a customer journey you know well (e.g., buying a car, opening a bank account). Rate the emotional intensity of each step and identify its peak and its end. Propose one change that improves the end without lengthening the journey, and one that reduces the worst moment.

25.8 Chapter Summary

KEY TAKEAWAYS

- **Positive emotions** broaden attention and thought and, over time, build personal resources (Fredrickson 2001). The claimed "critical positivity ratio" was retracted after critique (N. J. L. Brown, Sokal, and Friedman 2013), but the experimental evidence for broadening stands.
- **Hedonic adaptation** is real but incomplete and heterogeneous. Set points are positive, differ across people, and can change (Diener, Lucas, and Scollon 2006, R. E. Lucas 2007). The 50/10/40 "happiness pie" cannot be interpreted as a decomposition of individual happiness (N. J. L. Brown and Rohrer 2020).
- The **affect circumplex** organizes feelings by valence and arousal (Russell 1980). Arousal, not valence alone, predicts sharing (Berger 2011, Berger and Milkman 2012).
- **Appraisal theory** explains why emotions of similar valence and arousal, such as fear and anger, have opposite effects on risk perception (C. A. Smith and Ellsworth 1985, Lerner and Keltner 2001).
- How money is spent matters: prosocial and experiential spending are associated with greater happiness (Dunn, Aknin, and Norton 2008, Van Boven and Gilovich 2003). Higher income is associated with greater happiness for most people, with a plateau only among the least happy (Killingsworth, Kahneman, and Mellers 2023).
- The **peak-end rule** and duration neglect mean that remembered experience diverges from experienced utility (Kahneman et al. 1993, Kahneman, Wakker, and Sarin 1997), with direct implications for customer-journey design.

Table 25.6 presents the comparison of the chapter's theories.

Table 25.6. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Typical business use
Broaden-and-build	Positive emotions broaden cognition and build resources	Positive affect increases creativity and resources over time	Positive leadership, employee well-being programmes
Hedonic adaptation	Shifting adaptation level, desensitization	Effects of new circumstances fade	Retention design, variety and interruption
Affect circumplex	Valence and arousal structure feelings	Emotions arranged in a circle	Measuring emotional response to products and ads
Emotion and virality	Arousal mobilizes action	High-arousal content is shared more	Content marketing, campaign design

Table 25.6. Comparison of the chapter's theories. *(continued)*

Theory	Key mechanism	Signature prediction	Typical business use
Appraisal theory	Appraisals elicit emotions with appraisal tendencies	Fear raises and anger lowers perceived risk	Service recovery, crisis communication
Money and happiness	Spending patterns and income gradients	Prosocial and experiential spending raise happiness	Reward design, experiential offerings
Peak-end rule	Memory summarizes peak and end, duration neglected	Adding a milder ending improves evaluations	Customer-journey and service design

25.9 Discussion Questions

1. Broaden-and-build theory predicts that positive emotions build resources over time. Design a field study in an organization that could test the *build* hypothesis rather than only the *broaden* hypothesis.
2. The positivity-ratio episode involved a published claim that was later corrected. What features of the original claim should have prompted scepticism, and what lessons does it hold for evaluating applied research?
3. Using the revised set-point model, explain why a pay rise and a long commute might have different long-run effects on employee well-being. What data would you need to test your explanation?
4. A brand team wants a campaign to "go viral". Using the arousal × valence framework, evaluate three possible creative concepts and identify the risks of each.
5. Fear and anger are both negative, high-arousal emotions. How would you expect customers' emotions after a data breach to affect their risk perceptions and switching behaviour, and how should the firm's communication differ depending on which emotion dominates?
6. Kahneman and Deaton (2010) and Killingsworth (2021) reached different conclusions about income and emotional well-being. Explain how the adversarial collaboration reconciled them and what it implies for the design of employee compensation and benefits.

25.10 Further Reading

- Fredrickson (2001): the clearest statement of broaden-and-build theory and its hypotheses.
- N. J. L. Brown, Sokal, and Friedman (2013): a forensic critique of the positivity ratio. Essential reading on the misuse of mathematics in psychology.
- Diener, Lucas, and Scollon (2006): the standard reference on revising the hedonic treadmill.
- Russell (1980): the original circumplex model, short and influential.
- Berger and Milkman (2012): a field study of emotion and virality using *New York Times* articles. A model of combining text analysis with theory.
- Lerner et al. (2015): an authoritative review of emotion and decision making, including the appraisal-tendency framework.
- Killingsworth, Kahneman, and Mellers (2023): a model adversarial collaboration on income and emotional well-being.
- Kahneman, Wakker, and Sarin (1997): the conceptual foundation for distinguishing experienced from remembered utility.

Chapter 26

Persuasion and Communication

Why does a car manufacturer fill its print advertisements with fuel-economy figures, crash-test ratings, and warranty terms, while a chewing-gum brand runs a thirty-second spot with a catchy jingle and a famous face? Why does a sponsored post from an influencer persuade some followers and irritate others, and why does the word "#ad" sometimes change how the very same message is received? Why do some stories spread through social networks within hours while carefully produced corporate content goes unseen? And why can a single choice of words, "95% fat free" rather than "5% fat", "lives saved" rather than "lives lost", reverse what people prefer?

Each puzzle concerns how messages change minds, and each has been addressed by a distinct theoretical tradition. Social psychology contributed **dual-process (Section 23.6) models of persuasion**, which explain when people scrutinise arguments and when they rely on cues. Consumer research added the **persuasion knowledge model**, which treats audiences as active agents who recognise and cope with influence attempts. Marketing and sociology developed theories of **word of mouth and social transmission**, which ask why people pass information on. Mass-communication research produced **agenda setting**, **framing**, and **uses and gratifications**, which explain what media make salient, how they package issues, and why audiences choose the media they do. This chapter develops each tradition, shows how they connect, and translates them into hypotheses and managerial practice (Table 26.1).

Table 26.1. Chapter at a glance: theories of persuasion and communication.

Theory	Core question	Level of analysis	Key source
Elaboration likelihood model	When do arguments, rather than cues, change attitudes?	Individual	Petty & Cacioppo (1986)
Heuristic-systematic model	How do heuristic and systematic processing combine?	Individual	Chaiken (1980)
Persuasion knowledge model	How do targets recognise and cope with influence attempts?	Individual / dyad	Friestad & Wright (1994)
Social transmission (word of mouth)	Why do people share, and what gets shared?	Individual / network	Berger (2014)
Two-step flow	How does media influence travel through interpersonal ties?	Network / community	Katz & Lazarsfeld (1955)
Undirected communication (microblogging)	Why broadcast to no one in particular?	Individual	Buechel & Berger (2017)
Agenda setting	How do media shape what the public thinks is important?	Society / public	McCombs & Shaw (1972)

Table 26.1. Chapter at a glance: theories of persuasion and communication. *(continued)*

Theory	Core question	Level of analysis	Key source
Framing	How does the packaging of information change judgment?	Individual / society	Tversky & Kahneman (1981), Entman (1993)
Uses and gratifications	Why do audiences choose the media they do?	Individual	Katz, Blumler & Gurevitch (1973)

Two organising ideas recur throughout. First, **receivers are not passive**. The earliest research on mass communication imagined messages as injections whose effects could be read off their content. Each theory in this chapter replaces that image with an active audience: one that decides how much effort to invest (dual-process models), interprets the sender's motives (persuasion knowledge), selects and passes on content for its own reasons (social transmission, uses and gratifications), and constructs meaning from the frames it encounters. Second, **effects are conditional**. None of these theories predicts that a message will work everywhere. Each specifies the moderators, involvement, accessibility of motives, emotional arousal, issue obtrusiveness, audience goals, that determine when it will.

26.1 The Elaboration Likelihood Model

THEORY CARD

Core claim. Attitudes can change through two qualitatively different routes. When people are motivated and able to think about a message, persuasion depends on the quality of its arguments (the **central route**). When motivation or ability is low, persuasion depends on simple cues such as source attractiveness or the number of arguments (the **peripheral route**).
Origins. Social psychology. Petty and Cacioppo (1986), building on the Yale communication programme (Hovland, Janis, and Kelley 1953) and cognitive-response research. First tested in an advertising context by Petty, Cacioppo, and Schumann (1983).
Level of analysis. Individual message recipient.
Key constructs. Elaboration likelihood, motivation (personal relevance, involvement, need for cognition), ability (distraction, repetition, prior knowledge), argument quality, peripheral cues, attitude strength.
Boundary conditions. Assumes recipients hold or form an attitude toward a definable object. Applies most directly to message-based influence. Predictions about route depend on accurately measuring or manipulating elaboration rather than inferring it from outcomes.
Related theories. Heuristic-systematic model, Persuasion knowledge model, Dual-process theory (Section 23.6), Processing fluency (Section 23.8), Mere exposure (Section 29.1).

26.1.1 Origins and intellectual history

The Yale Communication Research Program after the Second World War organised persuasion research around the question "who says what to whom with what effect", and studied source credibility, message order, fear appeals, and audience characteristics as separate variables (Hovland, Janis, and Kelley 1953). The programme produced a large body of findings but also a frustrating pattern: the same variable, an expert source, a long message, a fear appeal, sometimes increased persuasion,

sometimes had no effect, and sometimes backfired. By the 1970s, the field lacked a framework for predicting *which* outcome would occur.

Petty and Cacioppo (1986) proposed that the inconsistencies arose because researchers had not taken into account how much recipients were thinking. They defined **elaboration** as the extent to which a person thinks about the issue-relevant arguments contained in a message, and argued that elaboration likelihood varies along a continuum. At the high end, people scrutinise arguments, generate their own favourable or unfavourable thoughts, and integrate those thoughts into an attitude. At the low end, people form attitudes by applying simple inferences, "experts are usually right", "if so many arguments are offered, the position must be sound", or through affective associations. Variables such as source expertise therefore have different effects at different levels of elaboration.

26.1.2 Core logic and propositions

The model can be summarised in a set of propositions that correspond closely to the postulates set out by Petty and Cacioppo (1986):

- **P1 (Correctness motive).** People are motivated to hold correct attitudes, but the amount of effort they are willing or able to devote to evaluating a message varies with individual and situational factors.
- **P2 (Elaboration continuum).** Elaboration varies continuously from low to high. The two "routes" describe the endpoints of this continuum rather than two discrete boxes.
- **P3 (Trade-off).** As elaboration increases, the impact of argument quality on attitudes increases and the impact of peripheral cues decreases.
- **P4 (Multiple roles).** A single variable can affect persuasion in different ways: it can serve as an argument, act as a peripheral cue, influence the amount of elaboration, or bias the direction of elaboration.
- **P5 (Attitude strength).** Attitudes formed or changed through the central route are more persistent over time, more resistant to counter-persuasion, and more predictive of behaviour than attitudes changed through the peripheral route.

DEFINITION

Definition 26.1. Argument quality is defined empirically in ELM research: strong arguments are those that, when people are instructed to think about them, elicit predominantly favourable thoughts. Weak arguments elicit predominantly unfavourable thoughts. **Peripheral cues** are features of the persuasion context that allow an attitude judgment without scrutiny of the issue-relevant merits: source attractiveness, celebrity, the sheer number of arguments, audience applause.

26.1.3 Mechanism

Figure 26.1 traces the two routes. Motivation and ability act as gatekeepers. If both are present, the recipient processes the arguments and the valence of the resulting thoughts drives attitude change. If either is missing, the recipient looks for a peripheral cue. Note the dashed path in the figure: when no cue is present, or when cognitive processing produces no net change, the person retains the initial attitude.

The signature prediction of the model is an **interaction** between elaboration and message features. The best-known demonstration is Petty, Cacioppo, and Schumann (1983), who showed participants

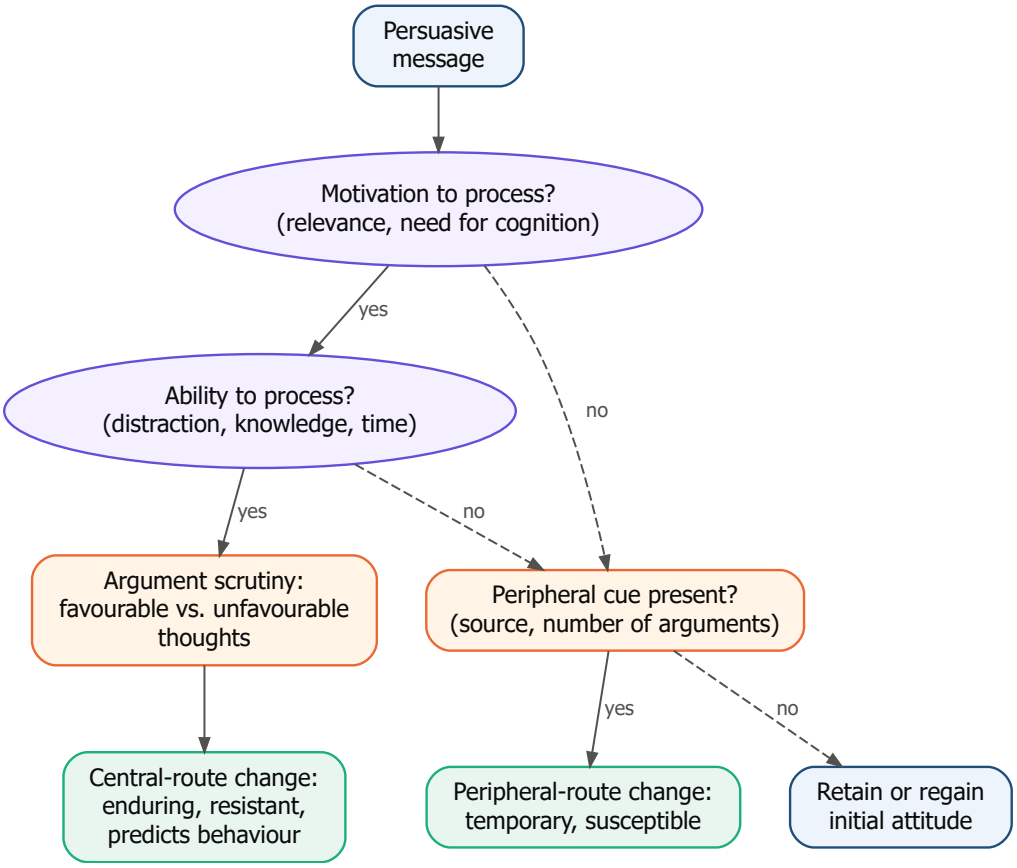


Figure 26.1. The elaboration likelihood model. Motivation and ability determine whether a message is processed centrally (argument scrutiny, enduring attitude change) or peripherally (cue-based, temporary change). Adapted from the logic in Petty and Cacioppo (1986).

a magazine advertisement for a disposable razor. Involvement was manipulated by telling some participants that the product would soon be available in their area and that they would later choose among razor brands (high involvement), and telling others the product would not be available locally (low involvement). The advertisement contained either strong or weak arguments and was endorsed either by famous athletes or by ordinary citizens. Argument quality had a much larger effect on attitudes under high involvement, whereas the celebrity endorsement mattered more under low involvement. An earlier study had found the same pattern for argument quality versus source expertise in a policy context: students evaluating a proposal to introduce comprehensive examinations either at their own university next year (high relevance) or at some later date (Petty, Cacioppo, and Goldman 1981).

Figure 26.2 illustrates the crossover logic with simulated data. The two panels are mirror images: the factor that matters most under high involvement (argument quality) matters least under low involvement, and vice versa for the peripheral cue.

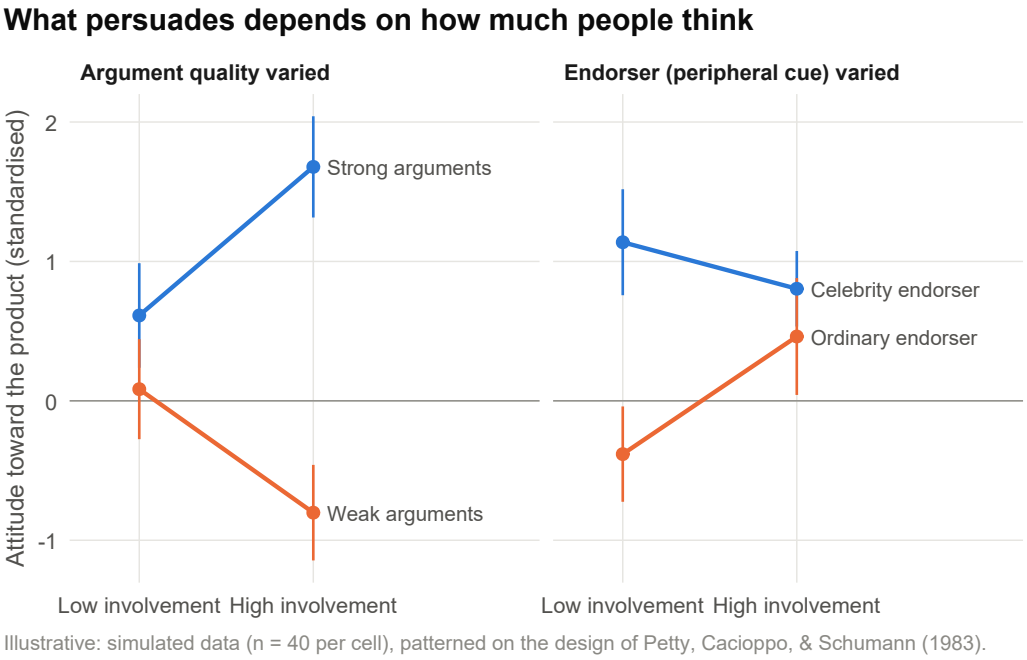


Figure 26.2. Illustrative simulation of the ELM interaction. Under high involvement, argument quality drives attitudes and the endorser matters little. Under low involvement, the pattern reverses. Points are cell means with 95% confidence intervals. Data are simulated, not empirical.

26.1.4 The heuristic-systematic model

The ELM's closest relative is the **heuristic-systematic model** (HSM), developed by Chaiken (1980) and elaborated in Eagly and Chaiken (1993). **Systematic processing** is a comprehensive, analytic treatment of judgment-relevant information. **Heuristic processing** applies simple decision rules ("experts can be trusted", "consensus implies correctness") that are stored in memory and cued by the situation. In her original experiments, Chaiken manipulated recipients' issue involvement together with the likability of the communicator and the number of arguments in the message. Communicator likability mattered more when involvement was low, whereas the content of the arguments carried more weight when involvement was high (Chaiken 1980).

The two models are often treated as interchangeable, but they differ in emphasis. The HSM's **sufficiency principle** holds that people exert as much effort as needed to reach a desired level of judgmental confidence, and its **co-occurrence assumption** holds that the two modes can operate simultaneously: heuristic cues can add to systematic processing, be attenuated by it, or bias it by shaping how ambiguous arguments are interpreted. Table 26.2 compares the models.

Table 26.2. Two dual-process models of persuasion compared.

Feature	Elaboration likelihood model	Heuristic-systematic model
Key source	Petty & Cacioppo (1986)	Chaiken (1980), Eagly & Chaiken (1993)
Low-effort mode	Peripheral route: any cue-based process (heuristics, affect, conditioning)	Heuristic processing: application of stored decision rules
High-effort mode	Central route: scrutiny of issue-relevant arguments	Systematic processing: comprehensive analysis of information

Table 26.2. Two dual-process models of persuasion compared. *(continued)*

Feature	Elaboration likelihood model	Heuristic-systematic model
Relation between modes	Trade-off: cue impact falls as elaboration rises	Co-occurrence: modes can add, attenuate, or bias one another
What drives effort	Motivation and ability	Sufficiency principle: gap between actual and desired confidence
Motives considered	Primarily accuracy (correct attitudes)	Accuracy, defence, and impression motives
Signature concept	Multiple roles of a single variable	Biased systematic processing
Typical managerial reading	Match creative strategy to involvement	Use cues that also shape how arguments are read

26.1.5 Evidence and extensions

The dual-process framework is among the most heavily tested in social psychology. Several lines of evidence stand out:

- **Involvement.** A meta-analysis by B. T. Johnson and Eagly (1989) found that the type of involvement matters. *Outcome-relevant* involvement (the issue has consequences for one's goals) interacts with argument quality in the way the ELM predicts, whereas *value-relevant* involvement (the issue engages core values) tends to reduce persuasion regardless of argument quality, because people defend their existing positions.
- **Individual differences.** People high in **need for cognition**, a stable tendency to engage in and enjoy effortful thinking, are more influenced by argument quality and less by peripheral cues (Cacioppo and Petty 1982).
- **Multiple roles.** The same variable can play different roles. An attractive endorser is a peripheral cue for a low-involvement snack, but an *argument* for a beauty product whose benefit is appearance. Under moderate elaboration, a surprising source can increase scrutiny itself (Petty and Cacioppo 1986).
- **Attitude strength.** Central-route attitudes are more durable and better predictors of behaviour, which is why the quality of the attitude, not only its valence, matters for long-run outcomes such as repeat purchase.

DEBATE AND EVIDENCE

One route or two? Kruglanski and Thompson (1999) argued in their **unimodel** that cues and arguments are not processed by different routes at all: both are "evidence" from which people draw conclusions, and the apparent dual-route pattern arises because studies typically present cues briefly and early and arguments at length and later. When cue information is made as lengthy and complex as argument information, the unimodel predicts that it, too, will matter most under high elaboration. Defenders of the dual-process view reply that the ELM was always defined in terms of the *process* (scrutiny of merits versus simple inference) rather than the *content* type, and that the multiple-roles postulate already anticipates a cue acting as an argument. The debate remains instructive for researchers: claims about "route" must be supported by process evidence, thought listings, argument-quality manipulations, response latencies, rather than inferred from which variable happened to predict attitudes.

26.1.6 Applications in business

The ELM translates directly into decisions about creative strategy, media selection, and sales communication. The core managerial question is not "which message is best?" but "how much will this audience be thinking when it encounters the message?"

BUSINESS CASE

Ad creative for high- versus low-involvement categories. Categories differ systematically in the elaboration they invite. A buyer comparing mortgages, enterprise software, or family cars faces significant financial and performance risk. The buyer is motivated to scrutinise claims and often has time to do so. Communication in these categories typically leads with verifiable arguments, specifications, total cost of ownership, independent ratings, case studies, and supplies them in formats that support processing (comparison tables, configurators, white papers). By contrast, a shopper choosing chewing gum, soft drinks, or snacks is rarely motivated to read argument-heavy copy. Advertising in these categories leans on cues that operate with minimal thought: celebrity and likeable characters, music, humour, brand repetition, and packaging that is easy to recognise at shelf. The ELM also explains why the *same* category can require both approaches. A first-time buyer of a smartphone may be highly involved, whereas a loyal repeat buyer is not. Brands therefore pair detailed launch communications with simpler reminder advertising.

26.1.7 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The effect of argument strength on brand attitude will be stronger when purchase is imminent (high outcome-relevant involvement) than when it is distant.
- *H2.* The effect of influencer popularity on purchase intention will be stronger for low-priced than for high-priced products.
- *H3.* Attitudes formed after argument-rich advertising will predict repeat purchase better than equally favourable attitudes formed after cue-based advertising.

Designs. Factorial experiments that cross argument quality with a motivation or ability manipulation (relevance, distraction, time pressure). Field experiments that vary creative executions across audiences with different involvement. Thought-listing protocols to measure elaboration directly.

Measures. Thought listing and cognitive-response coding, need for cognition scale (Cacioppo and Petty 1982), attitude strength (certainty, accessibility, resistance to counter-messages), recall of arguments versus cues.

Contexts. Advertising pretesting, health communication, sales presentations, online reviews (review content as arguments, star ratings as cues), and recommendation interfaces.

COMMON PITFALL

Inferring the route from the result. A common misuse is to observe that a cue predicted attitudes and conclude that people "used the peripheral route", or that argument quality predicted attitudes and conclude that they "used the central route". This is circular. Route should

be established independently, by manipulating motivation or ability, or by measuring elaboration, and the theory's evidence is the *interaction* between elaboration and message features, not a main effect of either.

26.2 The Persuasion Knowledge Model

THEORY CARD

Core claim. Consumers develop knowledge about persuasion, what marketers and other agents are trying to do, how their tactics work, and how to respond, and use that knowledge to interpret and cope with influence attempts. Recognising a behaviour as a persuasion tactic can change its meaning and its effect.
Origins. Consumer research, Friestad and Wright (1994), tested and extended by M. C. Campbell and Kirmani (2000).
Level of analysis. Individual target and dyad (target and persuasion agent).
Key constructs. Target, agent, persuasion attempt, persuasion episode, persuasion knowledge, agent knowledge, topic knowledge, coping behaviour, change-of-meaning principle, accessibility of ulterior motives, cognitive capacity.
Boundary conditions. Persuasion knowledge must be available (learned) and accessible (activated) at the moment of exposure, and the target must have the capacity to apply it. Children, novices, and distracted consumers use it less.
Related theories. Elaboration likelihood model, Signaling theory (Section 8.2), Heuristics and biases (Section 23.2).

26.2.1 Core logic

Friestad and Wright (1994) reframed persuasion from the target's point of view. Earlier theories asked how a message changes an attitude. The persuasion knowledge model (PKM) asks how ordinary people make sense of the fact that someone is trying to influence them. Over a lifetime, consumers acquire folk theories about advertising, selling, negotiation, and social influence: why agents do what they do, which tactics are fair, and which signal manipulation. These beliefs form **persuasion knowledge**, one of three knowledge structures that targets bring to an encounter (Table 26.3).

Table 26.3. The three knowledge structures in the persuasion knowledge model (after Friestad & Wright, 1994).

Knowledge structure	What it contains	Example question the target asks	Business illustration
Persuasion knowledge	Beliefs about agents' goals and tactics, their effectiveness, and their appropriateness	"Why is this salesperson complimenting me?"	Recognising scarcity claims as a pressure tactic
Agent knowledge	Beliefs about the specific agent's traits, competence, and goals	"Is this brand trustworthy?"	Prior experience with a retailer's return policy
Topic knowledge	Beliefs about the product, issue, or service being promoted	"Do these specifications make sense?"	An engineer evaluating a component supplier's claims

Two ideas carry most of the model's predictive power. The first is the **change-of-meaning prin-**

ciple: once a target recognises an agent's action as a persuasion tactic, the action can acquire a different meaning. Flattery that seemed sincere becomes ingratiation, a "limited-time offer" becomes a pressure tactic, a glowing product mention becomes advertising. The second idea is that targets are **coping**, not merely reacting. Their goal is to manage the episode in a way that serves their own interests, sometimes resisting, sometimes accepting the message while discounting the source, sometimes enjoying a clever tactic for what it is.

26.2.2 Mechanism

Persuasion knowledge does not operate automatically. M. C. Campbell and Kirmani (2000) showed that its use depends on the **accessibility** of the agent's ulterior motive and on the target's **cognitive capacity**. In their studies, consumers who encountered a salesperson's flattery were more likely to infer an ulterior motive when the motive was highly accessible, for example, when the flattery came *before* a purchase decision rather than after, and cognitively busy targets were less likely to draw on persuasion knowledge unless the motive was highly salient. Observers, who had spare capacity, were more likely than the targets themselves to see the tactic for what it was. Figure 26.3 summarises the mechanism.

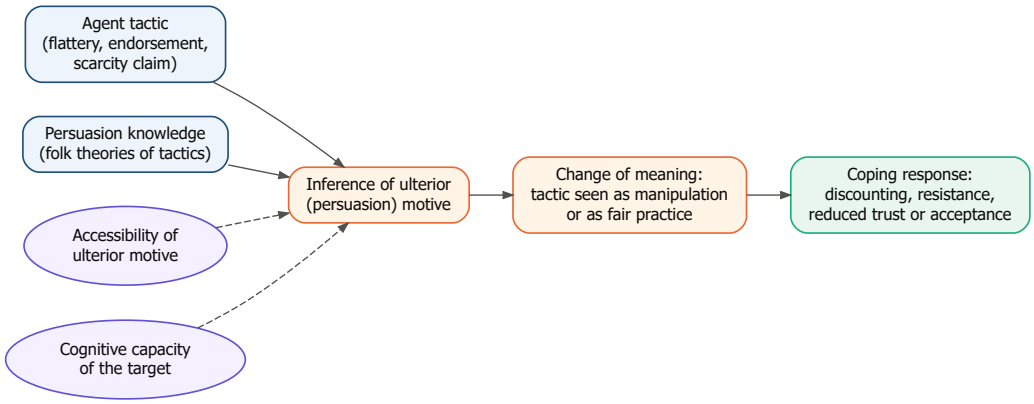


Figure 26.3. Mechanism of the persuasion knowledge model. An agent's tactic leads to persuasion-motive inference only when persuasion knowledge is accessible and the target has capacity to use it. Recognition changes the tactic's meaning and shapes coping.

The PKM and the ELM are complementary. Persuasion knowledge can be thought of as one of the things people think *about* when they elaborate. A highly involved target may scrutinise not only the arguments but also the agent's motives. Recognising a manipulative intent can then produce counter-arguments and resistance. Conversely, the PKM explains why peripheral cues sometimes fail: a celebrity endorsement that is recognised as purchased may stop working as a cue of quality and start working as a cue of persuasive intent.

26.2.3 Evidence and extensions

The model has been used to study advertising literacy among children and adolescents, covert marketing, product placement, native advertising, sales interactions, and corporate social responsibility claims. A recurring finding is that recognising persuasive intent is not uniformly harmful to the persuader. When a tactic is judged appropriate, an honest disclosure, a transparent comparison, recognition can increase perceived fairness and trust. When a tactic is judged inappropriate, a hid-

den payment, a deceptive claim, recognition produces discounting and a lasting penalty to agent knowledge ("this brand tries to trick people"). The model therefore predicts asymmetric long-run consequences: the short-run gains from covert tactics must be weighed against the costs incurred when they are discovered.

BUSINESS CASE

Influencer marketing and sponsorship disclosure. Influencer content is persuasive in part because it looks like ordinary peer communication rather than advertising. The PKM predicts that disclosing the commercial relationship, "#ad", "paid partnership", "sponsored", activates persuasion knowledge, increasing the likelihood that followers recognise the post as advertising. Regulators have taken the same view: the US Federal Trade Commission's guidance on endorsements requires that material connections between endorsers and brands be disclosed clearly and conspicuously, and major platforms provide built-in paid-partnership labels. For brands, the managerial question is not whether to disclose (that is a legal obligation in many jurisdictions) but how to design partnerships that survive disclosure. The model's answer is **fit and appropriateness**: when the influencer plausibly uses and values the product, the persuasion motive is seen as compatible with the influencer's genuine opinion. When the fit is poor, disclosure highlights the payment and invites discounting. Several practical rules follow: choose creators whose prior content demonstrates authentic topic knowledge, allow them to express qualified opinions, and avoid scripts that read as advertising copy.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Sponsorship disclosure increases recognition of advertising, and this recognition mediates a decrease in source trustworthiness.
- *H2.* The negative effect of disclosure on brand attitude is attenuated when influencer-product fit is high.
- *H3.* Cognitive load reduces the effect of an inappropriate sales tactic on perceived manipulative intent (M. C. Campbell and Kirmani 2000).

Measures. Advertising recognition, inferences of manipulative intent, perceived appropriateness of tactics, source trustworthiness, resistance (counter-arguing, sharing intentions).
Designs. Between-subject experiments varying disclosure presence, wording, and position. Field experiments with platform disclosure labels. Eye tracking to measure whether disclosures are noticed.

COMMON PITFALL

Assuming persuasion knowledge is always activated. Many studies equate the *presence* of a disclosure with *activation* of persuasion knowledge. Disclosures that are small, late, or buried in hashtags may not be noticed. Measure recognition directly, and remember that knowledge must be both available and accessible, and the target must have capacity to use it.

26.3 Word of Mouth and Social Transmission

THEORY CARD

Core claim. People share information, opinions, and content because sharing serves their own needs: managing impressions, regulating emotions, acquiring information, bonding with others, and persuading. These motives, together with features of the content and the audience, determine *what* is shared and how far it spreads.
Origins. Sociology of personal influence (E. Katz and Lazarsfeld 1955), marketing research on word of mouth (Godes and Mayzlin 2004, Chevalier and Mayzlin 2006), synthesised in Berger (2014) and Berger (2013).
Level of analysis. Individual sharer, dyad, and social network.
Key constructs. Word of mouth (WOM), sharer motives, content characteristics (emotional arousal, practical value, social currency), audience size, triggers, opinion leaders, cascades.
Boundary conditions. Predictions about content features assume people choose what to share. Network position, platform algorithms, and paid amplification can dominate content effects in some settings.
Related theories. Emotion and virality (Section 25.4), Diffusion of innovations (Section 10.1), Weak ties (Section 9.3), Identity signaling (Section 8.5), Microblogging.

26.3.1 From mass media to personal influence: the two-step flow

The study of social transmission began with a surprise. When Lazarsfeld and his colleagues studied how voters in Erie County, Ohio, made up their minds during the 1940 US presidential campaign, they expected to find strong direct effects of radio and print media. Instead, they found that many voters reported being influenced by other people, and that ideas often seemed to flow from the media to **opinion leaders** and from them to less active sections of the population (Lazarsfeld, Berelson, and Gaudet 1944). E. Katz and Lazarsfeld (1955) developed this **two-step flow** hypothesis in a study of women in Decatur, Illinois, examining influence in marketing, fashion, public affairs, and film-going. Opinion leaders were found at every social level, differed by domain, and were distinguished by greater media exposure and by their social position, not merely by status (Figure 26.4).

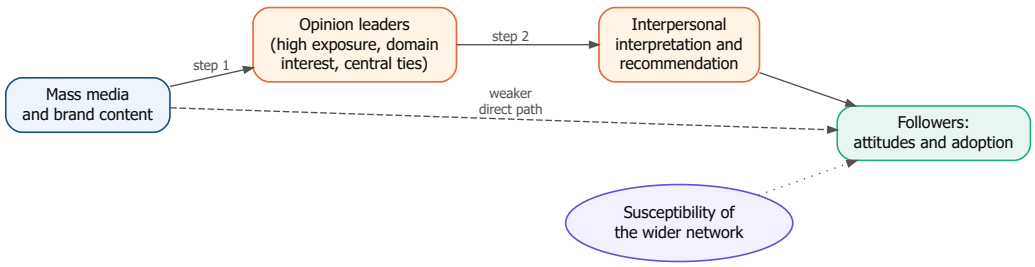


Figure 26.4. The two-step flow of communication (Katz & Lazarsfeld, 1955) and its modern revision. Media content reaches opinion leaders, who interpret and relay it through interpersonal ties. Network research suggests that large cascades depend on many easily influenced individuals rather than a few influentials (Watts & Dodds, 2007).

The two-step flow shaped decades of marketing practice, from seeding products with "influentials" to opinion-leader scales in consumer research. Later network research has qualified it. In simulations of influence networks, D. J. Watts and Dodds (2007) found that large cascades were usually

driven not by a small number of highly influential people but by a critical mass of easily influenced individuals. Influentials can trigger cascades, but only when the network is ready. The practical implication is that seeding strategy depends on network structure as much as on the identity of the seed.

26.3.2 Why people share

Berger (2014) organises the literature around the **functions** that word of mouth serves for the sharer. The framework explains why sharing is not a neutral relay of information: what people choose to say reflects what the act of saying it does for them (Table 26.4).

Table 26.4. Five functions of word of mouth (organised following Berger, 2014).

Function	What sharing does for the sharer	Predicted content features	Business implication
Impression management	Signals identity, expertise, and status, supports self-enhancement	Remarkable, novel, or insider content, positive self-relevant stories	Give customers something that makes them look good to share
Emotion regulation	Generates social support, venting, sense-making, and reducing dissonance	Emotionally intense experiences, positive and negative	Expect negative WOM after service failures, offer channels to vent
Information acquisition	Seeks advice and resolves uncertainty	Questions, requests for recommendations	Be present where customers ask for advice
Social bonding	Reinforces shared views and reduces loneliness	Common-ground content, shared experiences	Design for communal experiences and conversation
Persuasion	Convinces others or changes their behaviour	Arguments, strong opinions, polarised content	Equip advocates with credible arguments

Two further findings shape what spreads. First, **audience size** matters. Barasch and Berger (2014) showed that people who **broadcast** to many recipients focus on self-presentation and avoid sharing content that makes them look bad, whereas people who **narrowcast** to one recipient focus more on the recipient and share more useful content. Second, sharing is **intrinsically rewarding**: in neuroimaging and behavioural studies, Tamir and Mitchell (2012) found that disclosing information about the self engaged brain regions associated with reward, and that participants were willing to forgo small monetary payments to talk about themselves.

26.3.3 What gets shared: content characteristics

Berger and Milkman (2012) analysed which *New York Times* articles made the newspaper's most-emailed list over a three-month period. Controlling for placement and other features, positive content was more likely to be shared than negative content, but emotion was not simply a matter of valence. Content evoking high-arousal emotions, awe, anger, and anxiety, was more likely to be shared, whereas content evoking a low-arousal emotion, sadness, was less likely to be shared. A complementary experiment found that physiological arousal, even arousal induced by an unrelated activity such as jogging in place, increased participants' willingness to share content (Berger 2011). Figure 26.5 places these emotions in the valence-arousal space.

Content characteristics also determine *when* sharing happens. Berger and Schwartz (2011) found that products that were more publicly visible and more frequently cued by the environment received more word of mouth, and that interesting products received more immediate but not more

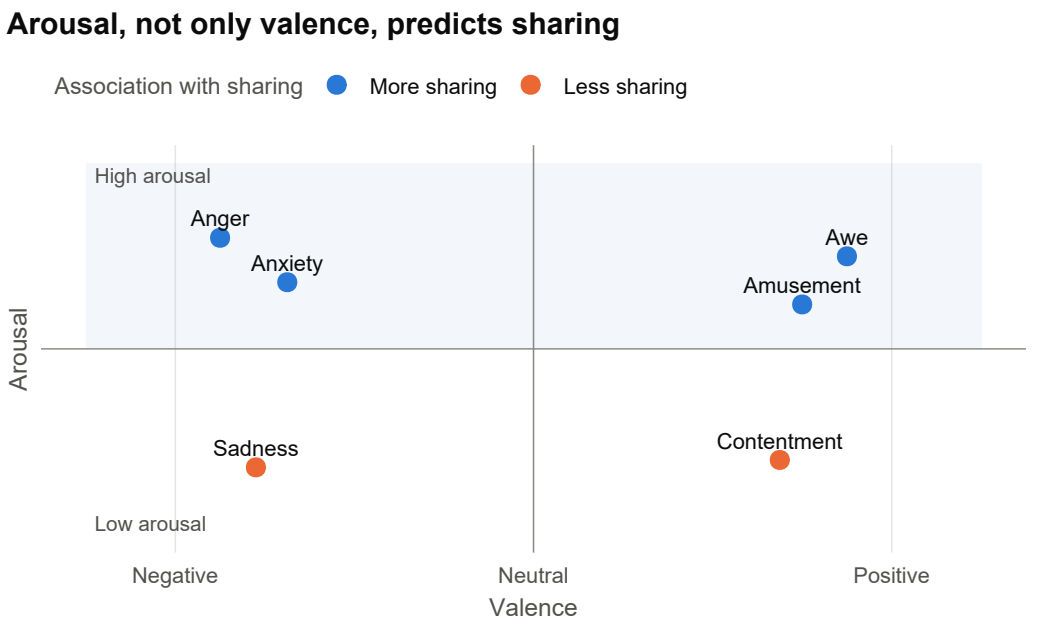


Figure 26.5. Conceptual map of emotions by valence and arousal, with the direction of their association with sharing reported by Berger & Milkman (2012) and Berger (2011). Positions are illustrative, not measured values.

ongoing word of mouth. Environmental **triggers**, cues that bring a product to mind, sustain conversation long after novelty fades.

26.3.4 The STEPPS framework

For practitioners, Berger (2013) synthesised this research into six principles, summarised in Table 26.5. The framework is a checklist rather than a formal theory, but each element maps onto a mechanism with independent empirical support.

Table 26.5. The STEPPS framework for designing shareable content (Berger, 2013).

Principle	Underlying mechanism	Design question	Illustrative tactic
Social currency	Impression management, self-enhancement	Does sharing make the sharer look good, smart, or in the know?	Insider access, remarkable facts, game mechanics
Triggers	Accessibility, environmental cueing	What in the everyday environment will bring this to mind?	Link the product to a frequent context or time of day
Emotion	High-arousal emotion increases transmission	Does the content evoke awe, amusement, anger, or anxiety?	Focus on feelings rather than features
Public	Observability, social proof	Can others see people using it?	Visible packaging, shareable badges, branded artefacts
Practical value	Other-oriented helping, usefulness	Is it useful enough to pass on?	Tips, deals that are framed as unusual, how-to guides
Stories	Narrative transportation carries information	Is the message embedded in a story people want to tell?	Make the brand integral to the narrative, not an add-on

BUSINESS CASE

Designing shareable content. A consumer-electronics firm wants its product launch to generate organic conversation rather than relying entirely on paid media. Social transmission research suggests several design choices. To create **social currency**, the launch might offer early access to a limited group, giving those customers something remarkable to report. To create **triggers**, the firm might tie the product to a recurring context, a commute, a weekly routine, so that it is cued repeatedly after launch. The firm would aim for **high-arousal emotion** (awe at a capability, amusement in a demonstration) rather than low-arousal contentment. It would make use **public** through visible design or a shareable output, and package **practical value** in tips that users would pass on to help friends. Finally, it would embed the brand in a **story**, the problem a customer solved, rather than adding a logo to unrelated entertainment. The theory also flags risks: content optimised for arousal can attract anger as easily as awe, and broadcast audiences reward self-presentation, so content that makes sharers look foolish will not travel.

26.3.5 Word of mouth and market outcomes

Does word of mouth actually move sales? Field evidence suggests that it does, but in ways that depend on the structure of conversation. Godes and Mayzlin (2004) found that the *dispersion* of online conversations about television shows across different communities was associated with subsequent ratings, suggesting that conversations that cross social boundaries carry more new information than conversations within one community. Chevalier and Mayzlin (2006) exploited differences in reviews between two online booksellers and found that improvements in a book's reviews on one site were associated with relative increases in its sales on that site, and that very negative reviews had a larger effect than very positive ones. Both studies illustrate a methodological challenge that recurs throughout this literature: because word of mouth and sales are both driven by product quality, credible identification requires comparison across sites, communities, or time.

26.3.6 Misinformation and the spread of falsehood

Social transmission theory also illuminates why false information spreads. Vosoughi, Roy, and Aral (2018) analysed the diffusion of roughly 126,000 rumour cascades on Twitter from 2006 to 2017, classifying each as true or false using the verdicts of six independent fact-checking organisations. Falsehood diffused significantly farther, faster, deeper, and more broadly than the truth in all categories of information, and the effects were more pronounced for false political news than for other categories. False stories were about 70% more likely to be retweeted than true ones. Two findings bear directly on mechanism. First, false news was more **novel** than true news, and replies to false stories expressed more surprise and disgust, consistent with the view that novelty and emotional arousal drive sharing. Second, automated accounts ("bots") accelerated the spread of true and false news at similar rates, implying that the difference was driven by human sharing decisions rather than by bots.

These findings connect directly to the functions of word of mouth. Novel information confers social currency on the sharer. Arousing content is more likely to be passed on, and broadcasting to many followers emphasises self-presentation over careful verification. Correcting misinformation is difficult because of the **continued influence effect**: retracted information often continues to shape reasoning, especially when no alternative explanation is offered (Lewandowsky et al. 2012). Interventions that shift attention to accuracy at the moment of sharing, for example, asking people

to rate the accuracy of an unrelated headline, have been found to improve the quality of news that people subsequently share, including in a field experiment on Twitter (Pennycook et al. 2021).

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Content that evokes high-arousal emotions will be shared more than content evoking low-arousal emotions of the same valence (Berger and Milkman 2012).
- *H2.* The effect of social-currency features on sharing will be stronger when sharers broadcast to large audiences than when they narrowcast (Barasch and Berger 2014).
- *H3.* Brands linked to frequent environmental triggers will receive more ongoing (but not more immediate) word of mouth (Berger and Schwartz 2011).

Designs. Archival analysis of shares, retweets, and reviews with content coded by text analysis. Field experiments with seeding strategies. Laboratory experiments that manipulate arousal or audience size. Natural experiments created by platform changes.
Measures. Share counts and cascade depth, dispersion of conversation across communities, sentiment and emotion dictionaries, self-reported sharing motives.

COMMON PITFALL

Treating virality as a property of content alone. Whether content spreads depends on content features *and* on network structure, platform ranking, paid seeding, and timing. Studies that correlate content features with share counts without accounting for exposure (how many people saw the content) confound the probability of sharing with the opportunity to share. Normalise by exposure where possible, and treat cascade size as a highly skewed outcome.

26.4 Microblogging and Undirected Communication

THEORY CARD

Core claim. Microblogging, status updates, posts, and short messages broadcast to one's network, is valuable precisely because it is **undirected**: it lets people reach out and invite social interaction without imposing communication on any particular person.
Origins. Consumer psychology, Buechel and Berger (2018).
Level of analysis. Individual communicator.
Key constructs. Directed versus undirected communication, social apprehension, fear of rejection, social connection.
Boundary conditions. Most relevant where people have a choice between directed channels (messages, calls) and undirected ones (public posts), and where the audience is a loose network rather than a close dyad.
Related theories. Social transmission, Uses and gratifications, Social comparison theory (Section 22.1).

Directed communication, a phone call, a direct message, an email to a specific person, carries an implicit demand: the recipient is expected to respond. For people who worry about being rejected or about imposing on others, that demand creates a barrier to reaching out. Buechel and Berger (2018) argue that undirected communication removes the barrier. A status update is addressed to no one in particular. Whoever responds does so voluntarily, so the poster can seek interaction

without the risk of a specific person declining it. Their studies suggest that **socially apprehensive** individuals, those who are anxious about communicating with others and concerned about rejection, find undirected communication particularly attractive.

The insight complements the broader transmission literature. Barasch and Berger (2014) showed that audience size changes *what* people share. Buechel and Berger (2018) show that directness changes *whether* some people communicate at all. Research on why people post on microblogging platforms also distinguishes intrinsic motives (enjoyment of posting) from image-related motives (the status of having an audience). Toubia and Stephen (2013) found evidence that image-related utility plays a substantial role for many non-commercial users.

BUSINESS CASE

Community design and customer voice. Brands that host customer communities often find that a small share of members post while most only read. Undirected communication research suggests one reason and one remedy. Directed requests ("Ask an expert", "Message our team") may deter socially apprehensive customers, whereas low-stakes undirected formats, open threads, reactions, "share your setup" posts, let them participate without making a specific demand on anyone. The same logic applies inside organisations: enterprise social platforms can surface questions from employees who would hesitate to email a senior colleague directly. Designers should pair such formats with norms that make responses likely, because an undirected post that receives no response can itself feel like rejection.

26.5 Agenda Setting

THEORY CARD

Core claim. The news media may not tell people *what to think*, but they are highly influential in telling people *what to think about*: the issues that receive the most coverage come to be seen by the public as the most important.
Origins. Political communication, McCombs and Shaw (1972), building on B. C. Cohen (1963) and on the voter studies of Lazarsfeld, Berelson, and Gaudet (1944), extended to priming by S. Iyengar and Kinder (1987).
Level of analysis. Society and public opinion (aggregate issue salience). Individual (issue importance judgments).
Key constructs. Media agenda, public agenda, issue salience, obtrusive versus unobtrusive issues, need for orientation, attribute (second-level) agenda setting, priming.
Boundary conditions. Strongest for **unobtrusive** issues that people cannot observe directly (foreign policy, macroeconomic trends) and among people with a high **need for orientation**. Weaker for issues experienced first-hand. Fragmented and personalised media environments complicate the notion of a single media agenda.
Related theories. Framing, Spiral of silence (Section 22.8), Pack journalism (Section 30.7), Availability heuristic (Section 23.2).

26.5.1 Core logic and evidence

B. C. Cohen (1963) observed that the press may often fail to tell people what to think but is remarkably successful in telling them what to think about. McCombs and Shaw (1972) turned this

observation into a testable hypothesis. During the 1968 US presidential campaign, they surveyed undecided voters in Chapel Hill, North Carolina, about which issues they considered most important, and content-analysed the news media those voters used. The rank order of issues on the media agenda corresponded closely to the rank order on the voters' agenda. Because undecided voters lacked strong partisan filters, they were a sensible population in which to look for media influence.

The correlational design left open the direction of causality: do media follow the public or lead it?and later work used time-series designs and experiments to address the question. S. Iyengar and Kinder (1987) manipulated the content of television news broadcasts that participants watched over several days and found that issues emphasised in the broadcasts became more important to viewers. They also documented **priming**: issues that were made salient became more heavily weighted in viewers' evaluations of the president. Agenda setting therefore has two consequences: it shapes which issues people think are important, and it shapes which criteria people use to evaluate leaders, firms, and products.

Later research distinguished **first-level** agenda setting (salience of objects such as issues or candidates) from **attribute** or **second-level** agenda setting (salience of particular attributes of those objects). Attribute agenda setting overlaps with framing, and Scheufele and Tewksbury (2007) argue that agenda setting and priming rest on accessibility, making some considerations easier to retrieve, whereas framing rests on applicability: the fit between a frame and how an issue is interpreted. Table 26.6 summarises the distinction.

Table 26.6. Agenda setting, priming, and framing distinguished (following Scheufele & Tewksbury, 2007).

Media effect	What media change	Psychological mechanism	Classic study	Business analogue
Agenda setting	Perceived importance of issues	Accessibility (salience)	McCombs & Shaw (1972)	Which attributes buyers consider at all
Priming	Criteria used to evaluate actors	Accessibility of evaluative standards	Iyengar & Kinder (1987)	Which attribute dominates brand evaluation
Framing	Interpretation of an issue	Applicability of an interpretive schema	Entman (1993), Tversky & Kahneman (1981)	How a price, risk, or crisis is understood

BUSINESS CASE

Issue salience and corporate communication. Agenda setting implies that firms compete not only on positions but on *which issues are discussed at all*. A manufacturer whose products excel on safety benefits when safety dominates media coverage of its category, while a rival that leads on price benefits when affordability dominates. Corporate communication functions therefore track media agendas, invest in research and data releases that journalists cover, and time announcements around the news cycle. The priming corollary is equally important: during a public debate about data privacy, consumers may weight privacy more heavily when evaluating a technology brand, even if their underlying preferences have not changed. Firms whose reputations rest on an attribute that becomes salient should expect amplified effects: positive or negative.

COMMON PITFALL

Equating coverage with persuasion. Agenda setting is about salience, not about direction of opinion. Heavy coverage of an issue can raise its importance while leaving attitudes toward it unchanged, or while polarising them. Research that treats volume of coverage as a measure of favourable influence conflates agenda setting with persuasion.

26.6 Framing Theory

THEORY CARD

Core claim. The way information is presented, which aspects are emphasised, and whether logically equivalent outcomes are described as gains or losses, systematically changes judgments and choices.
Origins. Two traditions: behavioural decision research on **equivalence framing** (Tversky and Kahneman 1981, Levin, Schneider, and Gaeth 1998), and communication and sociology research on **emphasis framing** (Entman 1993).
Level of analysis. Individual (choice and judgment) and societal (public discourse).
Key constructs. Gain versus loss frames, risky-choice, attribute, and goal framing, frame selection and salience, problem definition, causal interpretation, moral evaluation, treatment recommendation.
Boundary conditions. Equivalence-framing effects are strongest when people do not spontaneously translate between frames and when the decision is unfamiliar. Emphasis-framing effects are weaker when audiences hold strong prior attitudes or encounter competing frames.
Related theories. Prospect theory (Chapter 24), Agenda setting, Mental accounting (Section 23.5), Anchoring (Section 29.9).

26.6.1 Equivalence framing

Tversky and Kahneman (1981) presented participants with the "Asian disease problem". An outbreak is expected to kill 600 people, and participants choose between two programmes. In one version, outcomes are described in terms of lives saved: Programme A will save 200 people for certain, whereas Programme B offers a one-third probability of saving all 600 and a two-thirds probability of saving no one. In a second version, the same outcomes are described in terms of deaths: under Programme C, 400 people will die for certain. Under Programme D, there is a one-third probability that nobody will die and a two-thirds probability that all 600 will die. The programmes are numerically identical across versions, yet preferences reverse (Figure 26.6). In the gain frame most participants chose the certain option. In the loss frame most chose the gamble. The pattern follows from prospect theory (Chapter 24): a value function that is concave for gains and convex for losses produces risk aversion in the domain of gains and risk seeking in the domain of losses.

Levin, Schneider, and Gaeth (1998) distinguish three types of equivalence framing that differ in what is framed and in what they affect (Table 26.7). The distinction matters because the types have different mechanisms and different typical effect directions. **Attribute framing** is the most common in marketing: in a classic study, ground beef described as "75% lean" was rated more favourably than beef described as "25% fat", and the difference was attenuated, though not eliminated, after participants tasted the meat (Levin and Gaeth 1988).

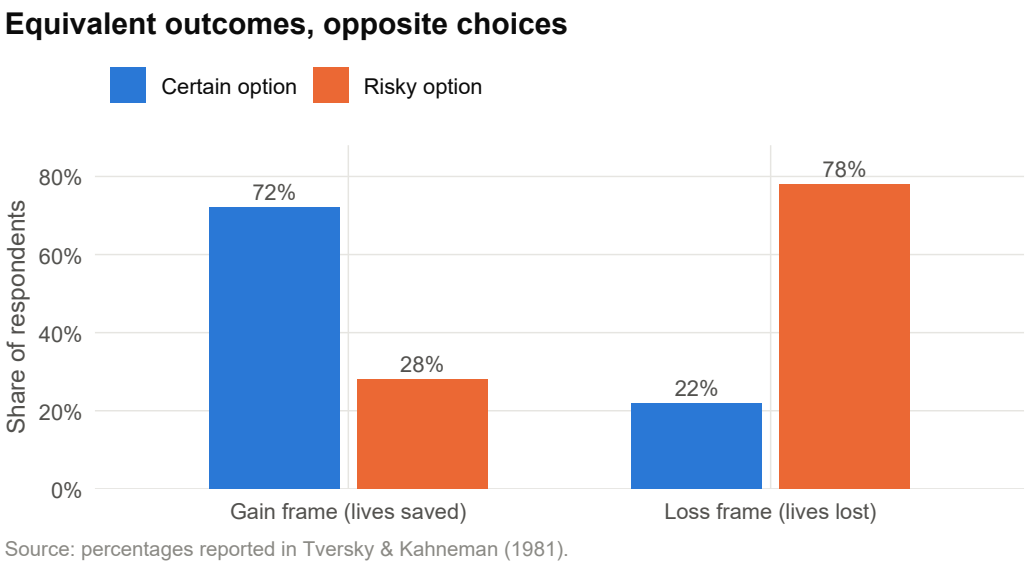


Figure 26.6. The framing of decisions. Percentages choosing the certain versus the risky programme in the Asian disease problem, as reported by Tversky & Kahneman (1981). The gain-frame problem was answered by 152 respondents and the loss-frame problem by 155.

Table 26.7. Three types of equivalence framing (Levin, Schneider, & Gaeth, 1998).

Frame type	What is framed	What is affected	Typical effect	Example
Risky-choice framing	Outcomes of options with different risk levels	Risk preference	Risk aversion for gains, risk seeking for losses	Asian disease problem
Attribute framing	A single attribute of an object	Evaluation of the object	Positive labels yield more favourable evaluations	”75% lean” vs. ”25% fat”
Goal framing	Consequences of performing or not performing an act	Persuasiveness of an appeal	Loss frames often more persuasive	”You lose X by not insulating” vs. ”You gain X by insulating”

26.6.2 Emphasis framing

In communication research, framing refers less to logically equivalent descriptions and more to *selection and emphasis*. Entman (1993) defined framing as selecting some aspects of a perceived reality and making them more salient in a communicating text so as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation. A news report about rising fuel prices can frame the issue as a problem of corporate profits, of geopolitical supply, of environmental policy, or of household budgets: each frame highlighting different causes and remedies. Entman’s four functions provide a coding scheme widely used in content analysis: does the text define a problem, diagnose causes, make a moral judgment, and suggest a remedy?

The two traditions are complementary rather than competing. Equivalence framing shows that even *logically identical* information can change choices. Emphasis framing shows that the selection of *which* true considerations to foreground shapes interpretation. Both are pervasive in business communication: in pricing (“less than a dollar a day”), in risk communication (“99% success rate”

versus "1% failure rate"), and in organisational change ("restructuring for growth" versus "cost cutting").

26.6.3 Applications in business

BUSINESS CASE

Crisis communication and framing. When a crisis threatens a firm's reputation, stakeholders ask who is responsible. **Situational crisis communication theory** (Coombs 2007) builds on attribution theory (Section 21.3) to argue that the appropriate response depends on how much responsibility stakeholders attribute to the organisation. Crises in the *victim* cluster (natural disasters, product tampering by outsiders) attract little attribution of responsibility. *Accidental* crises (technical errors) attract moderate attribution. *Preventable* crises (management misconduct, known risks that were ignored) attract strong attribution. Response strategies range from **deny** (attack the accuser, deny the crisis) through **diminish** (excuse, justification) to **rebuild** (compensation, apology). Framing is central: the firm's first communications shape which cluster stakeholders assign the crisis to. Johnson & Johnson's response to the 1982 Tylenol poisonings, a nationwide recall and the introduction of tamper-resistant packaging, despite the poisonings having been caused by an outside criminal, is widely taught as a case in which the firm framed itself as acting to protect customers rather than as a defendant. The theory also warns against mismatches: a firm that uses a "victim" frame for a crisis the public perceives as preventable invites further reputational damage.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Describing a product attribute in positive terms ("90% success") will increase evaluations relative to the equivalent negative description ("10% failure"), and the effect will be smaller for experts.
- *H2.* In crisis communication, rebuild strategies will protect reputation more than deny strategies when stakeholders attribute high responsibility to the firm (Coombs 2007).
- *H3.* News articles that frame a corporate event using a "moral" frame will generate more negative social-media sentiment than articles using an "economic" frame.

Designs. Randomised framing experiments. Content analysis of news and social-media frames (manual coding or supervised text classification). Event studies of market reactions to differently framed announcements.

Measures. Choice shares, risk preferences, evaluations, frame-coding schemes based on Entman's four functions, attributions of responsibility.

COMMON PITFALL

Calling every message difference a "frame". If two messages differ in the information they contain, not just in its presentation or emphasis, differences in outcome may reflect information rather than framing. Equivalence-framing designs require logically equivalent descriptions. Emphasis-framing designs should hold facts constant and vary only which considerations are foregrounded.

26.7 Uses and Gratifications

THEORY CARD

Core claim. Audiences are active and goal-directed: people select media and content to satisfy needs, and the effects of media depend on the gratifications people seek and obtain.
Origins. Mass-communication research, E. Katz, Blumler, and Gurevitch (1973), Blumler and Katz (1974), updated for digital media by Ruggiero (2000).
Level of analysis. Individual audience member.
Key constructs. Social and psychological needs, expectations of media, gratifications sought, gratifications obtained, media choice, functional alternatives (non-media sources of gratification).
Boundary conditions. Assumes people are aware of, and can report, their motives. Habitual and algorithmically curated media use may weaken the assumption of deliberate selection.
Related theories. Self-determination theory (Section 19.6), Microblogging, Social transmission, Technology acceptance model (Section 10.4).

26.7.1 Core logic

Most media-effects theories ask what media do to people. **Uses and gratifications** reverses the question and asks what people do with media. E. Katz, Blumler, and Gurevitch (1973) summarised the approach in a set of assumptions: the audience is active and its media use is goal-directed. The initiative in linking need gratification to media choice lies with the audience member. Media compete with other sources of need satisfaction. People are sufficiently self-aware to report their interests and motives, and value judgments about the cultural significance of media should be suspended while audience orientations are explored. The approach locates the origin of media effects in the social and psychological needs that generate expectations of media and lead to differential patterns of exposure.

Table 26.8 summarises a typology of gratifications commonly used in the literature, with contemporary business examples. The categories overlap, and the same medium can serve several of them at once.

Table 26.8. A typology of media gratifications with business examples.

Gratification	Underlying need	Traditional media example	Digital and business example
Information / surveillance	Understanding the environment, reducing uncertainty	News bulletins	Product reviews, industry newsletters
Personal identity	Reinforcing values, self-understanding	Lifestyle magazines	Following creators who express one's identity
Integration and social interaction	Belonging, conversation material	Watching the same broadcast as friends	Brand communities, group chats, live-streams
Entertainment / diversion	Escape, relaxation, emotional release	Television drama	Short-video platforms, mobile games
Self-expression and status	Being seen, building an audience	Letters to the editor	Posting reviews, microblogging

26.7.2 Evidence, extensions, and critique

The approach has proved adaptable. Each new medium, radio, television, the internet, mobile phones, social networks, has been studied by asking which needs it serves and how it competes with older media. Ruggiero (2000) argued that the interactivity, demassification, and asynchroneity of digital media make the active audience assumption more plausible than in the broadcast era, and that uses and gratifications would become more rather than less relevant. Critics have long noted that the approach can be descriptive and individualistic: lists of gratifications proliferate, self-reports of motives may be post-hoc rationalisations, and the theory says less about the content and institutional structure of media than about audience motives. Habitual use and recommendation algorithms also complicate the claim that media choice reflects deliberate goals.

DEBATE AND EVIDENCE

Active audience or captured attention? Uses and gratifications assumes that people choose media to satisfy needs. Critics of engagement-optimised platforms argue that recommendation systems shape the needs they then satisfy, so that what looks like goal-directed choice is partly the product of design. One way to reconcile the views is to distinguish **gratifications sought** from **gratifications obtained**: a platform may attract users by promising social connection (gratification sought) while delivering mainly diversion (gratification obtained). Persistent gaps between the two predict dissatisfaction and churn: a testable implication with clear managerial relevance.

BUSINESS CASE

Content strategy by gratification. A business-to-business software company publishes a blog, a podcast, and a community forum. Uses and gratifications suggests assessing each channel by the needs it serves rather than by reach alone. The blog might serve *information* needs (how-to guides), the podcast *personal identity* and *diversion* needs (listening to respected practitioners during a commute), and the forum *social interaction* and *self-expression* needs (helping peers, building a reputation). Metrics follow from the gratification: search traffic and task completion for information content, completion rates and subscriptions for the podcast, and contribution rates and reciprocal replies for the forum. Content that fails to deliver the gratification its audience seeks, a promotional podcast episode, an unanswered forum post, erodes use even if it is widely distributed.

TRY IT YOURSELF

List the three apps or media you used most yesterday. For each, write down the gratification you sought when you opened it and the gratification you actually obtained. Where they diverge, what design feature of the medium explains the gap, and what would a firm need to change to close it?

26.8 Chapter Summary

KEY TAKEAWAYS

- Persuasion depends on **how much people think**. Under high elaboration, argument quality dominates. Under low elaboration, peripheral cues and heuristics dominate, and the resulting attitudes are weaker (Petty and Cacioppo 1986, Chaiken 1980).
- Audiences recognise and cope with persuasion. The **persuasion knowledge model** explains why disclosed or recognised tactics can change meaning and lose force, and why appropriate tactics can survive disclosure (Friestad and Wright 1994, M. C. Campbell and Kirmani 2000).
- People share because sharing serves **their own functions**: impression management, emotion regulation, information, bonding, and persuasion (Berger 2014). High-arousal content travels further (Berger and Milkman 2012).
- Influence flows through **people as well as media** (E. Katz and Lazarsfeld 1955), but large cascades depend on network susceptibility more than on a few influentials (D. J. Watts and Dodds 2007).
- **Undirected communication** such as microblogging lets people reach out without imposing on anyone in particular (Buechel and Berger 2018).
- Falsehood spreads farther and faster than truth, driven by human sharing of novel, arousing content (Vosoughi, Roy, and Aral 2018).
- Media shape **what people think about** (agenda setting) and which criteria they use (priming) (McCombs and Shaw 1972, S. Iyengar and Kinder 1987). **Framing** shapes how issues and choices are interpreted (Tversky and Kahneman 1981, Entman 1993).
- **Uses and gratifications** treats audiences as goal-directed and explains media choice by the needs it serves (E. Katz, Blumler, and Gurevitch 1973).

Table 26.9 presents the comparison of the chapter's theories.

Table 26.9. Comparison of the chapter's theories.

Theory	Focal actor	Key mechanism	Signature prediction	Typical business use
ELM / HSM	Message recipient	Elaboration (motivation x ability)	Arguments matter under high, cues under low involvement	Creative strategy by category involvement
Persuasion knowledge model	Target of persuasion	Recognition of persuasive intent, change of meaning	Recognised inappropriate tactics backfire	Influencer disclosure, sales tactics
Social transmission	Sharer	Self-serving functions of sharing, arousal	High-arousal, socially valuable content spreads	Shareable content, referral design
Two-step flow	Opinion leader	Interpersonal relay of media content	Influence passes through social ties	Seeding and advocacy programmes
Undirected communication	Poster	Reaching out without imposition	Socially apprehensive people favour broadcasting	Community and platform design
Agenda setting / priming	Public	Accessibility of issues and criteria	Media emphasis predicts issue importance	Issue management, PR timing
Framing	Decision maker / public	Reference points, applicability of schemas	Equivalent descriptions change choices	Pricing, risk, and crisis communication

Table 26.9. Comparison of the chapter's theories. *(continued)*

Theory	Focal actor	Key mechanism	Signature prediction	Typical business use
Uses and gratifications	Audience member	Need-driven media selection	Media chosen for gratifications sought	Channel strategy and content metrics

26.9 Discussion Questions

1. The ELM predicts that argument-based attitudes are more durable than cue-based attitudes. Design a field experiment for a consumer brand that would test this prediction using purchase data rather than self-reported attitudes.
2. A regulator proposes that all influencer posts carry a prominent "Advertisement" label. Using the persuasion knowledge model, predict the effects on followers, influencers, and brands, and identify the conditions under which disclosure would *not* reduce persuasion.
3. Vosoughi, Roy, and Aral (2018) found that falsehood spreads farther than truth. Which functions of word of mouth help explain this pattern, and what platform design changes would follow from each?
4. Agenda setting was first documented in an era of a few dominant news outlets. How would you test the theory in a fragmented, algorithmically curated media environment? What is "the media agenda" in that setting?
5. Distinguish equivalence framing from emphasis framing. Give one business example of each, and explain how you would design an experiment that isolates the framing effect from an information effect.
6. Choose a product category. Using uses and gratifications, identify the dominant gratification sought by its customers and propose a content strategy that serves it. How would you measure whether the gratification is obtained?

26.10 Further Reading

- Petty and Cacioppo (1986): the definitive statement of the elaboration likelihood model, its postulates, and the early evidence.
- Petty, Cacioppo, and Schumann (1983): the canonical advertising experiment showing the involvement × argument quality × peripheral cue pattern.
- Friestad and Wright (1994): the original persuasion knowledge model. Essential for research on disclosure, covert marketing, and sales tactics.
- Berger (2014): a comprehensive review of word of mouth organised around the functions sharing serves.
- Berger and Milkman (2012): an influential empirical study of which content goes viral and why arousal matters.
- Vosoughi, Roy, and Aral (2018): a large-scale analysis of how true and false news spread online.
- Scheufele and Tewksbury (2007): a clear conceptual comparison of agenda setting, priming, and framing.
- E. Katz, Blumler, and Gurevitch (1973): the foundational statement of the uses and gratifications approach.

Part V

Consumers and Marketing

Chapter 27

Consumer Behavior and Branding

In April 1985 the Coca-Cola Company replaced the formula of its flagship drink with a sweeter recipe that had beaten both the original and Pepsi in extensive blind taste tests. The launch of "New Coke" provoked letters, protests, and a wave of press coverage, and within three months the company brought the original back as Coca-Cola Classic. Consumers had rejected a product they preferred when they could not see the label. The episode is the most famous demonstration of a point this chapter develops at length: what consumers buy is rarely just a bundle of physical attributes. They buy expectations that can be met or disappointed, chains of meaning that link attributes to personal values, memories of a brand, a partner in a relationship, a mirror for the self, a licence to indulge, a feeling of "mine", a reference point against which a price feels fair, and a share in a culture.

Similar puzzles appear across categories. Why do some owners of Harley-Davidson motorcycles tattoo the company's logo on their skin, while few owners of equally reliable machines feel any such loyalty? Why did Porsche's move into sport utility vehicles succeed while Bic's move from pens and lighters into perfume did not? Why do shoppers at outlet malls respond to "compare at" prices that they may not fully believe, and why did a large US department store chain lose customers when it abolished the discounts that its shoppers claimed to dislike? Why does touching a mug in a store make it feel more valuable, and why do people choose less-preferred options "for the sake of variety"?

This chapter introduces twelve theories that consumer researchers and brand managers use to answer such questions. We begin with the cognitive core of post-purchase evaluation (**expectancy-disconfirmation**) and the motivational logic linking product features to personal values (**means-end chain theory**). We then turn to the brand: how it lives in consumer memory (**customer-based brand equity**), how it acquires human characteristics (**brand personality**), how consumers form relationships with it (**consumer-brand relationships**), and how it expresses the self (**self-congruity**). Four theories address how consumers evaluate and choose: **schema congruity** (new products and brand extensions), the **hedonic-utilitarian** distinction, **psychological ownership**, and **variety seeking**. Two final theories widen the lens to price perception (**reference price**) and to the cultural meanings of consumption (**consumer culture theory**). Table 27.1 summarises the territory.

Table 27.1. Chapter at a glance: theories of consumer behavior and branding.

Theory	Core question	Level of analysis	Key source
Expectancy-disconfirmation	How do consumers judge whether they are satisfied?	Individual	Oliver (1980)
Means-end chain	How do product attributes connect to personal values?	Individual	Gutman (1982)
Customer-based brand equity	Why does the same marketing work better for some brands?	Individual / brand	Keller (1993), Aaker (1991)
Brand personality	Which human traits do consumers ascribe to brands?	Brand	Aaker (1997)
Consumer-brand relationships	Can consumers have relationships with brands, and of what kind?	Dyad (consumer-brand)	Fournier (1998)
Self-congruity	Why do consumers prefer brands that match who they are?	Individual	Sirgy (1982), Belk (1988)
Schema congruity	Why are moderately unexpected products evaluated favourably?	Individual	Meyers-Levy & Tybout (1989)
Hedonic and utilitarian consumption	How do pleasure and function shape choice and justification?	Individual	Hirschman & Holbrook (1982), Dhar & Wertenbroch (2000)
Psychological ownership	Why do people value what feels like theirs?	Individual	Pierce, Kostova & Dirks (2001)
Variety seeking	Why do consumers switch away from a favourite?	Individual	McAlister & Pessemier (1982)
Reference price	Against what standard do consumers judge a price?	Individual / market	Winer (1986), Mazumdar et al. (2005)
Consumer culture theory	How do consumers use goods and brands to make cultural meaning?	Individual / community / society	Arnould & Thompson (2005)

Three organising ideas recur. First, consumers evaluate offerings against **standards** (expectations, schemas, reference prices, the self), so the same objective product can delight one person and disappoint another. Second, brands are **knowledge structures** in memory that carry functional, symbolic, and emotional meaning, and that meaning is co-authored by firms, consumers, and the wider culture. Third, many of the most consequential effects in this chapter are **asymmetric**: losses loom larger than gains, transgressions matter more for some brands than others, and extensions can dilute as well as leverage. The chapter connects to Prospect theory (Chapter 24), the Endowment effect (Section 23.4), Social identity theory (Section 22.2), and the Elaboration likelihood model (Section 26.1), and it prepares the ground for the firm-level theories of the next chapter, such as Market orientation (Section 28.1) and Customer equity (Section 28.4).

27.1 Expectancy-Disconfirmation Theory

THEORY CARD

Core claim. Satisfaction results from comparing perceived performance with prior expectations. Performance that exceeds expectations (positive disconfirmation) raises satisfaction, and performance that falls short (negative disconfirmation) lowers it, over and above the levels of expectations and performance themselves.
Origins. Consumer psychology and marketing, building on adaptation-level and assimilation-contrast ideas (R. E. Anderson 1973). Formalised by R. L. Oliver (1980) and tested against rival specifications by Churchill and Surprenant (1982). Extended to firm-level data by E. W. Anderson and Sullivan (1993).
Level of analysis. Individual consumer, aggregated to firms, industries, and national indices.
Key constructs. Expectations (predictive standards), perceived performance, disconfirmation (subjective comparison), satisfaction, repurchase intention, complaining, word of mouth (Section 26.3).
Boundary conditions. Expectations must exist before consumption and performance must be evaluable. For complex durables or credence services, performance may dominate and disconfirmation matters less. Asymmetries are expected: shortfalls hurt more than equivalent surpluses help.
Related theories. Prospect theory (Chapter 24), Hedonic adaptation (Section 25.2), Service quality gap model (Section 28.3), Customer equity (Section 28.4), Cognitive dissonance theory (Section 21.1).

27.1.1 Origins and intellectual history

Early research on consumer satisfaction borrowed two ideas from psychology. The first was **assimilation**: consumers who expect a product to be good may perceive it as better than it is, pulling perceptions toward expectations. The second was **contrast**: a product that falls short may be perceived as worse than it is because the gap is magnified. R. E. Anderson (1973) set these rival predictions against each other experimentally and helped establish disconfirmed expectations as a central variable in satisfaction research.

R. L. Oliver (1980) reorganised the field around a simple but powerful structure. Expectations create a frame of reference. After consumption, the consumer judges whether performance was better or worse than expected. That **disconfirmation** judgment, together with the expectation level, determines satisfaction, and satisfaction in turn updates attitudes and purchase intentions. Satisfaction is therefore not a property of the product. It is a **relative judgment** that depends on what the consumer brought to the encounter.

27.1.2 Core logic and propositions

The theory involves four constructs and two paths to satisfaction: a direct effect of expectations (which anchor judgments) and an effect of disconfirmation (which captures surprise relative to that anchor).

- **P1 (Disconfirmation).** Holding performance constant, satisfaction increases with positive disconfirmation and decreases with negative disconfirmation (R. L. Oliver 1980).

- **P2 (Performance).** Perceived performance affects satisfaction both directly and through disconfirmation. For some products, the direct path dominates: Churchill and Surprenant (1982) found that satisfaction with a durable product was driven by performance alone, whereas expectations and disconfirmation also mattered for a nondurable product.
- **P3 (Asymmetry).** Negative disconfirmation has a larger effect on satisfaction than positive disconfirmation of the same size (E. W. Anderson and Sullivan 1993).
- **P4 (Evaluability).** Disconfirmation is more likely to be perceived when quality is easy to evaluate (E. W. Anderson and Sullivan 1993).
- **P5 (Consequences).** Satisfaction raises repurchase intention, but the link between stated satisfaction and actual repurchase is nonlinear and differs across customer groups (Mittal and Kamakura 2001) (Figure 27.1).

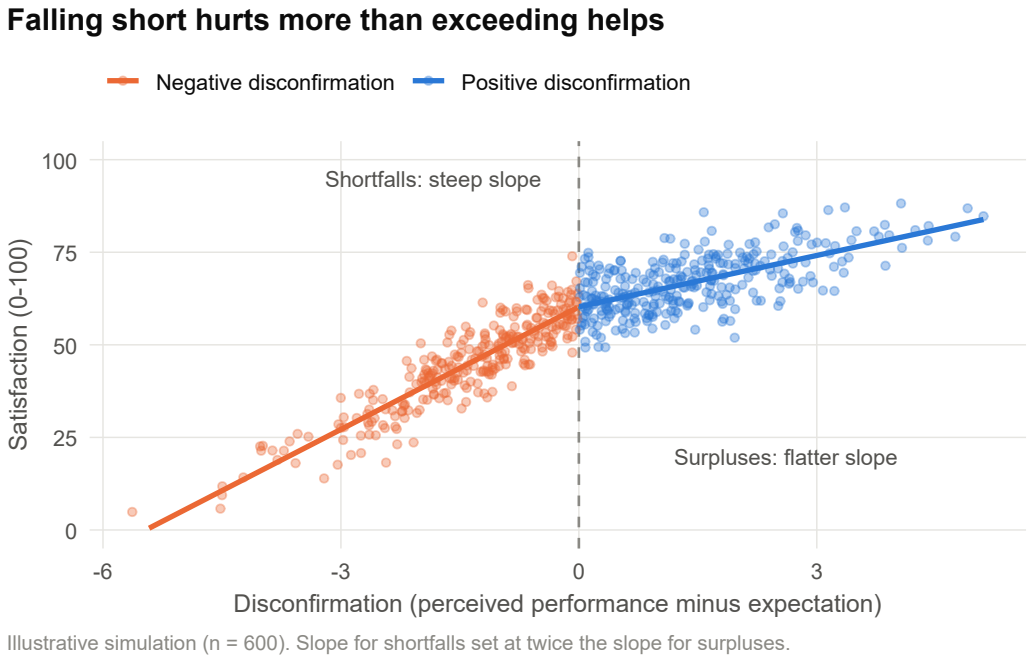


Figure 27.1. Stylised expectancy-disconfirmation relationship. Each point is a simulated customer. Satisfaction rises with disconfirmation (perceived performance minus expectations), but the slope is steeper for shortfalls than for surpluses, the asymmetry reported by Anderson and Sullivan (1993). Illustrative data, not empirical estimates.

27.1.3 Mechanism

Figure 27.2 shows the causal structure. The dashed edge from expectations to perceived performance represents assimilation: high expectations can colour the experience itself. The dotted moderator captures evaluability, which governs how much the comparison process, rather than raw performance, drives the judgment.

27.1.4 Evidence and extensions

The model has been tested in hundreds of studies. A meta-analysis by Szymanski and Henard (2001) found that disconfirmation and perceived fairness (equity) are among the strongest correlates of satisfaction, while the direct effect of expectations is weaker and more variable. Spreng,

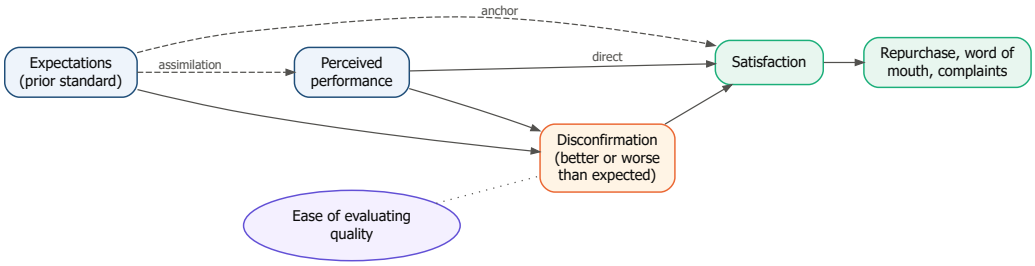


Figure 27.2. Mechanism of expectancy-disconfirmation. Expectations and perceived performance feed a subjective comparison (disconfirmation), which shapes satisfaction and downstream loyalty. Based on Oliver (1980), Churchill and Surprenant (1982), and Anderson and Sullivan (1993).

MacKenzie, and Olshavsky (1996) argued that expectations are not the only relevant standard: consumers also compare performance with their **desires**, and desire congruency adds explanatory power. At the firm level, E. W. Anderson and Sullivan (1993) used Swedish national survey data to show that perceived quality and disconfirmation drive satisfaction and that negative disconfirmation weighs more heavily, an asymmetry that echoes Loss aversion (Section 24.4). Satisfaction research also scaled up into national indices: the **American Customer Satisfaction Index** (ACSI) models satisfaction as the outcome of customer expectations, perceived quality, and perceived value, with complaints and loyalty as consequences (Fornell et al. 1996). Fornell et al. (2006) reported that a portfolio of firms with high ACSI scores outperformed the market, which suggested that satisfaction is an under-priced intangible asset.

DEBATE AND EVIDENCE

Is satisfaction a comparison or a feeling? Critics argue that the disconfirmation paradigm is too cognitive. Satisfaction also has affective roots: R. L. Oliver (1993) showed that positive and negative affect experienced during consumption contribute to satisfaction alongside disconfirmation and attribute-level judgments. A second debate concerns measurement: **inferred** disconfirmation (performance minus expectation scores) and **perceived** disconfirmation ("better or worse than expected") are not equivalent, and the perceived version is contaminated by the satisfaction judgment it is supposed to explain. A third debate concerns financial consequences. Jacobson and Mizik (2009) re-examined the claim that satisfaction predicts abnormal stock returns and found the effect concentrated in a small set of firms, mostly in computer and internet industries, which raised doubts about a general market mispricing. The exchange is a useful reminder that satisfaction-performance links depend on industry, time window, and model specification.

27.1.5 Applications in business

Expectancy-disconfirmation gives managers two levers rather than one. Firms can raise performance, but they can also **manage expectations**: promise what they can reliably deliver, then exceed it. Overpromising in advertising may win the first purchase but sets up negative disconfirmation, the costliest outcome in the asymmetric model.

BUSINESS CASE

Customer satisfaction measurement: the ACSI. The American Customer Satisfaction Index, launched in 1994 at the University of Michigan, surveys customers of companies across the US economy and estimates a structural model in which expectations, perceived quality, and perceived value drive satisfaction, which in turn drives complaints and loyalty (Fornell et al. 1996). Its design embodies the theory: it does not ask only "how satisfied are you" but measures the standards against which satisfaction is formed. Firms use ACSI benchmarks to compare themselves with competitors and to diagnose whether a fall in satisfaction reflects rising expectations or declining quality. The contrast with single-question loyalty metrics is instructive. The Net Promoter Score asks only about likelihood to recommend (Reichheld 2003), and Keiningham et al. (2007) did not find that it outperformed the ACSI in predicting revenue growth.

Table 27.2 presents the approaches to measuring customer satisfaction and their relation to expectancy-disconfirmation theory.

Table 27.2. Approaches to measuring customer satisfaction and their relation to expectancy-disconfirmation theory.

Approach	What is measured	Link to the theory	Main limitation
Inferred disconfirmation	Separate expectation and performance ratings, difference computed	Operationalises the comparison explicitly	Difference scores are unreliable
Perceived disconfirmation	Direct rating: better or worse than expected	Captures the subjective comparison	Overlaps with the satisfaction judgment
ACSI structural model	Expectations, quality, value, satisfaction, complaints, loyalty	Embeds satisfaction in its antecedents and consequences	Survey cost, annual cadence
Net Promoter Score	Likelihood to recommend (0-10)	Outcome only, no standard of comparison	Weak diagnostic value
Behavioural metrics	Repurchase, churn, complaints, returns	Tests the satisfaction-loyalty link	Confounded by switching costs

BUSINESS CASE

Illustrative scenario: launch messaging for a meal-kit service. A start-up tests two launch messages. Version A promises "restaurant quality in 20 minutes". Version B promises "a good weeknight dinner in about 30 minutes". Version A lifts trial, but customers report recipes taking 35-40 minutes. Under expectancy-disconfirmation, Version A converts a performance that is adequate in absolute terms into a shortfall. The theory predicts higher first-box churn for Version A, and an A/B test that tracks second-order rates rather than sign-ups is the appropriate check.

27.1.6 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Holding perceived performance constant, customers exposed to higher-expectation advertising will report lower satisfaction.

- *H2.* The effect of negative disconfirmation on churn will be larger than the effect of positive disconfirmation of equal magnitude.
- *H3.* The effect of disconfirmation on satisfaction will be weaker for credence services (e.g., financial advice) than for search or experience goods.

Established measures. Oliver's satisfaction items (R. L. Oliver 1980), perceived disconfirmation scales ("much worse than expected" to "much better than expected"), and the ACSI model and survey items (Fornell et al. 1996).

Designs. Experiments that manipulate expectations (advertising claims) and performance (product quality) orthogonally. Panel surveys linking satisfaction to observed repurchase (Mittal and Kamakura 2001). Firm-level panels using ACSI data with financial outcomes (Fornell et al. 2006, Jacobson and Mizik 2009).

Data sources. ACSI firm scores, customer review text (with star ratings as satisfaction and review content as disconfirmation cues), CRM churn data.

COMMON PITFALL

Treating satisfaction as absolute. A satisfaction score is relative to a standard. Comparing scores across segments, brands, or time without asking whether expectations moved can produce wrong conclusions: a luxury hotel and a budget motel can earn identical scores for very different performance. A related error is to measure disconfirmation and satisfaction in the same survey block and then report a large correlation as support for the theory. Separate the measures in time or method, or manipulate expectations directly.

27.2 Means-End Chain Theory

THEORY CARD

Core claim. Consumers value product attributes not for their own sake but as means to desired consequences, which in turn serve personal values. Product knowledge is organised as chains linking **attributes**, **consequences**, and **values**.

Origins. Consumer research. Gutman (1982), drawing on value theory and categorisation research. The **laddering** interview method was systematised by Reynolds and Gutman (1988). Zeithaml (1988) applied the logic to price, quality, and value.

Level of analysis. Individual consumer, aggregated to segments through hierarchical value maps.

Key constructs. Concrete and abstract attributes, functional and psychosocial consequences, instrumental and terminal values, ladders, implication matrix, hierarchical value map.

Boundary conditions. Most informative for involving products where consumers can articulate reasons. Less useful for habitual, low-involvement purchases where chains are short or absent.

Related theories. Schwartz values (Section 18.6), Expectancy theory (Section 19.2), Self-congruity, Hedonic and utilitarian consumption.

27.2.1 Core logic

Gutman (1982) proposed that consumers choose actions that produce desired consequences and minimise undesired ones, and that values give consequences their positive or negative valence. Product attributes matter because they are associated, in the consumer's mind, with consequences that are themselves linked to values. The result is a cognitive chain: *attribute* (what the product has) leads to *consequence* (what it does for me) leads to *value* (why that matters to me).

Figure 27.3 shows an illustrative ladder for a running shoe. The same attribute can lead to different values for different consumers, which is why means-end analysis is a segmentation tool as much as a theory of motivation.

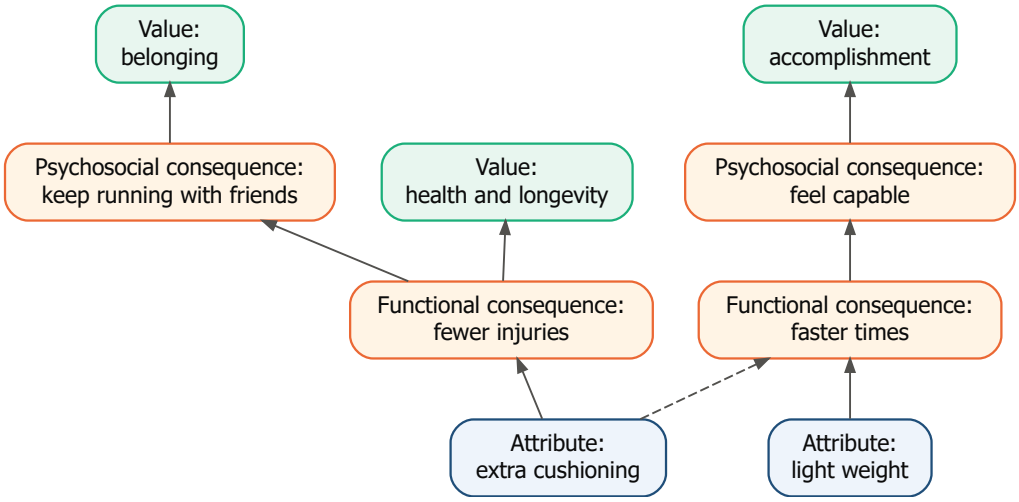


Figure 27.3. Illustrative means-end ladders for a running shoe. Two consumers start from the same attribute (cushioning) but reach different terminal values through different consequences. Laddering interviews elicit such chains by repeatedly asking why an element is important. Illustrative example, not interview data.

27.2.2 Laddering as method

Reynolds and Gutman (1988) described **laddering** as a semi-structured interview that begins with a distinction the consumer draws between products (often elicited by sorting or preference comparisons) and then repeatedly asks "why is that important to you?" until the respondent reaches a value or cannot go further. Responses are content-coded into attributes, consequences, and values. An **implication matrix** counts how often each element leads to every other, and a **hierarchical value map** (HVM) draws the dominant links above a cut-off. Table 27.3 summarises the levels.

Table 27.3. Levels of a means-end chain with an illustrative coffee ladder. Example responses are hypothetical.

Level	Definition	Typical laddering prompt	Coffee example	Marketing use
Concrete attribute	Physical, observable feature	What makes this one different?	Single-origin beans	Product specification

Table 27.3. Levels of a means-end chain with an illustrative coffee ladder. Example responses are hypothetical. *(continued)*

Level	Definition	Typical laddering prompt	Coffee example	Marketing use
Abstract attribute	Subjective characterisation of features	What does that mean for the product?	Rich flavour	Positioning claims
Functional consequence	Direct, tangible outcome of use	Why is that important to you?	Enjoyable morning ritual	Benefit copy
Psychosocial consequence	Social or psychological outcome	And why does that matter?	Moment to myself before work	Emotional appeals
Instrumental value	Preferred mode of conduct	Why is that important to you?	Being in control of my day	Brand character
Terminal value	Desired end state	And why is that important?	Inner harmony, self-respect	Leverage point for advertising

27.2.3 Evidence and critique

Means-end analysis has been used to study food, financial services, tourism, and advertising, and it shaped how agencies translate product features into value-based communication. Zeithaml (1988) used a means-end framework to show that consumers move from intrinsic and extrinsic cues to perceived quality, perceived sacrifice, and perceived value, a structure still used in pricing research.

DEBATE AND EVIDENCE

Are ladders discovered or constructed? Grunert and Grunert (1995) questioned whether laddering reveals stable cognitive structures or prompts respondents to construct plausible chains on the spot. "Why" probing may push people to higher abstraction than their actual decision process involves. Hard laddering (questionnaire formats) and the **association pattern technique**, which asks respondents to tick links in predefined attribute-consequence and consequence-value matrices, offer more scalable and comparable alternatives (ter Hofstede et al. 1998). The cost is that predefined lists can miss idiosyncratic meanings that soft interviews reveal.

BUSINESS CASE

From attribute claims to value platforms. Many mature categories compete on attribute parity: most toothpastes fight cavities, most cars are safe. Means-end analysis helps brand managers find the level of the ladder where differentiation is still possible. A car maker's safety engineering (attribute) can lead to "protecting my family" (psychosocial consequence) and to "being a responsible parent" (value). Volvo's long association with safety is a familiar public example of a brand that built its positioning around one ladder and kept it consistent across decades. The research task is to verify that the target segment's ladder actually runs through the claimed attribute, rather than assuming it.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Advertising that connects an attribute to a terminal value from the segment's dominant ladder will produce more favourable brand attitudes than advertising that stops at the functional consequence.
- *H2.* Ladders will be longer (reach more abstract levels) for high-involvement than for low-involvement categories.

Measures and designs. Soft laddering interviews coded into an implication matrix and HVM (Reynolds and Gutman 1988). The association pattern technique for large samples and cross-national segmentation (ter Hofstede et al. 1998). Value inventories such as the Schwartz value (Section 18.6) survey to validate the value level.

27.3 Customer-Based Brand Equity

THEORY CARD

Core claim. A brand has positive equity when consumers respond more favourably to its marketing (its product, price, promotion, or distribution) than they would to the same marketing for an unnamed or fictitious version. That differential response arises from **brand knowledge** in memory: awareness of the brand and the favourability, strength, and uniqueness of its associations (Keller 1993).

Origins. Marketing. D. A. Aaker (1991) framed brand equity as a set of assets and liabilities linked to a brand name. Keller (1993) developed the consumer-psychological account grounded in associative memory. Keller and Lehmann (2006) review the research programme.

Level of analysis. Individual consumer (brand knowledge), aggregated to the brand and the firm (price premiums, market share, firm value).

Key constructs. Brand awareness (recall, recognition), brand image, brand associations (type, favourability, strength, uniqueness), perceived quality, brand loyalty, differential response.

Boundary conditions. Equity is defined relative to a comparison (an unbranded or fictitious product) and to a consumer segment. Associations that are favourable in one segment or market may be neutral or negative in another.

Related theories. Signaling theory (Section 8.2), Market-based assets (Section 4.5), Brand personality, Schema congruity, Mere exposure effect (Section 29.1).

27.3.1 Origins and intellectual history

Brand equity became a managerial priority in the late 1980s, when a wave of acquisitions placed large financial values on brand names and practitioners asked what, exactly, they were paying for. D. A. Aaker (1991) answered with an asset view: brand equity comprises **brand loyalty**, **name awareness**, **perceived quality**, **brand associations**, and other proprietary assets such as trademarks and channel relationships. Each adds or subtracts value for the customer and the firm.

Keller (1993) moved the definition inside the consumer's head. Building on associative network models of memory, he defined **customer-based brand equity** as the differential effect that brand knowledge has on how consumers respond to the brand's marketing. Brand knowledge has two components. **Brand awareness** is the ability to recall or recognise the brand. **Brand image** is the

set of associations linked to the brand node in memory. The definition's power lies in its comparative structure: equity is not how much people like a brand, but how much *more* they respond to the same marketing action because the brand name is attached.

27.3.2 Core logic and propositions

- **P1 (Awareness as a threshold).** Brand awareness is necessary for equity: a brand must be in the consideration set before its associations can matter.
- **P2 (Image).** Given awareness, equity increases with the **favourability**, **strength**, and **uniqueness** of brand associations (Keller 1993).
- **P3 (Differential response).** High-equity brands enjoy higher willingness to pay, lower price elasticity, greater advertising and promotion effectiveness, and more favourable evaluations of brand extensions.
- **P4 (Credibility).** Brand investments make brand promises credible, which reduces perceived risk and information costs. Erdem and Swait (1998) formalise this as a signaling (Section 8.2) account of brand equity (Figure 27.4).

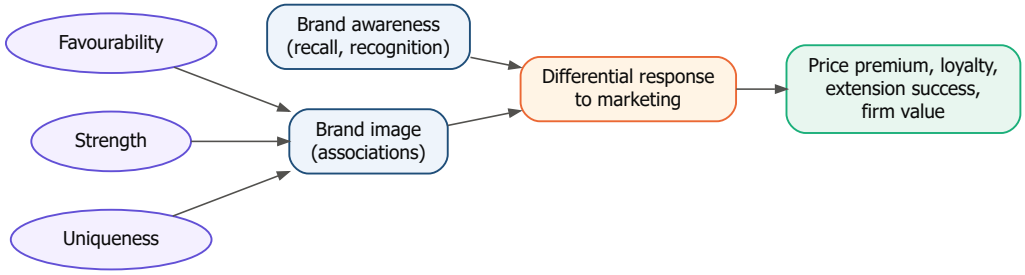


Figure 27.4. Customer-based brand equity as a memory structure. Awareness and the characteristics of brand associations jointly produce a differential response to marketing, which appears in market and financial outcomes. Based on Keller (1993) and Keller and Lehmann (2006).

Figure 27.5 translates P3 into a stylised demand comparison. The branded and unbranded products are physically identical. The branded product sells more at every price and loses fewer buyers when the price rises: brand equity shows up as both a level shift and a flatter demand curve.

27.3.3 Evidence and extensions

Evidence comes from three levels, summarised by Keller and Lehmann (2006) and in Table 27.4. At the **customer-mindset** level, survey scales measure awareness, associations, perceived quality, and loyalty (Yoo and Donthu 2001). At the **product-market** level, Ailawadi, Lehmann, and Neslin (2003) proposed the **revenue premium**, the extra revenue a brand earns relative to a private label, as an outcome measure that is objective, comparable across brands, and sensitive to changes in brand health. At the **financial-market** level, researchers relate brand metrics to firm value. Brand equity also predicts the success of brand extensions (D. A. Aaker and Keller 1990, Völckner and Sattler 2006), a topic we return to under schema congruity.

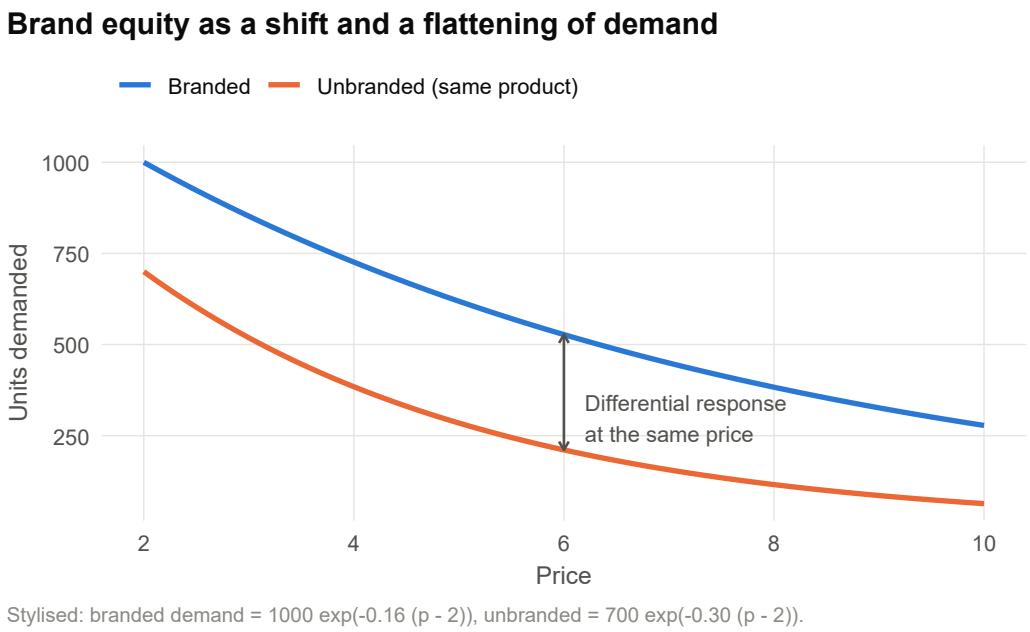


Figure 27.5. Stylised differential response. The same product sold under a strong brand name and without a brand faces different demand curves: more buyers at every price and lower price sensitivity. The vertical gap at a common price corresponds to the revenue premium measure of Ailawadi, Lehmann and Neslin (2003). Illustrative curves, not estimates.

Table 27.4. Three levels at which brand equity is measured, plus Aaker's asset view. Organised following Keller and Lehmann (2006).

Level	What is measured	Example measures	Strength	Weakness
Customer mindset	Awareness, associations, perceived quality, loyalty	Multidimensional brand equity scale (Yoo & Donthu, 2001), association elicitation	Diagnostic: shows why equity is high or low	Does not directly show market value
Product market	Price and volume premium over a benchmark	Revenue premium versus private label (Ailawadi et al., 2003), conjoint brand part-worths	Objective, comparable, tracks change	Needs a credible benchmark product
Financial market	Contribution of the brand to firm value	Event studies, brand value rankings	Speaks the language of finance	Hard to isolate the brand from other intangibles
Asset view	Loyalty, awareness, perceived quality, associations, proprietary assets	Aaker (1991) brand equity dimensions	Managerially complete	Mixes causes and outcomes

DEBATE AND EVIDENCE

Memory, signal, or market asset? The associative-memory account (Keller 1993) and the signaling account (Erdem and Swait 1998) make overlapping predictions but differ in mechanism. Memory accounts emphasise what consumers know and feel about a brand. Signaling accounts emphasise that brand investments are a bond the firm forfeits if it cheats, so eq-

uity should be greatest where quality is hard to observe. A second critique targets measurement proliferation: customer-mindset scales can be high while market outcomes stagnate, and many studies validate a scale against attitudes measured in the same survey. Linking mindset metrics to observed market behaviour remains the field's central measurement challenge (Keller and Lehmann 2006).

27.3.4 Applications in business

BUSINESS CASE

New Coke and the location of brand equity. Coca-Cola's 1985 reformulation was supported by blind taste tests in which the new formula was preferred. What the tests did not measure was the brand knowledge that consumers attached to the original product: associations with tradition, personal memories, and American identity. Replacing the formula threatened those associations, and the company restored the original under the Coca-Cola Classic name within months. In CBBE terms, the taste tests measured the response to the *unbranded* product, while the market responded to the *branded* product. The difference between the two is precisely what Keller's definition calls brand equity.

BUSINESS CASE

Private labels and the revenue premium. Retailers' own brands compete directly on the dimension where manufacturer brands claim superiority. Tracking the revenue premium of a national brand over the store brand in the same category, as Ailawadi, Lehmann, and Neslin (2003) propose, gives brand managers an early warning signal: a shrinking premium indicates eroding equity even when unit share is stable because of promotions.

27.3.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The price elasticity of demand will be lower for brands with stronger and more unique associations, controlling for perceived quality.
- *H2.* The positive effect of brand equity on extension evaluations will be stronger when perceived fit is high.
- *H3.* The effect of brand equity on choice will be larger for experience and credence goods than for search goods (signaling account).

Established scales and measures. Multidimensional brand equity scale (Yoo and Donthu 2001). Aaker's dimensions (D. A. Aaker 1991). Free association tasks coded for favourability, strength, and uniqueness. Revenue premium from scanner data (Ailawadi, Lehmann, and Neslin 2003).

Designs. Branded versus unbranded (or fictitious-brand) experiments. Choice models estimating brand intercepts and price coefficients. Panel data linking brand-tracking surveys to sales.

COMMON PITFALL

Measuring equity without the "differential". Many studies label brand attitude, awareness, or advertising spend "brand equity". Keller's construct is comparative: it requires a benchmark (an unbranded, fictitious, or private-label version) and a response to marketing. Without the comparison, a study measures liking or awareness, not equity.

27.4 Brand Personality

THEORY CARD

Core claim. Consumers ascribe human personality traits to brands. In the United States these perceptions are organised along five dimensions: **sincerity**, **excitement**, **competence**, **sophistication**, and **ruggedness** (J. L. Aaker 1997).
Origins. Consumer psychology, adapting trait psychology and the Big Five lexical approach to brands (J. L. Aaker 1997).
Level of analysis. The brand, as perceived by consumers.
Key constructs. Brand personality dimensions and facets, brand-user imagery, self-expression.
Boundary conditions. The five-factor structure was derived from ratings aggregated across brands from many categories. It does not necessarily hold for individual brands within one category or in other cultures.
Related theories. Self-congruity, Consumer-brand relationships, Customer-based brand equity, Identity signaling (Section 8.5).

27.4.1 Core logic

J. L. Aaker (1997) defined **brand personality** as the set of human characteristics associated with a brand. Personality traits can be transferred to a brand directly, through the people associated with it (typical users, employees, endorsers), and indirectly, through product attributes, logo, advertising style, and price. Brand personality serves a **self-expressive** function: consumers can use brands to express their actual or ideal selves, which links the theory to self-congruity.

Aaker developed her scale by having a large US sample rate well-known brands on a broad set of traits and then reducing the ratings by factor analysis. Three of the resulting dimensions (sincerity, excitement, competence) resemble Big Five human personality factors, while sophistication and ruggedness have no close human counterpart and appear to capture aspirational qualities that consumers want but do not necessarily have. Table 27.5 lists the dimensions and their facets.

Table 27.5. Aaker's (1997) five dimensions of brand personality and their facets.

Dimension	Facets	Human personality analogue	Typical positioning
Sincerity	Down-to-earth, honest, wholesome, cheerful	Agreeableness	Family, trust, care
Excitement	Daring, spirited, imaginative, up-to-date	Extraversion	Youth, novelty, energy
Competence	Reliable, intelligent, successful	Conscientiousness	Expertise, dependability
Sophistication	Upper class, charming	No close analogue	Luxury, refinement

Table 27.5. Aaker's (1997) five dimensions of brand personality and their facets. *(continued)*

Dimension	Facets	Human personality analogue	Typical positioning
Ruggedness	Outdoorsy, tough	No close analogue	Strength, outdoors, masculinity

Figure 27.6 shows the illustrative brand personality profiles for two hypothetical brands in the same category, measured on Aaker's five dimensions (1-5 scale).

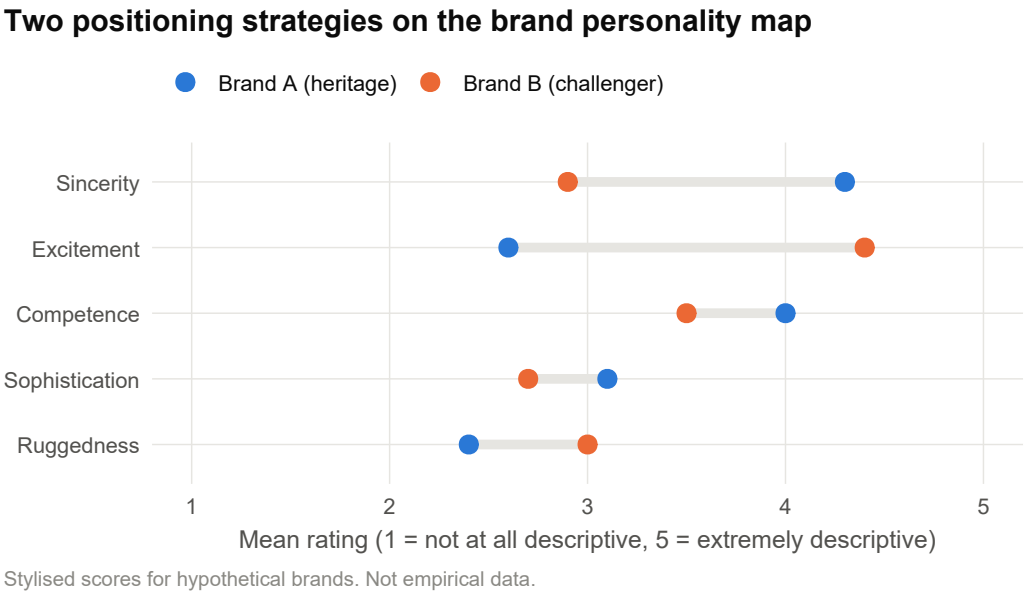


Figure 27.6. Illustrative brand personality profiles for two hypothetical brands in the same category, measured on Aaker's five dimensions (1-5 scale). Profile comparisons like this guide positioning, but within-category differences may not reproduce the five-factor structure. Stylised data.

27.4.2 Evidence and critique

Brand personality has been linked to brand attitudes, attachment, and loyalty, and it moderates how consumers respond to brand transgressions: relationships with **sincere** brands can be more fragile after a transgression than relationships with **exciting** brands (J. Aaker, Fournier, and Brasel 2004), which we discuss in the next section. Malär et al. (2011) show that a brand personality that matches the consumer's actual self strengthens emotional attachment.

DEBATE AND EVIDENCE

Does the scale measure personality, and does it generalise? Three critiques dominate. First, A. Azoulay and Kapferer (2003) argue that the scale's definition is too loose: items such as "upper class", "intelligent", or "successful" describe social status, competence, or user imagery rather than personality in the psychological sense. Second, Austin, Siguaw, and Mattila (2003) found that the five-factor framework did not replicate when applied to individual brands within a single product category, which is exactly the setting in which brand managers

want to use it. Third, the structure varies across cultures. In response, Geuens, Weijters, and De Wulf (2009) developed a shorter measure restricted to personality traits that performs better at the level of individual brands and across categories. Researchers should choose between the scales according to the comparison they need to make.

BUSINESS CASE

Personifying platforms: Apple's "Get a Mac" campaign. Between 2006 and 2009 Apple ran a long series of television spots in which two actors personified the Mac (casual, relaxed, creative) and the PC (formal, fussy, prone to problems). The campaign is an unusually literal application of brand personality: it assigned excitement and sincerity to one brand and a caricature of bureaucratic competence to its rival, and it invited viewers to choose the character they would rather be. It also illustrates the risk of personality positioning: a personality defined against a rival depends on consumers accepting the caricature, and a competitor can respond by reclaiming the traits, as Microsoft's own counter-campaign attempted.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Congruence between a brand's personality and the consumer's actual self will predict emotional brand attachment more strongly than congruence with the ideal self, especially for consumers low in self-esteem.
- *H2.* An advertising style manipulation (spontaneous versus formal) will shift perceived excitement but not perceived competence.

Established scales. Aaker's 42-item brand personality scale (J. L. Aaker 1997). The 12-item scale of Geuens, Weijters, and De Wulf (2009) for within-category comparisons.

Designs. Repeated-measures ratings of multiple brands with multilevel factor analysis (separating between-brand and within-respondent variance). Experiments manipulating personality cues in advertising or packaging.

COMMON PITFALL

Using the scale off the shelf for within-category positioning. Aaker's dimensions were derived across categories. Applying them to compare three banks or four running-shoe brands, and then interpreting each dimension as if the structure held, invites the problem documented by Austin, Siguaw, and Mattila (2003). Test the factor structure in your own data, consider the Geuens, Weijters, and De Wulf (2009) alternative, and do not treat "competence" ratings of a bank as a statement about its personality rather than its perceived performance.

27.5 Consumer-Brand Relationships

THEORY CARD

Core claim. Consumers can form relationships with brands that resemble interpersonal relationships in their forms, norms, and dynamics. The brand is an active partner whose marketing actions are interpreted as behaviour, and relationship quality predicts stability, forgive-

ness, and loyalty (Fournier 1998).
Origins. Consumer research. Fournier (1998), using phenomenological case studies and interpersonal relationship theory. Extended by work on relationship norms (Aggarwal 2004), transgressions (J. Aaker, Fournier, and Brasel 2004), emotional attachment (Thomson, MacInnis, and Park 2005), brand attachment (Park et al. 2010), and brand love (Batra, Ahuvia, and Bagozzi 2012).
Level of analysis. Dyad (consumer and brand), embedded in communities.
Key constructs. Relationship forms, brand relationship quality (BRQ), relationship norms (exchange and communal), brand attachment (brand-self connection, prominence), brand love, transgression and repair.
Boundary conditions. Most applicable when brands carry symbolic or emotional meaning and are used repeatedly. For many low-involvement purchases, consumers have habits, not relationships.
Related theories. Commitment-trust theory (Section 9.6), Social exchange theory (Section 9.1), Self-congruity, Brand personality.

27.5.1 Origins and core logic

Fournier (1998) argued that the relationship metaphor is legitimate only if brands can act as partners. Brands do act, she reasoned, because consumers interpret marketing actions (a new advertising campaign, a price increase, a product change) as the behaviour of a partner. Through in-depth interviews with three women about the brands in their lives, Fournier identified a typology of relationship forms, from arranged marriages and casual friendships to committed partnerships, flings, dependencies, and enmities, and a model of **brand relationship quality** with six facets: love and passion, self-connection, interdependence, commitment, intimacy, and brand partner quality.

Later work specified the mechanisms more precisely:

- **Norms.** Aggarwal (2004) showed that consumers apply interpersonal relationship norms to brands. In **exchange** relationships, consumers expect quid pro quo. In **communal** relationships, they expect care, and being charged for a small favour feels like a violation.
- **Transgressions.** J. Aaker, Fournier, and Brasel (2004) followed consumers' relationships with an online photo service in a longitudinal field experiment. Relationships with a brand given a **sincere** personality deepened over time but were damaged by a service transgression, while relationships with an **exciting** brand showed signs of reinvigoration after a transgression followed by repair.
- **Attachment.** Thomson, MacInnis, and Park (2005) developed a scale of emotional attachment to brands (affection, passion, connection). Park et al. (2010) defined **brand attachment** as the strength of the bond connecting the brand with the self, reflected in brand-self connection and the prominence of brand-related thoughts, and showed that attachment predicts behaviours that require consumers to invest their own resources better than attitude strength does (Figure 27.7).

Table 27.6 presents the selected relationship forms and constructs in consumer-brand relationship research.

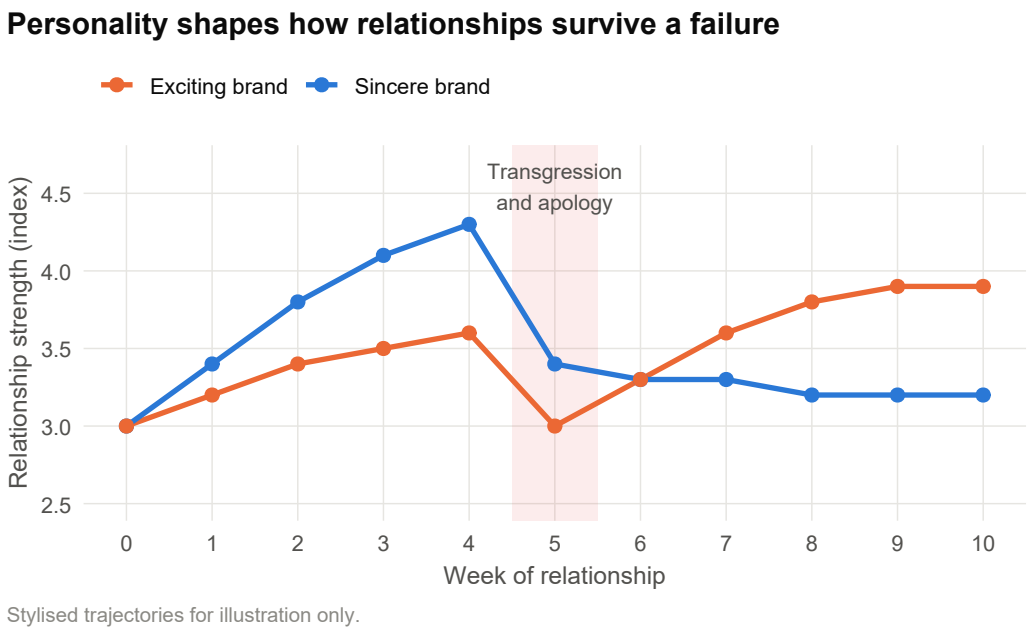


Figure 27.7. Stylised relationship trajectories after a brand transgression, following the pattern reported by Aaker, Fournier and Brasel (2004). Relationship strength with a sincere brand grows faster but does not recover after a service failure, while relationship strength with an exciting brand dips and then partially rebounds after repair. Illustrative values, not study data.

Table 27.6. Selected relationship forms and constructs in consumer-brand relationship research. Consumer statements are illustrative.

Construct	Definition	Illustrative consumer statement	Source
Committed partnership	Long-term, voluntary, socially supported union with high love and intimacy	I would never switch, this brand is part of how I cook	Fournier (1998)
Arranged marriage	Non-voluntary union imposed by a third party, with little affect	It is the detergent my partner always buys	Fournier (1998)
Fling	Short-term, high emotional reward, no commitment	I tried that energy drink all summer	Fournier (1998)
Enmity	Intense negative affect and desire to avoid or harm the partner	I tell everyone not to use that airline	Fournier (1998)
Exchange norm	Benefits given in expectation of comparable return	I pay, they deliver, that is the deal	Aggarwal (2004)
Communal norm	Benefits given out of concern for the partner's needs	They should help me without charging for it	Aggarwal (2004)
Brand attachment	Strength of the bond between brand and self (connection, prominence)	Thoughts about this brand come to me naturally	Park et al. (2010)
Brand love	Passion, self-brand integration, positive emotional connection, anticipated separation distress	I would feel anxious if this brand disappeared	Batra et al. (2012)

27.5.2 Evidence and critique

The relationship perspective has been productive: it explains why loyal customers sometimes punish a beloved brand harshly after a betrayal, why communal brands face backlash when they mone-

tise previously free services (Aggarwal 2004), and why attachment predicts costly behaviours such as recommending, defending, and paying a premium (Park et al. 2010).

DEBATE AND EVIDENCE

Relationship or habit? Sceptics argue that the metaphor overreaches. Most purchase behaviour in repeat-purchase categories follows regularities that require no relationship construct: small brands have fewer buyers who are also somewhat less loyal, the pattern known as **double jeopardy** (A. S. C. Ehrenberg, Goodhardt, and Barwise 1990). On this view, "loyalty" is mostly a by-product of penetration and habit, and investments in deep relationships pay off only for a minority of customers. Relationship theorists reply that the metaphor is most useful where it was developed, for symbolically rich brands and heavy users, and that it predicts responses to transgressions and norm violations that penetration models cannot. The productive position is to treat relationship strength as a variable that is high for some consumer-brand dyads and absent for most.

27.5.3 Applications in business

BUSINESS CASE

Harley-Davidson and Apple: relationship brands with communities. Harley-Davidson founded the Harley Owners Group (H.O.G.) in 1983, giving owners rides, rallies, and local chapters organised around the brand. An ethnography of new bikers by Schouten and McAlexander (1995) described a **subculture of consumption** whose core values (personal freedom, American heritage, and machismo) were expressed through the motorcycle, clothing, and rituals, and in which the company actively participated. Muniz and O'Guinn (2001) studied brand communities around Ford Bronco, Macintosh, and Saab and identified three markers of community: consciousness of kind, shared rituals and traditions, and a sense of moral responsibility to other members. Apple users' defence of the Mac against a dominant rival is a canonical example of the oppositional loyalty that such communities sustain. The managerial lesson from both brands is that a relationship is not only between a consumer and a brand. It is embedded in a community that the firm can support but not fully control.

BUSINESS CASE

Illustrative scenario: monetising a communal brand. A consumer electronics brand known for friendly, free support begins charging for phone help. The fee is modest and consistent with industry practice. Under Aggarwal (2004)'s norms account, customers in a communal relationship will perceive a norm violation, and the backlash will exceed what the money involved would predict. A better design reframes the change in exchange terms from the outset (a premium support tier) and preserves a free communal core.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* After a service failure, customers with high brand attachment will be more forgiving of minor failures but more punitive after failures framed as betrayals of trust.
- *H2.* Charging for a previously free service will lower brand evaluations more under com-

munal than under exchange relationship norms.
- *H3.* Brand attachment will predict share of wallet and advocacy better than brand attitude, controlling for satisfaction.

Established scales. Brand relationship quality facets (Fournier 1998). Emotional attachment scale (Thomson, MacInnis, and Park 2005). Brand attachment (connection and prominence) (Park et al. 2010). Brand love (Batra, Ahuvia, and Bagozzi 2012).
Designs. Longitudinal field experiments with manipulated brand personality and transgressions (J. Aaker, Fournier, and Brasel 2004). Relationship-norm priming experiments (Aggarwal 2004). Phenomenological interviews and netnography for discovery. Panel data on churn after service failures.

COMMON PITFALL

Assuming every customer wants a relationship. Relationship marketing programmes often treat all customers as potential partners. For most consumers in most categories, brand ties are weak and choice is habitual. Segment by relationship strength before investing in intimacy, and test whether relationship measures add predictive power beyond satisfaction and purchase history.

27.6 Self-Congruity Theory

THEORY CARD

Core claim. Consumers prefer products and brands whose image matches their self-concept. The match between brand-user image and self-image (**self-congruity**) motivates choice through the needs for self-consistency and self-esteem (Sirgy 1982). Possessions become part of an **extended self** (Belk 1988).
Origins. Consumer research, drawing on self theory and symbolic interactionism. Early statement by Grubb and Grathwohl (1967). Review and integration by Sirgy (1982). The extended self by Belk (1988).
Level of analysis. Individual consumer.
Key constructs. Actual self, ideal self, social self, ideal social self, brand-user image, self-congruity, self-consistency motive, self-esteem motive, self-brand connection, extended self.
Boundary conditions. Strongest for publicly consumed, symbolically rich products and for consumers high in self-monitoring or public self-consciousness. Weaker for functional, privately consumed products.
Related theories. Brand personality, Social identity theory (Section 22.2), Identity signaling (Section 8.5), Psychological ownership.

27.6.1 Core logic

Sirgy (1982) distinguished four components of the self-concept: the **actual self** (how I see myself), the **ideal self** (how I would like to see myself), the **social self** (how I believe others see me), and the **ideal social self** (how I would like others to see me). Products carry images, often stereotypes of their typical users. When product-user image matches a component of the self, consumption satisfies a motive: congruity with the actual self satisfies **self-consistency**, while congruity with

the ideal self satisfies **self-esteem**. Table 27.7 summarises the logic.

Table 27.7. Components of the self-concept and the motives served by self-congruity (Sirgy, 1982).

Self-concept component	Definition	Motive served by congruity	Positioning implication
Actual self	How I see myself	Self-consistency	Authentic, relatable users
Ideal self	How I would like to see myself	Self-esteem	Aspirational users and settings
Social self	How I believe others see me	Social consistency	Fit with the consumer's reference group
Ideal social self	How I would like others to see me	Social approval	Visible, status-conferring cues

Belk (1988) broadened the argument: people regard their possessions as parts of themselves. Possessions, including objects, places, people, and collections, are incorporated into an **extended self**, for example by controlling, creating, or coming to know them, and the involuntary loss of possessions can feel like a loss of self. Belk (2013) later extended the idea to digital possessions, online profiles, and shared digital spaces (Figure 27.8).

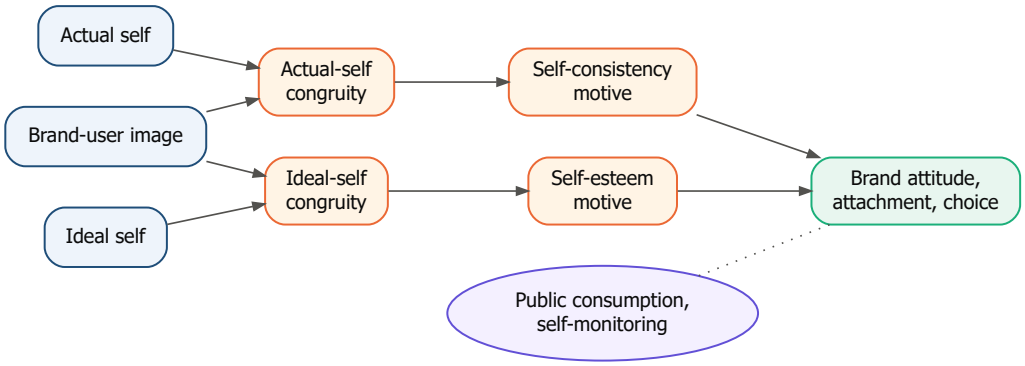


Figure 27.8. Mechanism of self-congruity. A brand-user image is compared with components of the self-concept. Actual-self congruity serves self-consistency and ideal-self congruity serves self-esteem, both feeding attachment and choice. Public consumption and self-monitoring strengthen the link. Based on Sirgy (1982) and Malar et al. (2011).

27.6.2 Evidence and extensions

Self-congruity predicts brand preference, store patronage, and loyalty across many categories. Three extensions refine it. First, Escalas and Bettman (2003) showed that consumers form **self-brand connections** with brands used by groups they belong to or aspire to join, and avoid brands associated with dissociative groups, which links the theory to Social identity theory (Section 22.2). Second, J. L. Aaker (1999) showed that the self is malleable: the traits that drive brand preference depend on which aspects of the self are made salient by the situation. Third, Malär et al. (2011) found that congruity with the actual self drives emotional brand attachment more strongly than congruity with the ideal self, with the ideal-self effect depending on involvement, self-esteem, and public self-consciousness. Aspirational branding is therefore not always the stronger strategy.

DEBATE AND EVIDENCE

How should congruity be measured? Traditional studies rated the brand and the self on the same list of adjectives and computed a discrepancy score. Sirgy et al. (1997) argued that this method suffers from the unreliability of difference scores, the use of predetermined images that may be irrelevant to the consumer, and an implausible assumption that consumers compute congruity attribute by attribute. They proposed a **direct** method: respondents imagine the typical user and rate how well that image matches their own, and the direct method showed stronger predictive validity.

BUSINESS CASE

Actual versus ideal self in beauty advertising. For decades, beauty advertising relied on ideal-self appeals: idealised models and aspirational settings. Dove's "Campaign for Real Beauty", launched in 2004, used women of varied ages and body types instead, an actual-self appeal. The campaign is widely studied as a repositioning that traded the self-esteem motive for self-consistency and authenticity. The research of Malär et al. (2011) offers one explanation for why such a move can strengthen attachment: actual-self congruity tends to be the more powerful driver of emotional bonds.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The effect of self-congruity on brand preference will be stronger for publicly consumed than for privately consumed products.
- *H2.* Actual-self congruity will predict emotional brand attachment more strongly than ideal-self congruity, and the difference will be larger for consumers low in public self-consciousness.

Established measures. Direct self-congruity items (Sirgy et al. 1997). Self-brand connection scale (Escalas and Bettman 2003). Brand personality ratings for the brand and the self (J. L. Aaker 1997).

Designs. Experiments that prime the actual or ideal self. Surveys linking congruity to observed choice. Qualitative studies of the extended self and possession rituals.

COMMON PITFALL

Equating congruity with similarity scores. Computing the absolute difference between self and brand ratings on a fixed list of traits is the default in many studies, but it inherits the problems of difference scores and may miss the images that actually matter to the consumer. Use direct congruity measures or polynomial regression with response surface analysis when you need to model fit between two separately measured profiles.

27.7 Schema Congruity Theory

THEORY CARD

Core claim. Consumers evaluate a new product by comparing it with the schema of the category or brand it belongs to. Products that are **moderately incongruent** with the schema can be evaluated more favourably than fully congruent or extremely incongruent ones, because resolving a modest surprise is itself rewarding (Mandler 1982, Meyers-Levy and Tybout 1989).
Origins. Cognitive psychology (Mandler 1982), brought into consumer research by Meyers-Levy and Tybout (1989). Applied to brand extensions by D. A. Aaker and Keller (1990) and Meyers-Levy, Louie, and Curren (1994).
Level of analysis. Individual consumer evaluating a product, brand extension, or advertisement.
Key constructs. Schema, congruity, moderate incongruity, extreme incongruity, incongruity resolution, elaboration, perceived fit, dilution.
Boundary conditions. The moderate incongruity advantage requires that consumers are motivated and able to resolve the incongruity. It weakens when perceived risk is high (M. C. Campbell and Goodstein 2001) and depends on prior knowledge (Peracchio and Tybout 1996).
Related theories. Goldilocks effect (Section 29.3), Customer-based brand equity, Processing fluency (Section 23.8), Mere exposure effect (Section 29.1).

27.7.1 Origins and core logic

Mandler (1982) proposed that the value people place on an object depends partly on how well it fits the schema they bring to it. A congruent object is familiar and predictable, which produces mild positive affect. An object that is moderately incongruent creates arousal and prompts cognitive work. When the incongruity can be resolved within the existing schema, the successful resolution is experienced as positive and evaluations rise above those for congruent objects. Extreme incongruity cannot be resolved without changing the schema, and the resulting frustration tends to lower evaluations.

Meyers-Levy and Tybout (1989) tested this account with new product descriptions that matched a category schema, departed from it moderately, or departed from it extremely. Moderately incongruent products were evaluated most favourably and also generated more elaborative thought, consistent with Mandler's resolution mechanism. The general pattern, an inverted U between incongruity and evaluation, is sketched in Figure 27.9. The same shape appears in research on attention to stimuli that are neither too simple nor too complex, the Goldilocks effect (Section 29.3).

27.7.2 Brand extensions as schema congruity problems

A **brand extension** uses an established brand name to enter a new category. The parent brand is a schema, and the extension's fit with that schema shapes evaluations. D. A. Aaker and Keller (1990) found that extensions were evaluated more favourably when the parent brand was perceived as high quality and when the extension fit the parent, with fit reflecting complementarity, substitutability, and the transferability of the parent's manufacturing skills. Extensions perceived as too easy to make were evaluated less favourably, as if consumers questioned what the brand added. A secondary analysis of eight studies across countries by Bottomley and Holden (2001)

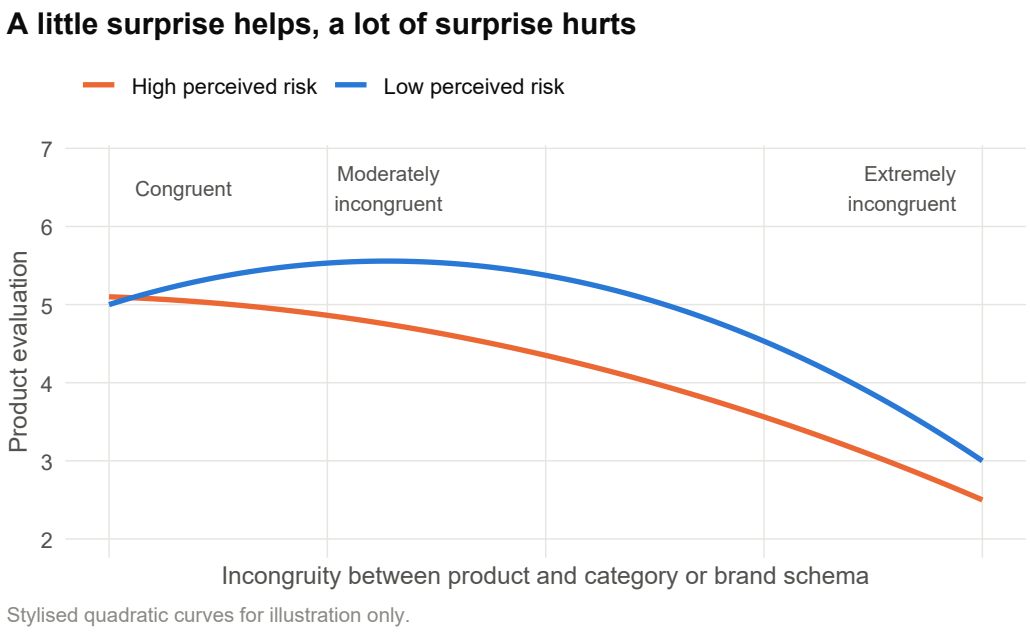


Figure 27.9. Stylised schema congruity effects. With low perceived risk, evaluations peak at moderate incongruity (resolution is rewarding). With high perceived risk, consumers prefer the norm and the advantage of moderate incongruity disappears. Pattern consistent with Meyers-Levy and Tybout (1989) and Campbell and Goodstein (2001). Illustrative curves.

confirmed that parent quality and fit matter in general but that the weights vary across contexts. Völckner and Sattler (2006) identified fit between the parent brand and the extension as the most important driver of extension success, followed by marketing support, parent-brand conviction, retailer acceptance, and parent-brand experience.

Extensions also feed back into the parent schema. Loken and John (1993) showed that extensions with attributes inconsistent with the family brand's beliefs can **dilute** those beliefs, especially when the extension is seen as typical of the brand. Meyers-Levy, Louie, and Curren (1994) found that brand names moderately incongruent with the extension category could produce more favourable evaluations than fully congruent names, which extends the moderate incongruity logic to naming (Table 27.8).

Table 27.8. Brand extensions from the public record interpreted through schema congruity. Outcomes are summarised at the level of public reporting.

Brand	Parent schema	Extension	Outcome	Schema congruity reading
Arm & Hammer	Baking soda: cleaning, deodorising	Toothpaste, deodorisers, laundry products	Long-lived extensions	Transfer of a core benefit makes moderate category distance resolvable
Honda	Engines, reliability	Cars, lawn mowers, generators	Durable multi-category brand	Fit through shared manufacturing competence
Porsche	Performance sports cars	Cayenne sport utility vehicle (2002)	Commercially successful despite enthusiast criticism	Moderate incongruity resolved by performance positioning

Table 27.8. Brand extensions from the public record interpreted through schema congruity. Outcomes are summarised at the level of public reporting. *(continued)*

Brand	Parent schema	Extension	Outcome	Schema congruity reading
Starbucks	Premium coffeehouse	Bottled Frappuccino in grocery channels (with PepsiCo)	Durable line	Complementary category, coffee credentials transfer
Bic	Cheap, disposable pens, lighters, razors	Low-priced perfume (late 1980s)	Withdrawn, widely cited failure	Extreme incongruity: disposability clashes with the symbolism of fragrance
Virgin	Challenger, irreverent service brand	Virgin Cola (1990s)	Failed to win lasting share	Attitude fit, but no capability transfer against entrenched leaders

DEBATE AND EVIDENCE

How robust is the moderate incongruity advantage? The inverted U is attractive but conditional. Peracchio and Tybout (1996) found that prior knowledge moderates the effect: consumers with richer category knowledge can accommodate or discount incongruities and may evaluate products on their attributes instead. M. C. Campbell and Goodstein (2001) showed that when perceived risk is high, consumers prefer congruent products, a "preference for the norm". Brand extension findings also travel imperfectly across cultures and product sets (Bottomley and Holden 2001). A further measurement concern is that studies rarely verify that participants actually resolved the incongruity, the mechanism the theory relies on. The safer claim is that moderate incongruity *can* be an advantage when consumers have the motivation, ability, and low risk needed to resolve it.

27.7.3 Applications in business

BUSINESS CASE

Brand extensions that worked and failed. Porsche's Cayenne, launched in 2002, was a sport utility vehicle from a brand defined by sports cars. Many enthusiasts objected, but the vehicle was engineered and marketed around performance, which allowed consumers to resolve the incongruity within the Porsche schema ("a sports car that happens to be an SUV"). It became one of the brand's highest-volume models. Bic's late-1980s entry into perfume offers the contrast. The Bic schema of cheap, functional, disposable products could not accommodate the sensory, emotional, and status meanings of fragrance, and the line was withdrawn. The managerial test suggested by the theory is whether a consumer can complete the sentence "It makes sense for this brand because…" with a reason drawn from the parent schema.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Moderately incongruent brand extensions will be evaluated more favourably than congruent extensions under low perceived risk, but not under high perceived risk.
- *H2.* The moderate incongruity advantage will be mediated by incongruity resolution, measured as the number of resolution-related thoughts.

- *H3.* Atypical extensions with negative performance will dilute parent-brand beliefs less than typical extensions with negative performance.

Measures. Perceived fit (complementarity, substitutability, transfer) (D. A. Aaker and Keller 1990). Thought listing coded for elaboration and resolution (Meyers-Levy and Tybout 1989). Typicality ratings. Parent-brand belief change (Loken and John 1993).
Designs. Between-subjects experiments with pretested congruity levels. Field data on extension survival (Völckner and Sattler 2006). Secondary analysis across studies (Bottomley and Holden 2001).

COMMON PITFALL

Confusing incongruity with novelty or quality. An "incongruent" stimulus often also differs in attractiveness, novelty, or information content. Without pretests that hold quality constant and a check that resolution occurred, a favourable evaluation of the moderately incongruent product may reflect a better product, not the resolution mechanism.

27.8 Hedonic and Utilitarian Consumption

THEORY CARD

Core claim. Products deliver **hedonic** value (pleasure, fun, fantasy, sensory experience) and **utilitarian** value (function, instrumentality, practicality). The two kinds of value are evaluated differently: hedonic consumption evokes a stronger need for justification and more guilt, which makes choice sensitive to framing, effort, and prior virtuous acts (Hirschman and Holbrook 1982, Dhar and Wertenbroch 2000, Okada 2005, Khan and Dhar 2006).
Origins. Consumer research. Hirschman and Holbrook (1982) and Holbrook and Hirschman (1982) introduced the hedonic and experiential perspective. Dhar and Wertenbroch (2000) and later work brought it into behavioural decision research.
Level of analysis. Individual consumer, product, and attribute.
Key constructs. Hedonic value, utilitarian value, guilt, justification, licensing, effort, acquisition versus forfeiture, mode of payment.
Boundary conditions. Most products have both dimensions. Effects depend on the salience of the hedonic dimension and on whether the decision context calls for justification (joint evaluation, accountability).
Related theories. Mental accounting (Section 23.5), Self-control (Section 20.4), [Feelings-as-information](#feelings-as-information-theory) (Section 23.7) theory (Section 23.7), Warm-glow heuristic (Section 29.2).

27.8.1 Origins and core logic

Hirschman and Holbrook (1982) defined **hedonic consumption** as the facets of consumer behaviour that relate to the multisensory, fantasy, and emotive aspects of product experience. Together with Holbrook and Hirschman (1982), who described an experiential view of consumption built on "fantasies, feelings, and fun", they challenged the information-processing paradigm that treated consumers as problem solvers choosing among attribute bundles. Behavioural decision researchers later asked how the two kinds of value trade off in choice.

Four propositions summarise the core findings:

- **P1 (Loss aversion differs).** Hedonic attributes loom larger when consumers give something up than when they acquire it. Dhar and Wertenbroch (2000) found that consumers were relatively more likely to choose a hedonic item in a forfeiture decision (which item to give up) than in an acquisition decision (which item to get).
- **P2 (Justification).** Hedonic options are harder to justify. Okada (2005) showed that a hedonic option tends to be rated more highly when evaluated on its own, while a utilitarian option is more often chosen when the two are compared side by side, and that consumers prefer to pay for hedonic goods with time and for utilitarian goods with money.
- **P3 (Licensing).** A prior virtuous act, or even the intention to act virtuously, boosts the self-concept and licenses a subsequent hedonic choice (Khan and Dhar 2006).
- **P4 (Earned indulgence).** Effort provides a justification. Kivetz and Simonson (2002) found that requiring more effort in a frequency programme shifted preferences toward luxury rather than necessity rewards (Figure 27.10).

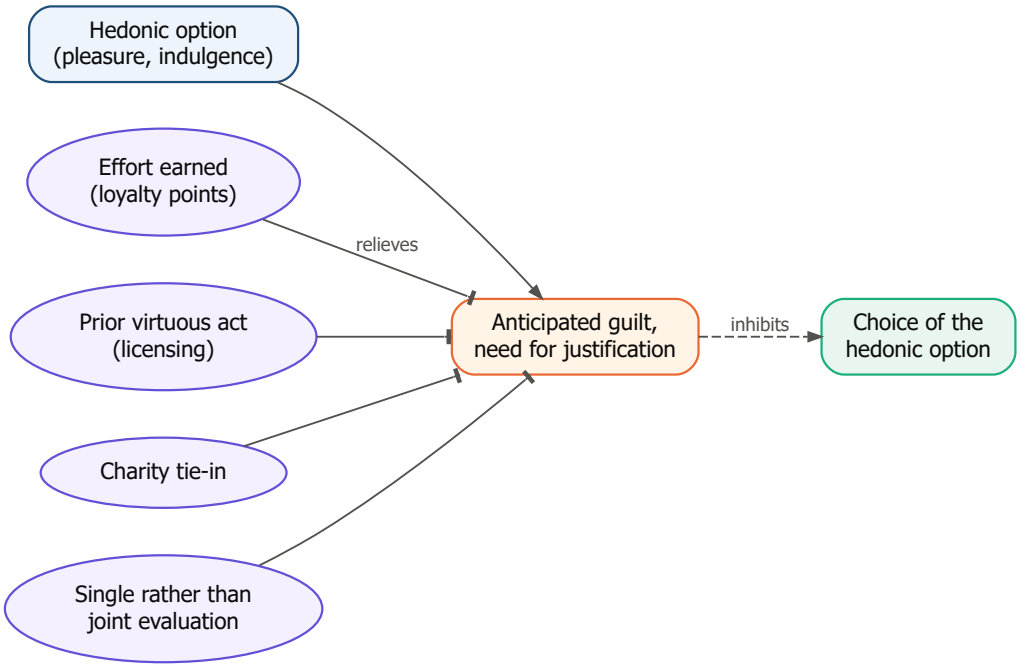


Figure 27.10. Justification mechanism in hedonic consumption. Hedonic options evoke anticipated guilt and a need for justification. Justification cues (effort, prior virtue, a charitable tie-in, evaluating the option on its own) relieve that need and increase hedonic choice. Based on Okada (2005), Khan and Dhar (2006), Kivetz and Simonson (2002), and Strahilevitz and Myers (1998).

Table 27.9 presents the how hedonic and utilitarian value differ in consumer decision making.

Table 27.9. How hedonic and utilitarian value differ in consumer decision making.

Dimension	Hedonic	Utilitarian	Key evidence
Nature of benefit	Pleasure, fantasy, sensory and emotional experience	Function, problem solving, instrumentality	Hirschman & Holbrook (1982)
Need for justification	High: evokes guilt	Low: easy to justify	Okada (2005)

Table 27.9. How hedonic and utilitarian value differ in consumer decision making. *(continued)*

Dimension	Hedonic	Utilitarian	Key evidence
Acquisition versus forfeiture	Weighs more when giving up	Weighs relatively more when acquiring	Dhar & Wertenbroch (2000)
Preferred payment	Time	Money	Okada (2005)
Effect of a prior virtuous act	Licensed: choice share rises	Little change	Khan & Dhar (2006)
Effect of required effort	Effort justifies luxury rewards	Necessity rewards preferred at low effort	Kivetz & Simonson (2002)
Charity incentive effectiveness	More effective	Less effective	Strahilevitz & Myers (1998)
Measurement	HED subscale	UT subscale	Voss et al. (2003)

27.8.2 Evidence and extensions

The distinction is now standard in consumer research and is reviewed by Alba and Williams (2013). Strahilevitz and Myers (1998) showed that charity incentives (a donation tied to purchase) are more effective for frivolous products than for practical ones, consistent with the idea that the donation offsets guilt. The two dimensions can be measured with the HED/UT scale of Voss, Spangenberg, and Grohmann (2003), which treats them as separate attitude components rather than poles of a single continuum, and shopping itself can deliver hedonic and utilitarian value (Babin, Darden, and Griffin 1994).

DEBATE AND EVIDENCE

Category or dimension? Much research classifies products as hedonic (chocolate, concerts) or utilitarian (detergent, printer cartridges). Critics note that the same product can be bought for either reason and that the hedonic-utilitarian distinction overlaps with other contrasts, such as vice versus virtue, wants versus shoulds, and luxuries versus necessities, which have different theoretical mechanisms (self-control, status, price). Studies that manipulate the salient goal of a purchase, or measure both dimensions with validated scales, yield cleaner inferences than studies that rely on category labels.

27.8.3 Applications in business

BUSINESS CASE

Hedonic purchase framing and loyalty programme design. The justification mechanism suggests several practical levers. **Effort framing**: points that consumers feel they have earned are more likely to be spent on indulgences, which supports offering luxury rewards (spa days, upgrades) in programmes that require sustained effort (Kivetz and Simonson 2002). **Progress framing**: loyalty cards that present progress as already under way increase completion (Nunes and Drèze 2006). **Programme mechanics**: Starbucks moved its Rewards programme in 2016 from earning per transaction to earning per dollar spent, a change that altered what customers had to do to earn a treat and illustrates how programme rules shape the effort-reward link. **Cause tie-ins**: the (PRODUCT)RED initiative, launched in 2006, partnered with brands such as Apple and Gap to contribute part of the proceeds from special-edition

products to the Global Fund, pairing discretionary consumer products with a charitable justification, consistent with Strahilevitz and Myers (1998).

BUSINESS CASE

Illustrative scenario: framing a premium headphone. A retailer tests two product pages for premium noise-cancelling headphones. Page A emphasises sound quality and immersive listening (hedonic). Page B adds a utilitarian frame: "focus at work, fewer distractions on your commute". The theory predicts that Page B lowers the need for justification and raises conversion among buyers who would otherwise feel guilty, while Page A performs better among buyers shopping for a gift, where justification concerns are weaker.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Requiring effort to earn a reward will increase the share of hedonic rewards chosen, and the effect will be mediated by reduced anticipated guilt.
- *H2.* A charitable tie-in will raise purchase likelihood more for products rated high on the HED subscale than for products rated high on the UT subscale.
- *H3.* Joint display of a hedonic and a utilitarian option will reduce the hedonic choice share relative to sequential display.

Established scales. HED/UT scale (Voss, Spangenberg, and Grohmann 2003). Hedonic and utilitarian shopping value (Babin, Darden, and Griffin 1994).
Designs. Choice experiments manipulating effort, prior virtue, and evaluation mode. Field experiments in loyalty programmes. Panel data on redemption choices.

COMMON PITFALL

Labelling categories instead of measuring value. Classifying a product as "hedonic" because it is chocolate, or "utilitarian" because it is a laptop, ignores that consumers buy laptops for gaming and chocolate as a gift. Measure the hedonic and utilitarian dimensions for the target consumers, and separate the hedonic-utilitarian contrast from vice-virtue and luxury-necessity contrasts when theorising mechanism.

27.9 Psychological Ownership

THEORY CARD

Core claim. People can feel that a target (an object, an idea, a brand, an organisation) is "theirs" without legal title. This feeling of **psychological ownership** arises from control, intimate knowledge, and investment of the self, and it raises valuation, attachment, and responsibility for the target (Pierce, Kostova, and Dirks 2001).
Origins. Organisational behaviour (Pierce, Kostova, and Dirks 2001, 2003), extended to consumers by research on touch (Peck and Shu 2009), the endowment effect (Shu and Peck 2011, Reb and Connolly 2007), and consumer empowerment (Fuchs, Prandelli, and Schreier 2010).
Level of analysis. Individual (with collective variants).

Key constructs. Feelings of ownership, routes (control, intimate knowledge, investment of self), motives (efficacy, self-identity, having a place), valuation, territoriality.
Boundary conditions. Requires a target that can be controlled, known, or invested in. Effects can turn negative, as territoriality, when ownership feels threatened.
Related theories. Endowment effect (Section 23.4), IKEA effect (Section 29.7), Self-congruity, Self-determination theory (Section 19.6).

27.9.1 Core logic

Pierce, Kostova, and Dirks (2001) defined psychological ownership as the state in which individuals feel as though the target of ownership, or a piece of it, is theirs. They proposed that the state serves three motives: **efficacy and effectance** (being in control), **self-identity** (expressing and maintaining the self), and **having a place** (a home or territory). Three routes produce it: **controlling** the target, **coming to know it intimately**, and **investing the self** in it through effort, time, or creativity. Figure 27.11 shows the mechanism and its consumer outcomes.

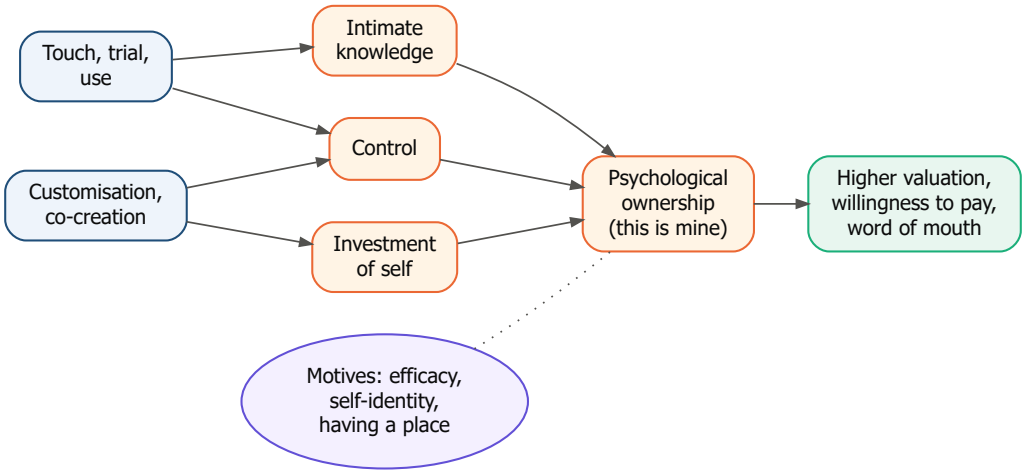


Figure 27.11. Routes to psychological ownership and their consumer outcomes. Touch, customisation, co-creation, and use produce feelings of ownership through control, intimate knowledge, and self-investment. Based on Pierce, Kostova and Dirks (2001) and Peck and Shu (2009).

27.9.2 Evidence

Peck and Shu (2009) showed that merely touching an object increases perceived ownership and valuation, and that vivid imagery of touching can substitute for touch when physical contact is not possible. Reb and Connolly (2007) separated legal ownership from physical possession and found that feelings of ownership tracked valuations, which places psychological ownership at the heart of the Endowment effect (Section 23.4). Shu and Peck (2011) showed that psychological ownership and the affective reactions it produces help explain the size of the endowment effect. Investment of the self is the mechanism behind the IKEA effect (Section 29.7): people value products they have assembled themselves more highly (Norton, Mochon, and Ariely 2012). Fuchs, Prandelli, and Schreier (2010) found that consumers empowered to select which products a firm should produce showed stronger demand for those products, with psychological ownership mediating the effect. Jussila et

al. (2015) review the marketing literature, and Morewedge et al. (2021) argue that the shift from owning to accessing (streaming, renting, sharing) reduces the routes to psychological ownership, with consequences for how much consumers value and care for what they use.

BUSINESS CASE

Designing for "mine". Retailers and brands use each route. **Touch and trial**: open displays, in-store demonstrations, test drives, and generous home-trial periods let consumers control and get to know a product before purchase (Peck and Shu 2009). **Customisation**: configurators such as Nike's online customisation service let buyers invest the self in a product's design. **Co-creation**: LEGO Ideas invites fans to submit and vote on set designs, giving participants a stake in products they did not legally own, consistent with Fuchs, Prandelli, and Schreier (2010). The theory also warns of a cost: customers with strong feelings of ownership can react territorially when a firm changes "their" product or community, as many brand-community backlashes show.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Allowing consumers to customise a product will raise willingness to pay, mediated by psychological ownership.
- *H2.* The effect of touch on valuation will be stronger for products with positive than with neutral sensory properties.
- *H3.* Access-based consumers (subscribers) will report lower psychological ownership and take less care of products than owners.

Measures. Psychological ownership items adapted from Van Dyne and Pierce (2004) ("I feel this is my product"). Willingness to pay and willingness to accept.
Designs. Touch versus no-touch experiments (Peck and Shu 2009). Empowerment manipulations (Fuchs, Prandelli, and Schreier 2010). Field experiments with trial periods.

COMMON PITFALL

Equating psychological ownership with legal ownership or liking. The construct is a feeling of possession, distinct from legal title and from attitude. Studies that infer it from ownership status or brand liking cannot test the theory. Measure the feeling directly and show that it mediates the outcome.

27.10 Variety Seeking

THEORY CARD

Core claim. Consumers sometimes switch away from a familiar or preferred option not because of a better offer but because variety itself has value, or because they misforecast how much variety they will want (McAlister and Pessemier 1982, Simonson 1990).
Origins. Consumer research and economics of choice. Typology by McAlister and Pessemier (1982). Diversification bias by Simonson (1990) and D. Read and Loewenstein (1995). Choosing less-preferred options for variety by Ratner, Kahn, and Kahneman (1999).
Level of analysis. Individual consumer, product category.

Key constructs. Derived versus direct variety, satiation, stimulation seeking, optimum stimulation level, diversification bias, sensory-specific satiety.
Boundary conditions. Strongest for hedonic, sensory, and frequently purchased categories, for simultaneous choices made for future consumption, and for consumers with a high optimum stimulation level.
Related theories. Hedonic adaptation (Section 25.2), Coolidge effect (Section 29.4), Choice architecture (Section 23.10), Hedonic and utilitarian consumption.

27.10.1 Core logic

McAlister and Pessemier (1982) distinguished **derived** varied behaviour, which results from external forces (several household members, changing situations, price promotions, new products), from **direct** variety seeking, in which variety itself is the goal. Direct variety seeking can be intrapersonal (satiation with an attribute, the desire for stimulation) or interpersonal (affiliation with or distinction from others). Table 27.10 summarises the typology.

Table 27.10. Typology of varied behaviour following McAlister and Pessemier (1982).

Type	Driver	Example	Managerial implication
Derived: multiple needs	Different users or uses in a household	One yoghurt for children, another for adults	Not true variety seeking: target uses
Derived: changing constraints	Promotions, stockouts, new products	Switching to whichever brand is on deal	Switching reflects price, not taste
Direct: intrapersonal	Satiation with attributes, stimulation seeking	Rotating snack flavours	Offer flavour or format variety within the brand
Direct: interpersonal	Affiliation with or distinction from others	Trying what friends drink, or avoiding it	Link to identity signaling and social influence

Variety seeking is not always a stable preference. Simonson (1990) showed that consumers who choose several items at once for future consumption choose more variety than those who choose one item at a time, a pattern D. Read and Loewenstein (1995) called **diversification bias** and demonstrated even among children choosing Halloween candy. Ratner, Kahn, and Kahneman (1999) found that people choose less-preferred songs for the sake of variety, even though doing so lowers their enjoyment during consumption. At the attribute level, Inman (2001) showed that variety seeking follows **sensory-specific satiety**: consumers switch flavours more than brands. Individual differences matter as well: consumers with a high **optimum stimulation level** engage in more exploratory behaviour, including variety seeking (Raju 1980, Steenkamp and Baumgartner 1992), and category characteristics shape how much true variety seeking occurs (van Trijp, Hoyer, and Inman 1996) (Figure 27.12).

BUSINESS CASE

Illustrative scenario: subscription boxes and flavour rotation. A snack subscription asks customers to choose a month of snacks at once. Diversification bias predicts that they will pick more variety than they will enjoy, which can raise later dissatisfaction with less-preferred items. Two design responses follow from the theory: allow week-by-week swaps (converting simultaneous into sequential choice), or default the box to a mix weighted toward each customer's revealed favourites. Snack and beverage brands also exploit genuine sensory-specific

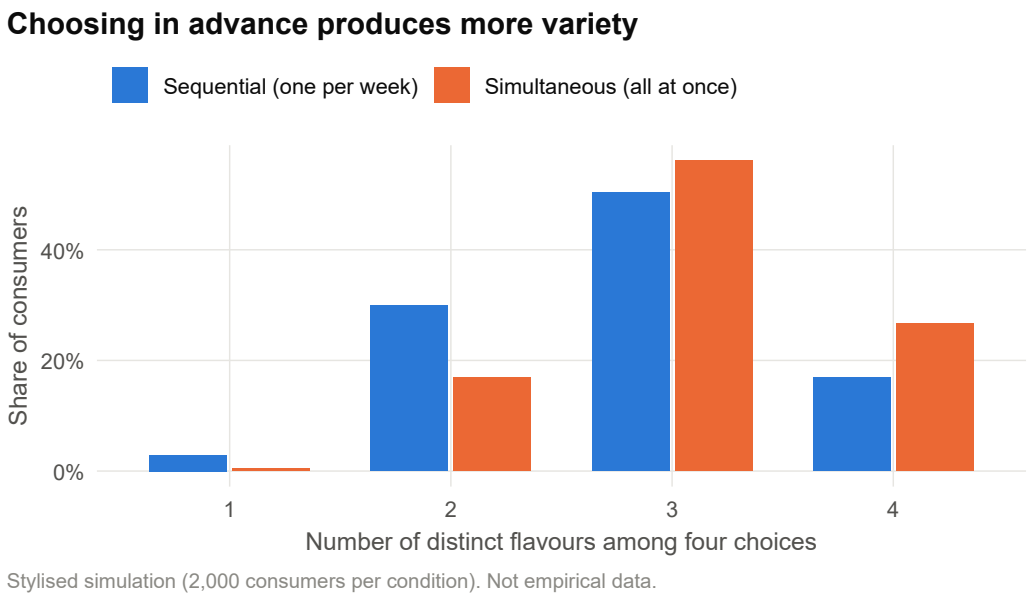


Figure 27.12. Illustrative simulation of diversification bias. Simulated consumers choose four snacks from six flavours either simultaneously (all at once, for the next four weeks) or sequentially (one each week). Simultaneous choosers spread choices across more distinct flavours, the pattern reported by Simonson (1990) and Read and Loewenstein (1995). Stylised parameters.

satiety with rotating limited-edition flavours that keep variety seekers inside the brand rather than switching to a competitor.

RESEARCH TOOLKIT

Sample hypotheses and designs. *H1.* Bundled, advance purchases will contain more distinct items than repeated single purchases by the same consumers. *H2.* Brand switching driven by satiation will occur more at the flavour than at the brand level (Inman 2001). Use scanner panels to separate true variety seeking from promotion-driven switching, choice experiments contrasting simultaneous and sequential choice (Simonson 1990), and optimum stimulation level scales (Raju 1980, Steenkamp and Baumgartner 1992).

COMMON PITFALL

Calling every switch variety seeking. Brand switching in panel data often reflects promotions, stockouts, or multiple household users, which McAlister and Pessemier (1982) classify as derived varied behaviour. Control for price and availability before attributing switching to a taste for variety.

27.11 Reference Price Theory

THEORY CARD

Core claim. Consumers judge a price against a **reference price**, an internal standard formed from past prices or an external standard supplied by the environment. Prices below the reference are coded as gains and prices above it as losses, and losses affect choice more strongly than equivalent gains (Winer 1986, Kalyanaram and Winer 1995, Mazumdar, Raj, and Sinha 2005).
Origins. Behavioural pricing and marketing science, drawing on adaptation-level theory, Prospect theory (Chapter 24) (Kahneman and Tversky 1979), and transaction utility in Mental accounting (Section 23.5) (R. Thaler 1985). Brand choice models by Winer (1986). Reviews by Kalyanaram and Winer (1995) and Mazumdar, Raj, and Sinha (2005).
Level of analysis. Individual consumer, aggregated to brand demand and retail pricing policy.
Key constructs. Internal reference price, external reference price, price expectation, gain and loss, loss aversion, transaction utility, promotion frequency.
Boundary conditions. Strongest for frequently purchased goods with salient price histories and for consumers who track prices. Reference effects are harder to identify when consumers differ widely in price sensitivity (Bell and Lattin 2000).
Related theories. Prospect theory (Chapter 24), Loss aversion (Section 24.4), Anchoring effect (Section 29.9), Mental accounting (Section 23.5), Decoy effect (Section 29.8).

27.11.1 Core logic and propositions

Classical demand theory predicts that consumers respond to the current price. Reference price theory adds that they respond to the current price *relative to a standard*. Winer (1986) built this idea into a brand choice model for frequently purchased products, in which the reference price is a consumer's expectation formed from the brand's past prices, and choice depends on both the observed price and its deviation from the reference. Mayhew and Winer (1992) showed with scanner data that both **internal** (memory-based) and **external** (point-of-purchase) reference prices affect choice.

- **P1 (Reference dependence).** Holding price constant, purchase probability rises when price is below the reference price and falls when it is above (Winer 1986, Kalyanaram and Winer 1995).
- **P2 (Adaptive formation).** Internal reference prices are formed largely from the brand's own past prices. Briesch et al. (1997) compared alternative specifications and found that models based on the brand's past prices fit best.
- **P3 (Asymmetry).** Consumers react more to prices above the reference (losses) than to prices below it (gains) (Hardie, Johnson, and Fader 1993, Kalyanaram and Winer 1995), an application of Loss aversion (Section 24.4).
- **P4 (External anchors).** Advertised regular or "compare at" prices raise perceived value and lower search, even when the reference is exaggerated (Urbany, Bearden, and Weilbaker 1988, Grewal, Monroe, and Krishnan 1998).
- **P5 (Promotion erosion).** Because reference prices adapt, frequent promotions lower the internal reference price, so that the regular price later feels like a loss.

Figure 27.13 illustrates P5 with an adaptive reference price. Under frequent promotions, the refer-

ence price drifts toward the deal price, and every return to the regular price becomes a perceived loss.

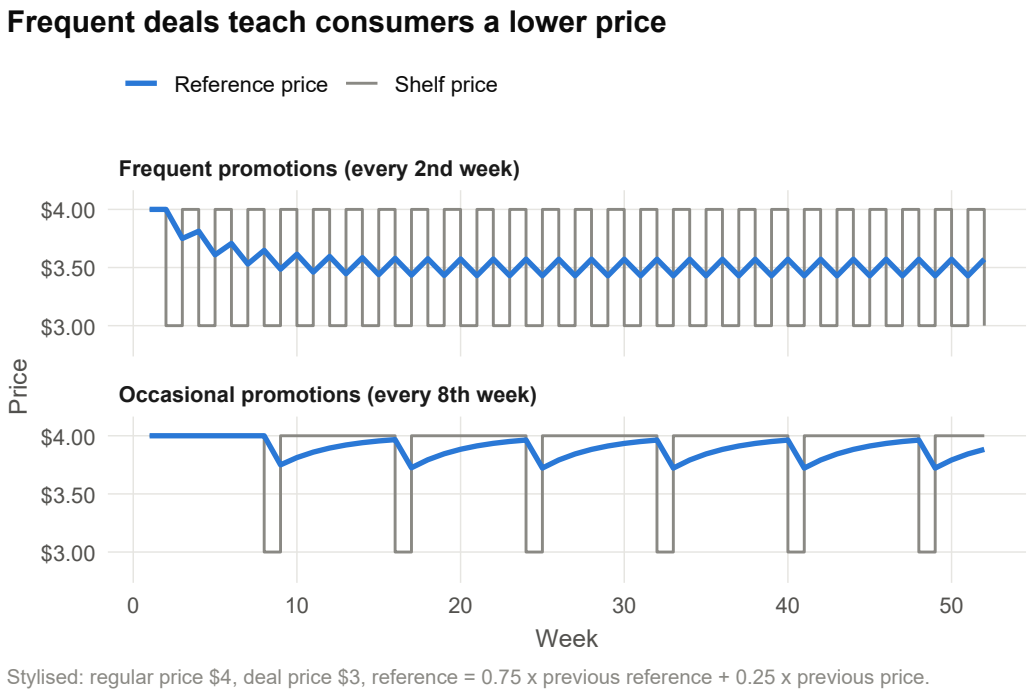


Figure 27.13. Illustrative adaptive reference prices under two promotion policies. The reference price is an exponentially weighted average of past prices. Frequent promotions pull the reference price toward the deal price, so the regular price is increasingly coded as a loss. Stylised simulation.

Table 27.11 presents the internal and external reference prices (organised following Mazumdar, Raj, and Sinha, 2005).

Table 27.11. Internal and external reference prices (organised following Mazumdar, Raj, and Sinha, 2005).

Feature	Internal reference price	External reference price
Definition	A price standard held in memory	A price standard presented in the purchase environment
Typical sources	Past prices paid or seen, expected price, perceived fair price	Advertised regular price, 'compare at' price, competitor prices, price of adjacent items
Formation	Adaptive, updated by recent prices	Supplied by the seller or context at the point of purchase
Typical measures	Exponentially smoothed past prices, stated expected price	Experimental manipulation of advertised reference prices
Managerial lever	Promotion depth and frequency, price consistency	Price-comparison advertising, assortment and display
Regulatory concern	Low	High: fictitious former prices can be deceptive
Key sources	Winer (1986), Briesch et al. (1997)	Urbany et al. (1988), Grewal et al. (1998)

27.11.2 Evidence and critique

Kalyanaram and Winer (1995) summarised a set of empirical generalisations: reference prices affect brand choice, past prices are a central input to their formation, and the response to price losses is larger than the response to gains. Mazumdar, Raj, and Sinha (2005) organise the literature around how reference prices are formed, how they are retrieved, and how they affect evaluations and choice, and they offer propositions for future research.

DEBATE AND EVIDENCE

Is loss aversion in scanner data real? Bell and Lattin (2000) showed that heterogeneity in price sensitivity can masquerade as loss aversion. Price-sensitive households tend to buy when prices are low (gains) and price-insensitive households when prices are high (losses), so a pooled model can find asymmetric responses even when no individual is loss averse. Once heterogeneity was accounted for, estimated loss aversion shrank substantially. The debate does not overturn reference dependence, but it means that asymmetry estimates from aggregate data require careful modelling of consumer heterogeneity. A second debate concerns which standard consumers actually use: the brand's own past prices, the category's prices, or the external price at the shelf (Briesch et al. 1997, Mayhew and Winer 1992).

27.11.3 Applications in business

BUSINESS CASE

Outlet stores and "compare at" prices. Outlet malls allow manufacturers to segment price-sensitive shoppers who will travel and accept less service from consumers who buy at full-price stores (Coughlan and Soberman 2005). Price tags in outlets often display a "compare at" or "original" price alongside the selling price. Reference price theory explains why: an external reference raises perceived transaction value even when consumers discount it (Urbany, Bearden, and Weilbaker 1988, Grewal, Monroe, and Krishnan 1998). The same mechanism creates regulatory risk. The US Federal Trade Commission's guides against deceptive pricing address former-price comparisons, and in 2014 several members of the US Congress asked the Commission to examine whether outlet pricing and labelling misled shoppers, partly because some outlet goods are made specifically for outlets and were never sold at the reference price. For managers, the lesson is that reference prices create value only while they remain credible.

BUSINESS CASE

JCPenney's "Fair and Square" pricing. In 2012 the US department store chain JCPenney, under a new chief executive, replaced most of its frequent sales and coupons with simpler everyday prices. The strategy assumed that shoppers disliked the promotional game. Sales and store traffic fell steeply, the chief executive was replaced in 2013, and promotions returned. Reference price theory and transaction utility (R. Thaler 1985) suggest why: years of promotions had taught customers low internal reference prices and a habit of enjoying the "deal". Removing external reference prices eliminated the perceived gain without lowering the price consumers expected to pay.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The effect of a price increase on brand choice will be larger than the effect of an equivalent price decrease, after controlling for household heterogeneity in price sensitivity.
- *H2.* Increasing promotion frequency will lower consumers' stated expected price for the brand.
- *H3.* Plausible external reference prices will raise perceived transaction value more than implausible ones, and the effect will be weaker among consumers with high price knowledge.

Measures. Internal reference price from exponentially smoothed past prices (Winer 1986, Briesch et al. 1997). Stated expected and fair prices. Acquisition and transaction value scales (Grewal, Monroe, and Krishnan 1998).

Designs. Brand choice models with reference price terms estimated on scanner panels, with latent-class or random-coefficient heterogeneity (Bell and Lattin 2000). Experiments manipulating advertised reference prices (Urbany, Bearden, and Weilbaker 1988). Field experiments varying promotion schedules.

COMMON PITFALL

Estimating loss aversion from pooled data. Asymmetric price responses in aggregate data can arise from heterogeneity, endogenous purchase timing, and stockpiling rather than loss aversion. Model heterogeneity, account for purchase timing, and treat asymmetry as a hypothesis to be tested rather than a parameter to be assumed.

27.12 Consumer Culture Theory

THEORY CARD

Core claim. Consumption is a cultural practice. Consumers use goods, brands, and services to construct identities, form communities, and negotiate social structures and ideologies, and the meanings of goods are produced and circulated by culture, firms, and consumers together (Arnould and Thompson 2005).

Origins. Interpretive consumer research drawing on anthropology, sociology, and cultural studies (McCracken 1986, Belk, Wallendorf, and Sherry 1989, Holt 2002). The label "consumer culture theory" was introduced by Arnould and Thompson (2005) to name a family of perspectives rather than a single theory.

Level of analysis. Individual, community, market, and society.

Key constructs. Cultural meaning, meaning transfer, consumption rituals, identity projects, marketplace cultures and subcultures, brand communities, cultural capital, marketplace ideologies, authenticity.

Boundary conditions. Best suited to explaining meaning, identity, and practice. It aims at theoretical transferability across contexts rather than statistical generalisation, and it is less suited to predicting the size of effects.

Related theories. Self-congruity, Consumer-brand relationships, Social identity theory (Section 22.2), Institutional theory (Section 7.2), Identity signaling (Section 8.5).

27.12.1 Origins and core logic

McCracken (1986) offered the foundational account of how goods acquire meaning. Cultural meaning resides first in the **culturally constituted world**, the categories and principles through which a society organises experience. Two institutions, **advertising** and the **fashion system**, transfer that meaning into consumer goods. Consumers then draw meaning from goods into their own lives through **rituals**: possession, exchange (such as gift giving), grooming, and divestment. McCracken (1989) applied the same logic to celebrity endorsement: celebrities carry meanings from their roles, which advertising transfers to products and consumption transfers to consumers. Figure 27.14 shows the model.



Figure 27.14. McCracken's (1986) model of the movement of cultural meaning. Meaning moves from the culturally constituted world into goods through advertising and the fashion system, and from goods into consumers' lives through possession, exchange, grooming, and divestment rituals.

Arnould and Thompson (2005) reviewed two decades of interpretive research and organised it into four research programmes (Table 27.12). The common thread is that consumers are not merely choosers among attribute bundles but active producers of meaning who work with, and sometimes against, the meanings that marketers offer. Holt (2002) developed this point into a dialectical theory of branding: consumers increasingly resist brands they perceive as cultural engineering, and brands retain value when they serve as authentic cultural resources for consumers' own identity projects. In later work, Holt (2004) argued that iconic brands succeed by offering myths that address contradictions in a society's culture at particular historical moments.

Table 27.12. The four research programmes of consumer culture theory identified by Arnould and Thompson (2005).

Research programme	Central question	Illustrative study	Managerial relevance
Consumer identity projects	How do consumers use market resources to construct coherent selves?	Extended self (Belk, 1988)	Positioning brands as identity resources
Marketplace cultures	How do consumers forge communities and subcultures around consumption?	Harley-Davidson bikers (Schouten & McAlexander, 1995), brand communities (Muniz & O'Guinn, 2001)	Community management, co-creation
Sociohistoric patterning of consumption	How do class, gender, ethnicity, and institutions structure consumption?	Cultural capital and taste (Holt, 1998)	Segmentation beyond demographics
Mass-mediated marketplace ideologies	How do consumers interpret, adopt, or resist commercial meanings?	Doppelganger brand images (Thompson et al., 2006)	Monitoring brand meaning and backlash

27.12.2 Evidence and critique

CCT research has produced influential concepts: the **sacred and profane** in consumption (Belk, Wallendorf, and Sherry 1989), **subcultures of consumption** (Schouten and McAlexander 1995),

brand communities (Muniz and O'Guinn 2001), the structuring role of **cultural capital** in taste (Holt 1998), and the **doppelganger brand image**, a parody of a brand's image circulated by critics that signals when emotional branding has lost authenticity (C. J. Thompson, Rindfleisch, and Arsel 2006). Methodologically, it introduced ethnography, phenomenological interviewing, and **netnography**, the ethnographic study of online communities (Kozinets 2002), into marketing.

DEBATE AND EVIDENCE

Theory or label, and how much context? Sceptics from the positivist tradition question whether findings from small, purposive samples generalise, and whether a "family of perspectives" is a theory at all. Defenders respond that CCT seeks analytic rather than statistical generalisation and that its concepts have proved transferable across markets. A critique from within the tradition is equally pointed. Askegaard and Linnet (2011) argue that CCT has focused too heavily on individual consumers' lived experience and needs to attend to the "context of context": the social, historical, and institutional structures that shape what consumers can experience in the first place.

BUSINESS CASE

Meaning transfer in celebrity endorsement and cultural branding. Nike's Air Jordan line, launched in 1985 around a then-rookie basketball player, is a textbook case of McCracken's meaning transfer: the athlete's cultural meanings (excellence, style, defiance) moved into the product through advertising and into consumers' lives through use, collecting, and display (McCracken 1989). CCT also explains the limits of such strategies. When consumers perceive a brand's meanings as manufactured, they produce counter-meanings, the doppelganger images that C. J. Thompson, Rindfleisch, and Arsel (2006) identified around Starbucks. The managerial implication is to monitor not only what the brand says but also what consumers and critics are saying about it, and to ground brand meaning in practices consumers can recognise as authentic (Holt 2002).

RESEARCH TOOLKIT

Typical research questions and propositions. CCT studies usually build theory rather than test hypotheses, but they generate propositions that other methods can test, for example: *P1.* Brand communities that share rituals and a sense of moral responsibility will show greater resistance to competitor switching than communities based only on product use. *P2.* Consumers with high cultural capital will favour brands whose meanings are subtle and require interpretive knowledge.

Designs and data. Ethnography and participant observation (Schouten and McAlexander 1995). Phenomenological interviews. Netnography of online communities (Kozinets 2002). Historical and archival analysis of advertising and media. Mixed methods that pair interpretive discovery with survey or experimental tests.

COMMON PITFALL

Using interpretive research as decoration. A handful of quotations does not make a CCT study. The tradition's contribution comes from sustained engagement with a context, theoretically motivated sampling, and an explicit account of how meanings are produced and

contested. Conversely, quantitative researchers should not dismiss CCT findings for lacking statistical generalisation: they are best treated as theoretical mechanisms to be tested and bounded.

27.13 Chapter Summary

KEY TAKEAWAYS

- Satisfaction is a **relative judgment**. Expectations, performance, and their comparison all matter, and shortfalls hurt more than surpluses help (R. L. Oliver 1980, E. W. Anderson and Sullivan 1993).
- Product attributes matter because they lead to **consequences and values**, and laddering reveals these chains (Gutman 1982, Reynolds and Gutman 1988).
- **Brand equity** is a differential response to marketing that originates in brand knowledge, and it should be measured against a benchmark (Keller 1993, D. A. Aaker 1991).
- Brands acquire **personalities** and become **relationship partners** whose transgressions, norms, and attachment shape loyalty (J. L. Aaker 1997, Fournier 1998, Park et al. 2010).
- Consumers prefer brands that fit their **self-concept**, and possessions become part of an **extended self** (Sirgy 1982, Belk 1988).
- **Moderate incongruity** with a schema can be an advantage, but only when consumers can resolve it and risk is low (Meyers-Levy and Tybout 1989, M. C. Campbell and Goodstein 2001).
- **Hedonic** consumption requires justification, so effort, prior virtue, and charity tie-ins change choices (Dhar and Wertenbroch 2000, Okada 2005, Khan and Dhar 2006).
- **Psychological ownership** and **variety seeking** explain valuation and switching beyond price and quality (Pierce, Kostova, and Dirks 2001, Simonson 1990).
- Prices are judged against **reference prices**, and promotions reshape those standards (Winer 1986, Mazumdar, Raj, and Sinha 2005).
- **Consumer culture theory** treats consumption as meaning-making that firms shape but do not control (McCracken 1986, Arnould and Thompson 2005).

Table 27.13 presents the comparison of the chapter's theories.

Table 27.13. Comparison of the chapter's theories.

Theory	Key mechanism	Signature prediction	Typical business use
Expectancy-disconfirmation	Comparison with expectations	Shortfalls hurt more than surpluses help	Satisfaction measurement, expectation management
Means-end chain	Attribute to consequence to value	Value-level appeals outperform attribute claims for involving products	Positioning, advertising strategy
Customer-based brand equity	Brand knowledge in memory	Same marketing works better for the brand than for an unbranded version	Brand tracking, price premiums
Brand personality	Trait inference and self-expression	Personality shapes attitudes and relationship dynamics	Brand identity, creative briefs

Table 27.13. Comparison of the chapter's theories. *(continued)*

Theory	Key mechanism	Signature prediction	Typical business use
Consumer-brand relationships	Relationship norms and attachment	Transgression effects depend on relationship type	CRM, communities, crisis response
Self-congruity	Self-consistency and self-esteem	Preference for brands matching the self	Target user imagery
Schema congruity	Incongruity resolution	Inverted U in evaluations	Brand extensions, new product design
Hedonic and utilitarian consumption	Guilt and justification	Justification cues raise hedonic choice	Loyalty rewards, cause marketing, framing
Psychological ownership	Control, knowledge, self-investment	Touch and customisation raise valuation	Trials, configurators, co-creation
Variety seeking	Satiation and anticipated satiation	More variety in simultaneous choices	Assortment, flavour rotation, subscriptions
Reference price	Gains and losses against a standard	Frequent deals lower reference prices	Promotion policy, price communication
Consumer culture theory	Cultural meaning transfer and resistance	Brands thrive as authentic cultural resources	Cultural branding, community strategy

27.14 Discussion Questions

1. A hotel chain's satisfaction scores fall after a successful advertising campaign while service quality is unchanged. Use expectancy-disconfirmation theory to explain the pattern and design a study to test your explanation.
2. Conduct three short laddering interviews about a product you recently bought. Draw a hierarchical value map. Which level of the ladder offers the best positioning opportunity, and why?
3. Keller defines brand equity as a *differential* response. Design a study that measures brand equity for a national brand of breakfast cereal at the customer-mindset and product-market levels. What benchmark will you use?
4. Aaker, Fournier, and Brasel found that sincere and exciting brands respond differently to transgressions. How should a crisis communication plan differ for the two kinds of brands? State a testable hypothesis.
5. Choose a brand extension announced in the past five years. Predict its success using schema congruity theory and the drivers identified by Völckner and Sattler. What evidence would falsify your prediction?
6. A retailer plans to cut its promotion frequency in half. Using reference price theory, forecast the short-run and long-run effects on sales and perceived value, and explain how you would separate loss aversion from consumer heterogeneity in the data.

27.15 Further Reading

- R. L. Oliver (1980): the classic statement of expectancy-disconfirmation and the starting point for satisfaction research.
- Keller (1993): the foundational definition of customer-based brand equity and its memory-based logic.

- J. L. Aaker (1997): the brand personality scale, best read together with the critique in A. Azoulay and Kapferer (2003).
- Fournier (1998): the paper that legitimised the study of consumer-brand relationships, rich in qualitative detail.
- Meyers-Levy and Tybout (1989): a compact and elegant test of schema congruity with a clear mechanism.
- Mazumdar, Raj, and Sinha (2005): the most comprehensive review of reference price research, with testable propositions.
- Arnould and Thompson (2005): the map of consumer culture theory and its research programmes.
- Alba and Williams (2013): a review of hedonic consumption that connects the hedonic-utilitarian distinction to broader research on pleasure.

Chapter 28

Marketing Strategy and Services

In the early 1990s many large manufacturers still described themselves by what came off their assembly lines. Thirty years later, aircraft-engine makers sell "hours of thrust", software firms sell subscriptions rather than licences, and retailers compete on what they know about each customer as much as on what they stock. Why do some firms sense these shifts early and others late? Why does a bank with modern branches and competitive rates still lose customers to a rival that simply answers the phone? Why is a subscriber who stays one extra year worth so much more than a few points of price? And why do so many advertising plans still assume that consumers move neatly from awareness to purchase?

These questions belong to **marketing strategy**: the set of theories that explain how firms understand markets, create and deliver value, and allocate marketing resources over time. This chapter covers seven of the field's most influential frameworks. **Market orientation** explains how organizations generate and act on market intelligence. **Service-dominant logic** reframes what is exchanged in markets. The **gaps model of service quality** diagnoses where service failures originate. **Customer equity** treats customers as financial assets. The **hierarchy of effects** models how advertising is supposed to work. The **product life cycle** describes how markets evolve, and the **marketing mix** organizes the decisions a manager controls. Each is widely taught and widely used, and each has attracted serious criticism. Understanding both the logic and the critique is what separates a theory-informed strategist from a framework user.

The seven frameworks fall into three groups. Market orientation and service-dominant logic are theories of how a firm should *understand* markets and value. The gaps model and customer equity are theories of how service and customer relationships translate into *outcomes*, perceived quality in one case and financial value in the other. The hierarchy of effects, the product life cycle, and the marketing mix are planning frameworks for *managing* programs over consumers' decision processes, over a market's evolution, and across the decisions a manager controls. The first group draws on the resource-based (Section 4.1) and learning traditions. The second draws on psychology and finance. The third is closest to everyday practice, and for that reason it is the most in need of the critical scrutiny that later research has supplied (Table 28.1).

Table 28.1. Chapter at a glance: theories of marketing strategy and services.

Theory	Core question	Level of analysis	Key source
Market orientation	Do firms that generate and act on market intelligence perform better?	Firm / business unit	Kohli & Jaworski (1990), Narver & Slater (1990)
Service-dominant logic	What is exchanged in markets, and who creates value?	Market / service ecosystem	Vargo & Lusch (2004)

Table 28.1. Chapter at a glance: theories of marketing strategy and services. *(continued)*

Theory	Core question	Level of analysis	Key source
Gaps model of service quality	Where do service-quality shortfalls originate?	Firm / customer	Parasuraman, Zeithaml & Berry (1985)
Customer equity	What are customers worth, and how should marketing be valued?	Customer / firm	Blattberg & Deighton (1996), Gupta et al. (2004)
Hierarchy of effects	Through which mental stages does advertising work?	Individual consumer	Lavidge & Steiner (1961)
Product life cycle	How do sales, competition, and strategy change as markets mature?	Product class / market	Levitt (1965)
Marketing mix	Which controllable variables make up a marketing program?	Firm / offering	Borden (1964), McCarthy (1960)

28.1 Market Orientation

THEORY CARD

Core claim. Businesses that systematically generate intelligence about customers and competitors, share it across functions, and respond to it achieve superior performance.
Origins. Marketing strategy. Two parallel 1990 articles in the *Journal of Marketing*: a behavioural conceptualization by Kohli and Jaworski (1990) and a cultural one by Narver and Slater (1990). Meta-analysed by Kirca, Jayachandran, and Bearden (2005).
Level of analysis. Strategic business unit or firm.
Key constructs. Intelligence generation, intelligence dissemination, responsiveness, customer orientation, competitor orientation, interfunctional coordination, market turbulence, competitive intensity.
Boundary conditions. Benefits depend on whether customer needs are knowable and changing, on the cost of collecting intelligence, and on whether responding to current customers diverts resources from disruptive opportunities.
Related theories. Resource-based theory (Section 4.1), Market-based assets (Section 4.5), Dynamic capabilities (Section 5.1), Absorptive capacity (Section 5.2), Organizational learning (Section 14.3), Disruptive innovation (Section 12.2).

28.1.1 Origins and intellectual history

The **marketing concept**, the idea that a business should be organized around satisfying customer needs profitably, had been a staple of textbooks since the 1950s. What it lacked was a definition precise enough to measure. In 1990 two research teams independently supplied one. Kohli and Jaworski (1990) reviewed the literature and interviewed managers across functions and industries. They defined **market orientation** as the organization-wide *generation* of market intelligence about current and future customer needs, the *dissemination* of that intelligence across departments, and organization-wide *responsiveness* to it. Their definition is behavioural: market orientation is something a business *does*.

Narver and Slater (1990) took a cultural view. Market orientation, in their account, is the organizational culture (Section 14.6) that most effectively creates the behaviours needed to deliver superior value to buyers. It has three behavioural components, **customer orientation**, **competitor**

orientation, and **interfunctional coordination**, and two decision criteria, a long-term focus and profitability. Using data from 140 strategic business units of a single corporation, they reported a substantial positive association between market orientation and profitability.

The two conceptualizations were soon operationalized in widely used scales: MARKOR for the behavioural view (Kohli, Jaworski, and Kumar 1993) and the Narver and Slater scale for the cultural view. Day (1994) then reframed the construct as a set of **capabilities**, notably *market sensing* and *customer linking*, connecting market orientation to the resource-based view (Section 4.1) and to later work on market-based assets (Section 4.5) (Srivastava, Shervani, and Fahey 1998) (Table 28.2).

Table 28.2. Two conceptualizations of market orientation.

Feature	Behavioural view	Cultural view
Key source	Kohli & Jaworski (1990)	Narver & Slater (1990)
Definition	Generation, dissemination, and use of market intelligence	Culture that creates behaviours for superior customer value
Components	Intelligence generation, dissemination, responsiveness	Customer orientation, competitor orientation, interfunctional coordination
Typical measure	MARKOR (Kohli, Jaworski & Kumar, 1993)	Narver & Slater multi-item scale
Unit of focus	Activities and processes	Values, norms, and coordination
Managerial lever	Information systems, cross-functional routines, incentives	Leadership, shared values, hiring and socialization

28.1.2 Core logic and propositions

The logic links information to action. Firms that know more about what customers value, and about what competitors are about to offer, can design better offerings and react faster. That knowledge has value only if it reaches the people who design, price, and deliver the offering, and only if they act on it. Hence the theory's central propositions:

- **P1 (Performance).** Market orientation is positively related to business performance (Narver and Slater 1990, Jaworski and Kohli 1993).
- **P2 (Antecedents).** Top-management emphasis, interdepartmental connectedness, and market-based reward systems increase market orientation. Interdepartmental conflict and centralization reduce it (Jaworski and Kohli 1993).
- **P3 (Mediation).** The effect on performance runs largely through innovativeness, product and service quality, and customer loyalty (Kirca, Jayachandran, and Bearden 2005).
- **P4 (Environment).** Market orientation should matter more when customer preferences change quickly and competition is intense, although empirical support for this moderation is weak (Jaworski and Kohli 1993).
- **P5 (Employee outcomes).** Market orientation is associated with stronger employee commitment and esprit de corps (Jaworski and Kohli 1993).

Figure 28.1 summarizes this structure.

28.1.3 Evidence and extensions

Hundreds of studies have tested the market orientation-performance link across industries and countries. The meta-analysis by Kirca, Jayachandran, and Bearden (2005) synthesised this liter-

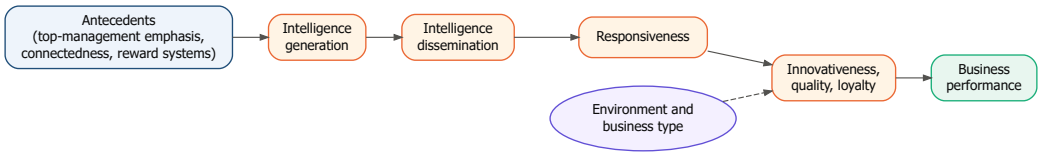


Figure 28.1. Mechanism of market orientation. Organizational antecedents shape the generation, dissemination, and use of market intelligence. Market orientation improves performance mainly through innovativeness, quality, and customer loyalty. Environmental conditions and the type of business moderate the effect. Synthesised from Jaworski and Kohli (1993) and Kirca et al. (2005).

ature and found a positive relationship with overall business performance, with innovativeness, customer loyalty, and quality acting as mediators. The relationship was stronger in manufacturing than in service firms and in national cultures with lower power distance and uncertainty avoidance, which links the theory to research on culture (Section 18.2). Top-management emphasis, interdepartmental connectedness, and market-based reward systems emerged as reliable antecedents.

A second stream asks what kind of market orientation matters. Narver, Slater, and MacLachlan (2004) distinguish **responsive** market orientation, which addresses customers' expressed needs, from **proactive** market orientation, which tries to discover latent needs that customers cannot yet articulate. In their data, the proactive form was the stronger predictor of new-product success.

DEBATE AND EVIDENCE

Can listening to customers be dangerous? Studies of disk-drive makers by Christensen and Bower (1996) showed that leading firms failed in part *because* they listened carefully to their largest existing customers, who did not want the smaller, lower-performing drives that later displaced them. Read narrowly, this looks like a refutation of market orientation. Read carefully, it shows that the theory's boundary conditions matter: a responsive orientation focused on current mainstream customers can blind a firm to emerging segments and technologies (Christensen 1997). The proactive extension (Narver, Slater, and MacLachlan 2004) and the original definition's emphasis on *future* needs (Kohli and Jaworski 1990) partly address the critique. Two methodological concerns remain. Most evidence relies on cross-sectional, single-informant surveys in which managers rate both orientation and performance, which inflates correlations through common-method variance. And because successful firms can afford more market research, reverse causality is hard to rule out.

28.1.4 Applications in business

Market orientation is a practical diagnostic. A business can audit each step in the intelligence chain. Is customer information collected beyond satisfaction surveys, including complaints, front-line observations, and lost-sale analysis? Does it reach product, operations, and finance, or does it stay in the marketing department? And is anyone accountable for acting on it within a defined time?

BUSINESS CASE

Market orientation in retailing. Two well-documented retailers illustrate different routes to the same capability. Tesco's Clubcard loyalty programme, launched in the United Kingdom in the mid-1990s and analysed with the data science firm dunnhumby (which Tesco later took

over), turned transaction data into intelligence used for ranging, promotions, and personalized offers. This is a strong form of *intelligence generation and dissemination*. Inditex's Zara built *responsiveness*: store managers report what customers ask for and what sells, central design teams turn those signals into new items, and short production cycles let the firm act on intelligence within weeks rather than seasons. The lesson for managers is that data collection alone is not market orientation. Clubcard data created value because it changed decisions, and Zara's store reports mattered because the supply chain could respond to them.

28.1.5 Using the theory in research

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Market orientation will be positively related to new-product performance, and this effect will be mediated by product innovativeness.
- *H2.* The positive effect of responsive market orientation on performance will be weaker, and the effect of proactive orientation stronger, in markets with high technological turbulence.
- *H3.* Market-based reward systems will strengthen the link between intelligence dissemination and responsiveness.

Measures. MARKOR (Kohli, Jaworski, and Kumar 1993), the Narver and Slater scale (Narver and Slater 1990), responsive and proactive scales (Narver, Slater, and MacLachlan 2004). Objective performance from financial databases rather than self-reports where possible.
Designs. Multi-informant surveys with lagged objective outcomes, panel designs that track changes in orientation over time, and text analysis of annual reports or earnings calls as an unobtrusive measure of customer focus.

COMMON PITFALL

Equating market orientation with "doing what customers say". The construct concerns the *process* of generating, sharing, and acting on intelligence about current *and future* needs and about competitors. Measuring only customer satisfaction or survey activity misses dissemination and responsiveness, and treating stated preferences as the whole of market intelligence invites the trap Christensen documented.

28.2 Service-Dominant Logic

THEORY CARD

Core claim. Service, the application of knowledge and skills for the benefit of another party, is the fundamental basis of all exchange. Goods are vehicles for service, and value is co-created by multiple actors and determined by the beneficiary in use.
Origins. Marketing theory. Vargo and Lusch (2004), refined in Vargo and Lusch (2008) and Vargo and Lusch (2016), building on services-marketing work such as Shostack (1977).
Level of analysis. Market, service ecosystem, and actor-to-actor exchange.

Key constructs. Service (singular), operant and operand resources, value co-creation, value-in-use, value proposition, resource integration, service ecosystem, institutions.
Boundary conditions. A lens or meta-theory rather than a set of directly testable hypotheses. Most useful where value depends heavily on how customers use and combine resources.
Related theories. Market orientation, Resource-advantage theory (Section 4.4), Social exchange theory (Section 9.1), Commitment-trust theory (Section 9.6), Institutional theory (Section 7.2).

28.2.1 Origins and core logic

Marketing grew out of economics, and economics grew up studying the production and exchange of tangible goods. Services were long treated as a residual category, defined by what they lacked: they were *intangible*, *heterogeneous*, *inseparable* from consumption, and *perishable*. Shostack (1977) argued that marketing needed to "break free" from product thinking, because most offerings combine tangible and intangible elements.

Vargo and Lusch (2004) went further. They argued that the goods-services distinction itself is the problem. In their **service-dominant (S-D) logic**, *service*, in the singular, is the process of using one's competences for the benefit of another. People exchange service for service, and goods are simply distribution mechanisms for service: a car provides the service of transportation, and money is a claim on future service. The theory distinguishes **operand resources**, which must be acted on to be useful (raw materials, machines), from **operant resources**, which act on other resources (knowledge, skills, relationships). Operant resources, S-D logic argues, are the fundamental source of competitive advantage, which connects the theory to the resource-based view (Section 4.1) and to resource-advantage theory (Section 4.4) (Hunt and Morgan 1995).

The second core idea concerns value. Under **goods-dominant logic**, value is created by the firm, embedded in the product, and realized at the point of sale (**value-in-exchange**). Under S-D logic, value is realized when the beneficiary integrates the offering with its own resources in a context of use (**value-in-use**). The firm cannot deliver value on its own. It can only offer **value propositions**, and the customer is always a co-creator of value (Vargo and Lusch 2008). The 2016 update consolidated the framework into five axioms and added a fifth element: value co-creation is coordinated through actor-generated **institutions** and institutional arrangements, such as norms, rules, and shared meanings (Vargo and Lusch 2016).

DEFINITION

Definition 28.1. The five axioms of service-dominant logic (Vargo and Lusch 2016). (1) Service is the fundamental basis of exchange. (2) Value is co-created by multiple actors, always including the beneficiary. (3) All social and economic actors are resource integrators. (4) Value is always uniquely and phenomenologically determined by the beneficiary. (5) Value co-creation is coordinated through actor-generated institutions and institutional arrangements.

The axioms shift the unit of analysis away from the firm-customer dyad. Because every actor integrates resources from many sources, a customer's value-in-use depends on complementary offerings, other customers, and the rules of the setting, not only on the focal firm. A smartphone delivers value through apps, networks, and the user's own skills. A hospital stay delivers value through clinicians, family members, and insurance arrangements. S-D logic calls these nested systems **service**

ecosystems, which links it to research on network effects (Section 17.6) and two-sided markets (Section 17.7) (Table 28.3).

Table 28.3. Goods-dominant versus service-dominant logic, synthesised from Vargo and Lusch (2004, 2008, 2016).

Dimension	Goods-dominant logic	Service-dominant logic
Unit of exchange	Goods (and services as a residual category)	Service: applied knowledge and skills
Role of goods	End product that embodies value	Vehicle or appliance for delivering service
Primary resource	Operand resources (materials, equipment)	Operant resources (knowledge, skills, relationships)
Locus of value	Value-in-exchange, realized at sale	Value-in-use, determined by the beneficiary
Role of customer	Target and recipient of value	Co-creator and resource integrator
Firm's role	Produce and deliver value	Offer value propositions and support co-creation
Relationship focus	Transactions	Ongoing relationships within service ecosystems
Business model example	Selling an engine	Charging per hour of engine availability

Figure 28.2 represents value co-creation as resource integration within an institutional context.

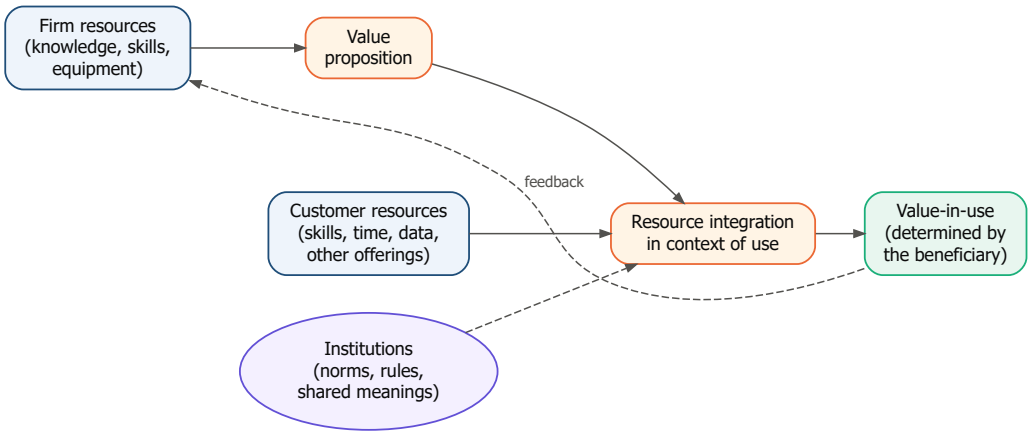


Figure 28.2. Value co-creation in service-dominant logic. The firm offers a value proposition. The customer integrates it with its own operant and operand resources in a context of use, and value-in-use is determined by the beneficiary. Institutions coordinate the process. Redrawn from the logic of Vargo and Lusch (2008, 2016).

28.2.2 Evidence and extensions

Because S-D logic is a perspective rather than a falsifiable theory, its "evidence" is mostly conceptual uptake and the empirical programmes it has inspired. Two are especially productive. The first is **servitization**, the movement of manufacturers into services. Vandermerwe and Rada (1988) coined the term to describe firms offering "bundles" of goods, services, support, and knowledge. Baines et al. (2009) review the literature on servitized manufacturing and identify the challenges firms face, including designing integrated product-service offerings and building service cultures within product organizations. The second is research on **service ecosystems** and platforms, where value emerges from interactions among many actors rather than from a firm-customer dyad.

DEBATE AND EVIDENCE

Insight or relabelling? Critics raise three concerns. First, if *everything* is service, the concept risks losing discriminating power: a theory that classifies every exchange the same way may explain little about why exchanges differ. Second, the claim that customers *always* co-create value is true by definition under S-D logic, and therefore cannot be tested. Empirical work instead studies variation in the *degree* of customer participation. Third, S-D logic's foundational premises were revised several times (Vargo and Lusch 2008, 2016), which supporters see as healthy evolution and critics see as a moving target. A fair assessment is that S-D logic is most valuable as a reframing device. It directs attention to use contexts, customer resources, and business models that capture value over time, and specific mid-range theories are needed to generate testable predictions.

28.2.3 Applications in business

BUSINESS CASE

Servitization at Rolls-Royce. The aircraft-engine business is the best-known illustration of S-D logic in practice. Rolls-Royce traces its "Power by the Hour" concept to the early 1960s and today offers engine care under its TotalCare programme. Instead of buying an engine and then paying separately for spare parts and repairs, an airline pays for engine availability, typically per flying hour, and the manufacturer assumes responsibility for maintenance. The shift changes incentives on both sides. Under a parts-and-repair model, the supplier earns more when engines break. Under a per-hour model, it earns more when engines stay on the wing, which aligns its incentives with the airline's value-in-use. The model also changes what the firm must know: engine health monitoring data become an operant resource, and long-term service contracts create the relationship continuity that S-D logic emphasises. The risk moves too. The manufacturer now bears the cost of unexpected failures, so pricing requires accurate reliability forecasting, a point that links servitization to transaction cost economics (Section 6.3) and agency theory (Section 6.1).

RESEARCH TOOLKIT

Research uses. Treat S-D logic as a framing that motivates specific hypotheses. Examples: (1) outcome-based contracts will increase supplier investment in reliability relative to transactional contracts, (2) customers with greater operant resources (expertise, complementary assets) will realise more value-in-use from the same offering, (3) servitized manufacturers will show more stable revenue but face a service-paradox period in which profitability falls before it rises. **Designs.** Contract-level data from B2B firms, field studies of usage data, comparative case studies of servitization transitions.

COMMON PITFALL

Using "co-creation" as a slogan. Asking customers to do more work (assembly, self-service, data entry) is not value co-creation unless it increases value-in-use for them. Researchers should measure the customer's resources and outcomes, not only the firm's intent.

28.3 The Gaps Model of Service Quality

THEORY CARD

Core claim. Customers judge service quality by comparing what they expected with what they perceive they received. That customer gap results from four internal gaps in the provider's understanding, design, delivery, and communication.
Origins. Services marketing. Parasuraman, Zeithaml, and Berry (1985) proposed the gaps model and Parasuraman, Zeithaml, and Berry (1988) developed the SERVQUAL measurement instrument. Challenged by Cronin and Taylor (1992).
Level of analysis. The firm (internal gaps) and the individual customer (perceived quality).
Key constructs. Expected service, perceived service, the five gaps, tangibles, reliability, responsiveness, assurance, empathy.
Boundary conditions. Applies to services in which customers form expectations and can evaluate delivery. Measurement of expectations is contested, and the five dimensions do not replicate cleanly in every industry.
Related theories. Expectancy disconfirmation (Section 27.1), Peak-end rule (Section 25.7), Market orientation, Service-dominant logic.

28.3.1 Origins and intellectual history

Services posed a measurement problem. A car's quality can be assessed by inspecting it, but a bank teller's courtesy or an airline's handling of a missed connection exists only as an experience. European work by Grönroos (1984) distinguished **technical quality** (what the customer receives) from **functional quality** (how it is delivered). In the United States, Parasuraman, Zeithaml, and Berry (1985) conducted focus groups with consumers and interviews with executives in several service categories, including retail banking, credit cards, securities brokerage, and product repair. They concluded that customers assess quality as the difference between expectations and perceptions, and that the size of that difference depends on a chain of internal shortfalls.

28.3.2 Core logic and the five gaps

The model identifies five gaps:

1. **Gap 1, the knowledge gap.** Management's perception of customer expectations differs from what customers actually expect.
2. **Gap 2, the standards gap.** Service-quality specifications do not reflect management's perception of expectations.
3. **Gap 3, the delivery gap.** Actual service delivery departs from the specifications.
4. **Gap 4, the communication gap.** External communications promise something other than what is delivered.
5. **Gap 5, the customer gap.** Perceived service falls short of expected service. This is perceived service quality, and the model treats it as a function of gaps 1 to 4 (Parasuraman, Zeithaml, and Berry 1985).

Customer expectations are shaped by word of mouth (Section 26.3), personal needs, past experience, and the firm's own communications. This last link is what makes Gap 4 so consequential: advertising that over-promises raises expectations and widens Gap 5 even when delivery is unchanged. Figure 28.3 redraws the model.

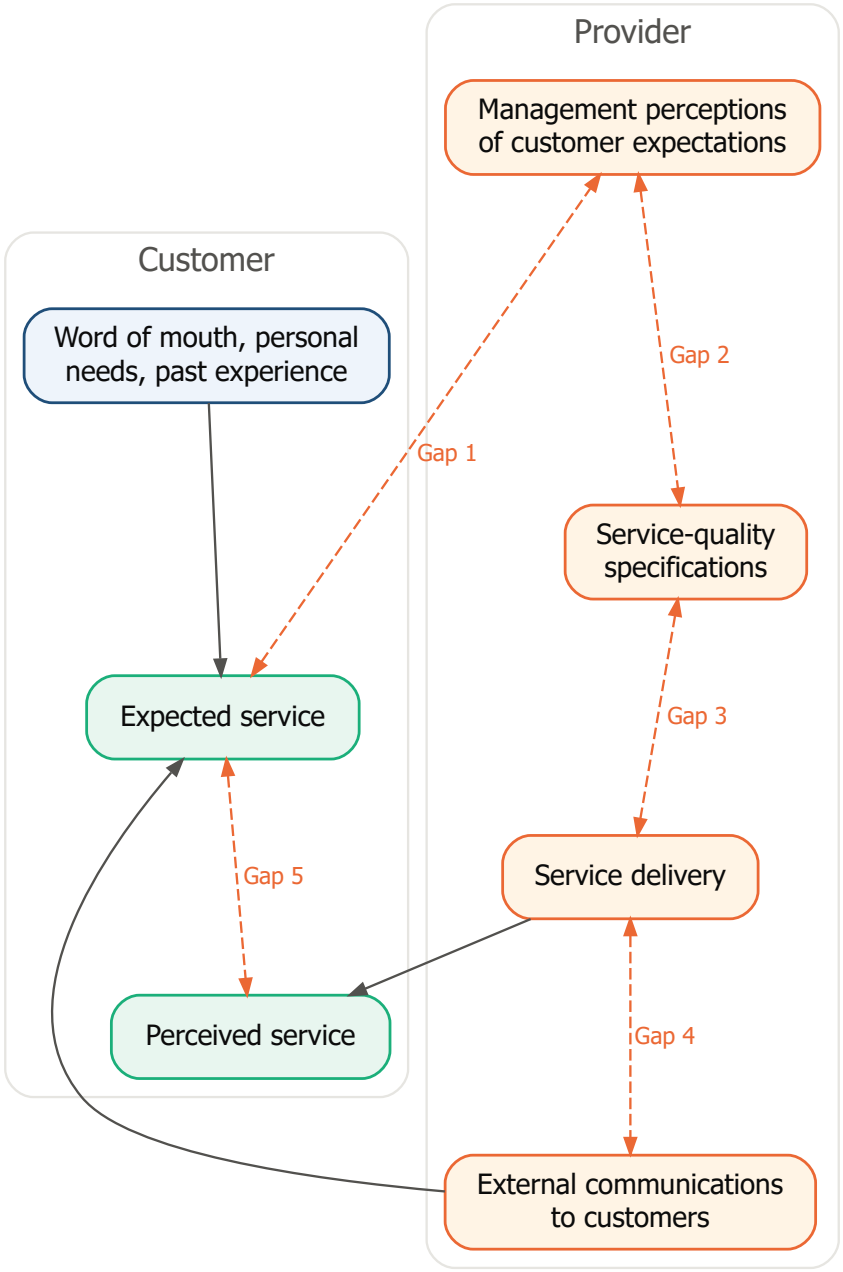


Figure 28.3. The gaps model of service quality. Solid arrows show influences on customer expectations and perceptions. Dashed arrows mark the five gaps, including the three links along which the provider translates perceived expectations into specifications, delivery, and communications. The customer gap (Gap 5) between expected and perceived service is a function of four provider gaps. Redrawn from Parasuraman, Zeithaml and Berry (1985).

28.3.3 Measuring the customer gap: SERVQUAL

To measure Gap 5, Parasuraman, Zeithaml, and Berry (1988) developed **SERVQUAL**, a 22-item instrument in which respondents rate both what they expect from excellent firms in a category and what they perceive from the focal firm. The difference score, perception minus expectation, is computed for each item. The ten determinants identified in the 1985 study were consolidated into five dimensions (Table 28.4).

Table 28.4. The five SERVQUAL dimensions (definitions paraphrased from Parasuraman, Zeithaml and Berry, 1988) with illustrative applications. The illustrations are not scale items.

Dimension	Definition	Retail banking illustration	Airline illustration
Tangibles	Physical facilities, equipment, and appearance of personnel	Clean branches, a usable app	Cabin condition, gate facilities
Reliability	Ability to perform the promised service dependably and accurately	Error-free statements and transfers	On-time departures, baggage arrives
Responsiveness	Willingness to help customers and provide prompt service	Fast resolution of a blocked card	Rebooking after a cancellation
Assurance	Knowledge and courtesy of employees and their ability to inspire trust	Competent, confident advisers	Calm, knowledgeable crew
Empathy	Caring, individualized attention	Advice suited to a customer's situation	Help for families or passengers with special needs

Figure 28.4 shows what a gap analysis looks like in practice, using simulated survey data for two services.

28.3.4 Evidence and the SERVQUAL-SERVPERF debate

SERVQUAL became one of the most widely applied instruments in marketing, and also one of the most criticised. Cronin and Taylor (1992) argued that the difference-score formulation was neither theoretically nor empirically justified. Testing alternatives in banking, pest control, dry cleaning, and fast food, they found that a **performance-only** measure, later called **SERVPERF**, explained more variance in overall quality than the gap scores. They also concluded that service quality is an antecedent of customer satisfaction, and that satisfaction had a stronger influence on purchase intentions than service quality did. Teas (1993) questioned what respondents mean by "expectations": an ideal standard, a predicted level, or what a firm *should* provide. Parasuraman, Zeithaml, and Berry (1994) responded that expectation measures retain diagnostic value, because they show managers *where* shortfalls are largest relative to what customers want, even if performance-only scores predict overall judgements better (Table 28.5).

Table 28.5. The SERVQUAL versus SERVPERF debate.

Issue	SERVQUAL position	SERVPERF position
What is measured	Perceptions minus expectations	Perceptions only
Theoretical basis	Disconfirmation of expectations	Quality as an attitude toward performance
Predictive validity	Lower in head-to-head comparisons	Higher variance explained in overall quality

Table 28.5. The SERVQUAL versus SERVPERF debate. *(continued)*

Issue	SERVQUAL position	SERVPERF position
Diagnostic value	Identifies where shortfalls relative to expectations are largest	Identifies weak dimensions, not relative shortfalls
Respondent burden	Two ratings per item	One rating per item
Key sources	Parasuraman et al. (1988, 1994)	Cronin & Taylor (1992)

DEBATE AND EVIDENCE

Are the gaps real, and are there five dimensions? Three issues remain open. First, difference scores have known psychometric problems: they are often less reliable than their components and correlate with them mechanically. Second, the five-dimension structure does not replicate consistently across industries. Dimensions often merge, and researchers frequently adapt items to context. Third, expectations are dynamic: they rise with good experiences and with competitors' improvements, which makes the comparison standard a moving target. These critiques weaken SERVQUAL as a universal measure, but they do not undermine the gaps model as a managerial diagnostic. The internal gaps (1 to 4) remain a useful way to locate the organizational source of a quality problem.

The downstream evidence is more consistent. Service quality is associated with favourable behavioural intentions, such as loyalty, willingness to pay more, and positive word of mouth (Zeithaml, Berry, and Parasuraman 1996). The **service-profit chain** extends the argument inside the firm: internal service quality drives employee satisfaction and retention, which improves the value delivered to customers, which drives loyalty and profit (Heskett et al. 1994).

28.3.5 Applications in business

BUSINESS CASE

Service quality in banking and airlines. Retail banking was one of the categories in which the gaps model was developed (Parasuraman, Zeithaml, and Berry 1985), and it remains a natural fit: much of what customers value, accurate transactions and prompt problem resolution, is invisible until it fails. A bank that invests in new branches (tangibles) while call-centre waits lengthen (responsiveness) can see satisfaction fall despite rising spending, because it has improved the dimension with the smallest gap. Airlines illustrate Gap 4. Advertising that emphasises premium cabins and attentive service raises expectations for every passenger, including those in economy who will experience delays and crowded boarding. In the United States, the Department of Transportation publishes carriers' on-time performance and complaint statistics, which turns reliability, a dimension customers weight heavily, into a publicly observable attribute. The practical sequence is the same in both industries: measure the customer gap by dimension, trace each large gap to its internal source (knowledge, standards, delivery, or communication), and fix the source rather than the symptom.

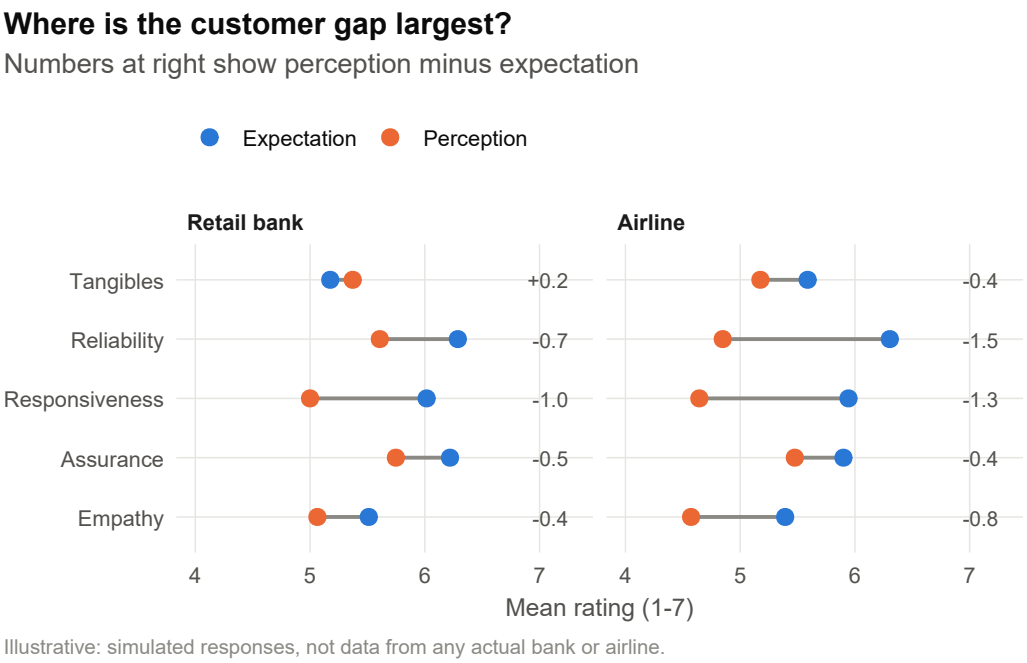


Figure 28.4. Illustrative SERVQUAL gap analysis with simulated survey data (250 respondents per service, 7-point scales). Each line connects the mean perception (orange) with the mean expectation (blue). Longer lines indicate larger customer gaps. The largest stylised gaps are in reliability and responsiveness, the dimensions most directly under operational control.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Advertising intensity that emphasises service features will widen the customer gap when operational capacity is constrained (Gap 4). *H2.* The effect of perceived reliability on loyalty will be stronger than the effect of tangibles. *H3.* Employee satisfaction will mediate the effect of internal service quality on customer-perceived quality.
Measures. SERVQUAL and SERVPERF scales (Parasuraman, Zeithaml, and Berry 1988, Cronin and Taylor 1992), behavioural intentions (Zeithaml, Berry, and Parasuraman 1996), operational metrics (wait time, first-contact resolution, on-time rates) as objective counterparts.
Designs. Matched employee-customer surveys across branches or stores, multilevel models, field experiments on service scripts, and natural experiments from public disclosure of service metrics.

COMMON PITFALL

Running SERVQUAL without the gaps model. A survey that reports gap scores but never asks *why* the gap exists treats a diagnostic tool as a scorecard. The value of the model lies in tracing the customer gap back to one of the four provider gaps, each of which calls for a different remedy: research for Gap 1, standards and design for Gap 2, staffing and training for Gap 3, and communication discipline for Gap 4.

28.4 Customer Equity and Customer Lifetime Value

THEORY CARD

Core claim. Customers are assets whose value equals the discounted stream of profits they are expected to generate. The sum of these values, **customer equity**, is a proxy for firm value and the right yardstick for marketing decisions.

Origins. Marketing science and database marketing. Blattberg and Deighton (1996) proposed the customer equity test. S. Gupta, Lehmann, and Stuart (2004) linked customer value to firm value and Rust, Lemon, and Zeithaml (2004) linked customer equity to marketing investments.

Level of analysis. Individual customer, cohort, and firm.

Key constructs. Customer lifetime value (CLV), margin, retention rate, discount rate, acquisition cost, customer equity, value equity, brand equity (Section 27.3), relationship equity.

Boundary conditions. Works best when customers can be identified and their purchases observed. Estimating retention is harder in noncontractual settings, where defection is unobserved.

Related theories. Market-based assets (Section 4.5), Customer-based brand equity (Section 27.3), Commitment-trust theory (Section 9.6), Service quality gap model, Resource-based theory (Section 4.1).

28.4.1 Origins and core logic

Traditional marketing metrics, such as market share and sales, count transactions. **Customer lifetime value (CLV)** counts relationships. The idea came from direct and database marketers, who could track each customer's purchases and therefore estimate what a customer was worth over time. Blattberg and Deighton (1996) argued that firms should manage marketing by a **customer equity test**: spending on acquisition and retention should be set to maximise the total value of the customer base, not short-term sales. That framing changes practical decisions. Acquiring a customer at a loss can be rational if expected future margins are high enough, and a retention programme that looks expensive per period can be cheap per unit of lifetime value.

S. Gupta, Lehmann, and Stuart (2004) took the idea to the capital markets. Following their formulation, the lifetime value of a customer acquired today is the discounted sum of expected margins, weighted by the probability that the customer is still active, minus the cost of acquiring the customer:

$$\mathrm{CLV} = \sum_{t=0}^{T} \frac{(p_t - c_t)\, r^t}{(1+i)^t} - \mathrm{AC},$$

where $p_t - c_t$ is the margin in period t, r is the per-period retention rate, i is the per-period discount rate, T is the horizon, and AC is the acquisition cost. With a constant margin m and an infinite horizon, the sum is a geometric series with the closed form

$$\mathrm{CLV} = m\,\frac{1+i}{1+i-r} - \mathrm{AC}.$$

The formula makes the theory's key implication visible. Because r sits in the denominator, CLV is *convex* in retention: each additional point of retention is worth more than the last. S. Gupta, Lehmann, and Stuart (2004) valued the customer bases of several public firms and found that a 1% improvement in retention improved firm value by about 5%, compared with about 1% for an equivalent improvement in margin and about 0.1% for acquisition cost. Customer equity, the sum of CLV across current and future customers, then becomes a lower-bound estimate of the value of the firm's operating assets.

Rust, Lemon, and Zeithaml (2004) opened the black box of retention. They decomposed customer equity into three drivers: **value equity** (the customer's objective assessment of quality, price, and convenience), **brand equity** (subjective, emotional appraisal, see customer-based brand equity (Section 27.3) (Keller 1993)), and **relationship equity** (the customer's tendency to stay beyond objective and subjective assessments, built through loyalty programmes and switching costs). Linking each marketing initiative to these drivers, and the drivers to switching probabilities, yields a projected **return on marketing**.

The customer equity perspective also reframes the classic budgeting question of acquisition versus retention. Blattberg and Deighton (1996) argued that each has diminishing returns: the first acquisition dollars reach the most responsive prospects, and the first retention dollars save the customers who were easiest to keep. The optimal allocation equates the marginal contribution of the last dollar spent on each to customer equity. Two consequences follow. A firm that has underinvested in retention will see a high return from shifting money toward it, but the advantage shrinks as spending grows. And because the value of acquiring a customer depends on how long that customer will stay, acquisition and retention budgets cannot be set independently. Raising retention raises the value of every new customer, which justifies more acquisition spending (Table 28.6).

Table 28.6. Drivers of customer equity (after Rust, Lemon and Zeithaml, 2004).

Driver	What it captures	Typical levers	Example metric
Value equity	Objective assessment of what the offering delivers for its cost	Quality, price, convenience	Perceived value ratings, price index
Brand equity	Subjective, emotional appraisal beyond objective value	Advertising, brand experience, ethics	Brand awareness and attitude
Relationship equity	Propensity to stay beyond objective and subjective appraisal	Loyalty programmes, community, switching costs	Retention rate, share of wallet

28.4.2 Computing customer lifetime value in R

The code below implements both forms of the formula for an illustrative subscription business and compares the effect of 1% relative improvements in each input. The numbers are stylised, not estimates for any real firm.

```
# Customer lifetime value with constant margin, retention, and discount rate.
# Margins arrive at the start of each period (t = 0, 1, 2, ...).
clv <- function(margin, retention, discount, acquisition_cost = 0, horizon = Inf) {
  if (is.infinite(horizon)) {
    value <- margin * (1 + discount) / (1 + discount - retention)
  } else {
    t <- 0:horizon
    value <- sum(margin * retention^t / (1 + discount)^t)
```

```
  }
  value - acquisition_cost
}

# Illustrative subscription: monthly margin 6, monthly retention 0.95,
# monthly discount rate 0.008 (about 10% a year), acquisition cost 40
base <- clv(margin = 6, retention = 0.95, discount = 0.008, acquisition_cost = 40)
three_years <- clv(6, 0.95, 0.008, acquisition_cost = 40, horizon = 35)

# Effect of a 1% relative improvement in each input
scenarios <- data.frame(
  change = c("Base case", "Retention +1%", "Margin +1%",
             "Acquisition cost -1%", "Discount rate -1%"),
  clv = c(base,
          clv(6, 0.95 * 1.01, 0.008, 40),
          clv(6 * 1.01, 0.95, 0.008, 40),
          clv(6, 0.95, 0.008, 40 * 0.99),
          clv(6, 0.95, 0.008 * 0.99, 40))
)
scenarios$pct_change <- round(100 * (scenarios$clv / base - 1), 1)
scenarios$clv <- round(scenarios$clv, 2)
round(c(infinite_horizon = base, three_year_horizon = three_years), 2)
```

```
#>   infinite_horizon three_year_horizon
#>              64.28              51.93
```

```
scenarios
```

```
#>                 change   clv pct_change
#> 1            Base case 64.28        0.0
#> 2        Retention +1% 84.70       31.8
#> 3           Margin +1% 65.32        1.6
#> 4 Acquisition cost -1% 64.68        0.6
#> 5    Discount rate -1% 64.41        0.2
```

Even in this toy example the ordering reported by S. Gupta, Lehmann, and Stuart (2004) appears: a 1% improvement in retention raises CLV far more than the same relative change in margin, acquisition cost, or the discount rate. The magnitudes differ from theirs because the inputs differ, which is itself a lesson. Retention sensitivity grows as retention rises, as Figure 28.5 shows.

TRY IT YOURSELF

Use the `clv()` function to answer three questions. (1) At what monthly retention rate does the illustrative subscriber exactly repay the acquisition cost of 40? (2) How much would the firm have to raise the monthly margin to offset a fall in retention from 0.95 to 0.93? (3) Repeat question 1 with a 36-month horizon. Why does the break-even retention rate rise?

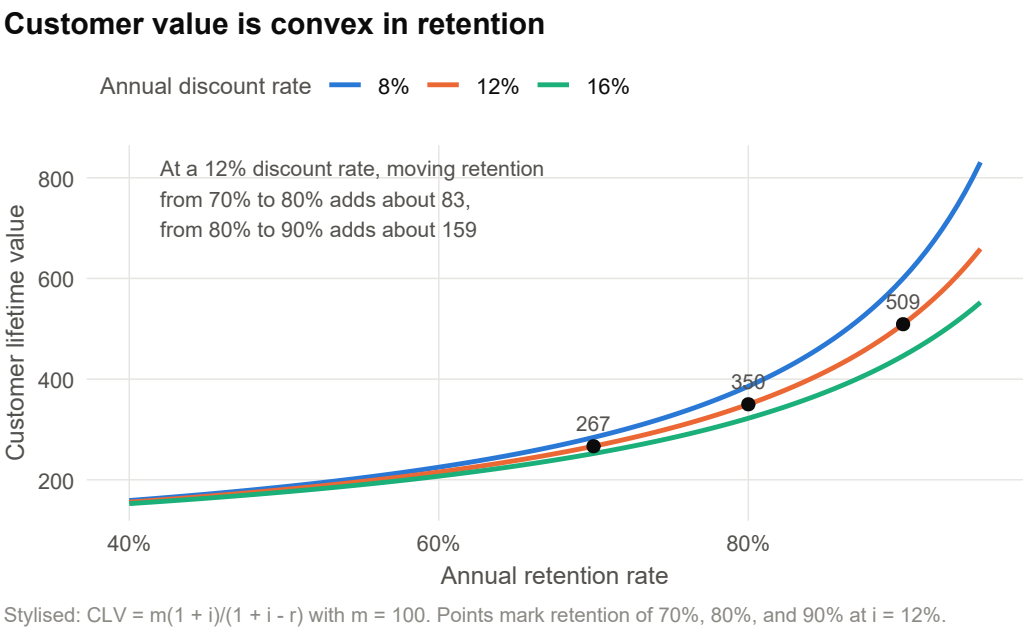


Figure 28.5. Stylised sensitivity of customer lifetime value to the annual retention rate for an annual margin of 100 and three discount rates (infinite horizon, acquisition cost excluded). Value rises slowly at low retention and steeply at high retention, so the same retention gain is worth more to a firm that already retains well.

28.4.3 Evidence and extensions

Two empirical programmes have refined the theory. The first concerns **noncontractual settings**, such as retailing, in which customers do not announce that they have left. Schmittlein, Morrison, and Colombo (1987) developed the Pareto/NBD model to infer from purchase histories how many customers are still "alive", and Fader, Hardie, and Lee (2005) offered the simpler BG/NBD alternative that is now widely used for CLV estimation. The second concerns the link between loyalty and profit. Reinartz and Kumar (2000) studied a catalogue retailer and found that long-life customers were not necessarily the most profitable, which challenges the assumption that retention always pays. Retention is valuable only when retained customers generate margins that exceed the cost of keeping them.

DEBATE AND EVIDENCE

How far can customer metrics go? Critics raise three issues. First, CLV models are only as good as their retention estimates, and small errors in r produce large errors in value because of the convexity shown above. Second, valuing a firm by its customers requires forecasting the acquisition of *future* customers, which reintroduces the uncertainty that customer metrics were supposed to remove. S. Gupta, Lehmann, and Stuart (2004) found that customer-based valuations tracked market values closely for some firms but not for others, such as fast-growing internet firms. Third, a single-minded focus on CLV can lead firms to "fire" low-value customers or to price-discriminate in ways that damage trust and brand equity. Customer equity is a powerful *accounting* lens, but it does not by itself say how to build value.

28.4.4 Applications in business

BUSINESS CASE

Subscription businesses and CLV. Streaming services, software-as-a-service providers, telecom carriers, and meal-kit firms report churn, average revenue per user, and customer-acquisition cost because these quantities map directly onto the CLV formula. The convexity result explains several common practices. Firms offer discounts to customers who try to cancel, because a saved high-tenure customer is worth more than a new one. They track cohort retention curves rather than averages, because early-life churn and late-life churn have different causes. And they compare the ratio of lifetime value to acquisition cost by channel to decide where to spend. The pitfalls are equally common: counting revenue rather than margin, assuming that a promotional cohort will retain like an organic one, and using an implausibly low discount rate. A useful discipline is to report CLV as a range across plausible retention and discount assumptions, as in Figure 28.5, rather than as a single number.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Customers acquired through price promotions will have lower retention and CLV than customers acquired through referrals. *H2.* Improvements in service quality will raise customer equity mainly through the relationship-equity driver. *H3.* Firms that disclose customer metrics will have lower analyst forecast dispersion.
Data and methods. Transaction logs and CRM data, BG/NBD and Pareto/NBD models for non-contractual settings (Fader, Hardie, and Lee 2005, Schmittlein, Morrison, and Colombo 1987), survival models for contractual churn, brand-switching matrices (Rust, Lemon, and Zeithaml 2004), and firm-level valuation from public disclosures (S. Gupta, Lehmann, and Stuart 2004).

COMMON PITFALL

Confusing revenue with value and averages with individuals. CLV uses *margins*, not revenue, and should be computed by cohort or segment. An average CLV hides the heterogeneity that makes customer equity actionable, and a retention rate estimated in one period can change sharply when the competitive environment or the acquisition channel changes.

28.5 Hierarchy of Effects

THEORY CARD

Core claim. Advertising moves consumers through a sequence of mental stages, from awareness and knowledge through liking, preference, and conviction to purchase. Each stage is a prerequisite for the next and can serve as an intermediate measure of advertising effectiveness.
Origins. Advertising research. Lavidge and Steiner (1961), drawing on a long tradition of sales-stage models such as AIDA (attention, interest, desire, action). Reviewed critically by Barry and Howard (1990) and Vakratsas and Ambler (1999).
Level of analysis. Individual consumer, aggregated to campaign metrics.

Key constructs. Cognitive stage (awareness, knowledge), affective stage (liking, preference), conative stage (conviction, purchase), involvement.
Boundary conditions. The "learn, feel, do" order is most plausible for high-involvement, considered purchases. For low-involvement and habitual purchases, experience often precedes attitude.
Related theories. Elaboration likelihood model (Section 26.1), Mere exposure effect (Section 29.1), Information processing model (Section 31.1), Theory of planned behavior (Section 10.3), SOR framework (Section 31.5).

28.5.1 Core logic

Lavidge and Steiner (1961) proposed that advertising should be evaluated not only by sales, which are influenced by many factors, but by its ability to move consumers up a series of steps: **awareness**, **knowledge**, **liking**, **preference**, **conviction**, and **purchase**. The six steps map onto the three classic components of attitude: the *cognitive* (thoughts), the *affective* (feelings), and the *conative* (motives and behaviour). The model had two lasting consequences. It justified intermediate advertising objectives, such as raising awareness, which could be measured before sales effects materialised. And it gave rise to the **marketing funnel**, the most widely used planning heuristic in advertising and digital marketing (Figure 28.6).

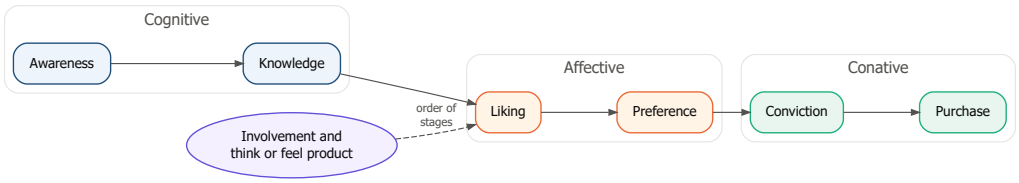


Figure 28.6. The hierarchy of effects. Advertising is assumed to move consumers sequentially through cognitive, affective, and conative stages. Involvement and product type moderate whether the sequence holds. Redrawn from Lavidge and Steiner (1961), with the moderator drawn from Vaughn (1980).

28.5.2 Evidence and critiques

The sequence was questioned almost from the start. Research on low-involvement learning suggested that for many inexpensive, frequently purchased products, consumers buy first and form attitudes afterwards, and exposure alone can raise liking without much thought (see the mere exposure effect (Section 29.1)). Vaughn (1980) organised these alternatives in the **FCB grid**, which crosses involvement (high versus low) with the nature of the decision (thinking versus feeling) and assigns a different order of stages to each quadrant (Table 28.7).

Table 28.7. The hierarchy of effects and its main alternatives.

Model	Assumed sequence	Where it fits best	Source
Lavidge-Steiner hierarchy	Awareness, knowledge, liking, preference, conviction, purchase	High-involvement, considered purchases	Lavidge & Steiner (1961)
FCB grid: informative	Learn, feel, do	High involvement, thinking (insurance, appliances)	Vaughn (1980)
FCB grid: affective	Feel, learn, do	High involvement, feeling (fashion, perfume)	Vaughn (1980)

Table 28.7. The hierarchy of effects and its main alternatives. *(continued)*

Model	Assumed sequence	Where it fits best	Source
FCB grid: habitual	Do, learn, feel	Low involvement, thinking (household staples)	Vaughn (1980)
FCB grid: satisfaction	Do, feel, learn	Low involvement, feeling (snacks, soft drinks)	Vaughn (1980)
Integrative space	Cognition, affect, and experience interact without a fixed order	All categories, filtered by motivation and ability	Vakratsas & Ambler (1999)

The most influential critique came from Vakratsas and Ambler (1999), who reviewed more than 250 studies on how advertising works. They found little support for a fixed temporal sequence of effects. They proposed instead that advertising effects are best understood in a space defined by **cognition**, **affect**, and **experience**, with consumers' motivation and ability acting as filters, and with effects varying by product category, competitive context, and consumers' prior experience with the brand.

DEBATE AND EVIDENCE

Is the funnel dead? Defenders argue that the hierarchy remains a useful *planning* model even if it is a poor *process* model: firms still need to create awareness before consideration, and stage metrics are easier to track than sales. Critics respond that treating stages as sequential leads managers to over-invest in awareness, to ignore post-purchase experience as a driver of attitudes, and to misattribute sales to the last touchpoint in a digital funnel. Barry and Howard (1990) concluded that the hierarchy has intuitive and practical appeal but weak empirical support for strict sequencing. The productive middle ground is to keep the *constructs*, which are measurable and meaningful, while dropping the assumption that they occur in a fixed order.

Digital marketing has given the hierarchy a second life and a new set of problems. Online funnels track impressions, clicks, visits, add-to-cart events, and purchases, which look like a measurable hierarchy of effects. Yet these stages are behavioural traces, not mental states, and they suffer from selection: people who were already likely to buy are more likely to click. Attributing a purchase to the last ad clicked, or to each stage in proportion to its position, assumes the sequential causal structure the hierarchy proposes. Randomised holdout experiments, in which some consumers are prevented from seeing a campaign, are the most reliable way to test whether each stage actually adds incremental effect.

28.5.3 Applications in business

BUSINESS CASE

Illustrative scenario: diagnosing a campaign with stage metrics. A consumer-electronics brand launches a new pair of earbuds with a television and online video campaign. Brand tracking shows prompted awareness rising sharply, but consideration and purchase intent barely move, and sales lag. A strict hierarchy reading would recommend more reach. A Vakratsas and Ambler reading asks a different question: is the bottleneck cognitive (consumers do not know why the product is better), affective (they do not like the brand), or experiential (they have tried a competitor and are satisfied)? If reviews and retail demos show that people

who try the earbuds rate them highly, the binding constraint is experience, and in-store trials or a generous return policy may outperform additional advertising. The stage constructs remain valuable as diagnostics. What changes is the assumption about the order in which they must be built.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Advertising effects on purchase will be mediated by awareness and attitude for high-involvement categories, but will operate mainly through trial and post-purchase evaluation for low-involvement categories. *H2.* Exposure frequency will raise brand liking even without changes in brand knowledge in low-involvement categories. **Designs.** Brand-tracking panels linked to purchase data, geo-randomised advertising experiments, and path models that compare sequential and non-sequential specifications.

COMMON PITFALL

Treating funnel metrics as causal stages. Correlated movements in awareness, liking, and purchase do not show that one causes the next. Purchase and experience can drive attitudes, and the same advertising can affect several stages at once. Test the order rather than assume it.

28.6 Product Life Cycle

THEORY CARD

Core claim. Products pass through a predictable sequence of stages, introduction, growth, maturity, and decline, each with characteristic patterns of sales, profit, competition, and appropriate strategy.
Origins. Marketing strategy. Popularised by T. Levitt (1965). Critically assessed by Dhalla and Yuspeh (1976) and Day (1981).
Level of analysis. Product class or product form (weaker for individual brands).
Key constructs. Introduction, growth, maturity, decline, takeoff, slowdown, product class versus product form versus brand.
Boundary conditions. The curve describes product classes better than brands, and its shape and timing vary widely. It is partly endogenous to marketing decisions.
Related theories. Bass model (Section 10.2), Diffusion of innovations (Section 10.1), Dominant design (Section 12.5), Disruptive innovation (Section 12.2), Competitive dynamics (Section 11.3).

28.6.1 Core logic

T. Levitt (1965) urged managers to "exploit" the product life cycle by anticipating each stage rather than reacting to it. In the **introduction** stage, sales grow slowly as buyers learn about the product, and profits are negative because of development and launch costs. In **growth**, sales accelerate as adoption spreads and competitors enter. In **maturity**, sales level off, competition intensifies, and price pressure squeezes margins. In **decline**, sales fall as substitutes and changing tastes erode demand. Levitt's practical advice was to plan, while a product is still growing, the moves that will

extend its life: more frequent use by current users, more varied use, new users, and new uses for the basic material.

The PLC is the managerial counterpart of the diffusion process studied in diffusion of innovations (Section 10.1) (Rogers 2003) and formalised in the Bass model (Section 10.2) (F. M. Bass 1969): the growth stage corresponds to the steep part of the adoption S-curve, and maturity to the period when most potential adopters have adopted and sales depend on replacement and repeat purchase. Figure 28.7 shows the stylised curve.

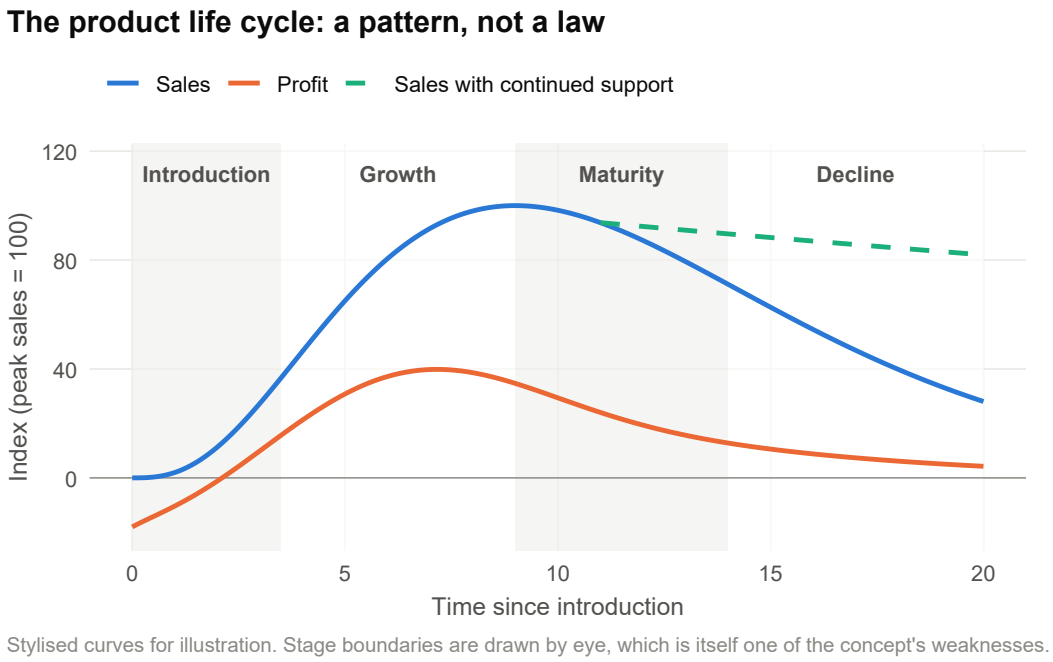


Figure 28.7. Stylised product life cycle for a product class. Sales (blue) follow an S-shaped rise and eventual decline. Profit (orange) is negative at launch and peaks before sales do, as competition compresses margins. The dashed line shows the alternative emphasised by Dhalla and Yuspeh (1976): continued support or repositioning can sustain a mature product, so decline is partly a consequence of strategy.

Table 28.8 presents the conventional characteristics of product life cycle stages, as described in Levitt (1965) and later textbook treatments.

Table 28.8. Conventional characteristics of product life cycle stages, as described in Levitt (1965) and later textbook treatments.

Stage	Sales	Profits	Competition	Typical marketing emphasis
Introduction	Low, slow growth	Negative	Few rivals	Build awareness and trial, educate the market
Growth	Rapid growth	Rising, often peaking late in the stage	Entry increases	Build preference, expand distribution
Maturity	Slowing, then flat	Declining as price competition grows	Many rivals, shake-out	Defend share, differentiate, extend uses
Decline	Falling	Low or negative	Exit by weaker rivals	Harvest, reposition, or divest

28.6.2 Evidence and critiques

DEBATE AND EVIDENCE

Forget the product life cycle? Dhalla and Yuspeh (1976) published one of the most cited attacks on a marketing concept. They argued that the PLC lacks empirical support at the brand level, that its stages cannot be identified in advance, and that its prescriptions can be self-fulfilling. A manager who believes a brand has entered decline cuts its support, and sales then decline *because* support was withdrawn. The PLC, in their view, is a *dependent* variable shaped by marketing decisions, not an independent force that should dictate them. Day (1981) offered a more balanced review. He distinguished **product classes**, **product forms**, and **brands**, and argued that the concept is most defensible for product forms, least defensible for brands, and useful mainly as a framework for thinking about how market conditions evolve rather than as a forecasting tool.

Later empirical research has replaced the generic curve with models of specific **turning points**. Golder and Tellis (1997) modelled the **takeoff** of new consumer durables, the point at which slow early sales give way to rapid growth, and Golder and Tellis (2004) studied the **slowdown** that ends the growth stage. They explained the PLC's sharp turns partly through informational cascades, in which consumers' adoption decisions follow the observed decisions of others. These studies keep the PLC's core insight, that markets pass through qualitatively different phases, while making the phases measurable and predictable.

28.6.3 Applications in business

BUSINESS CASE

Extending mature products. Levitt's strategy of finding new uses has a classic illustration in Arm & Hammer baking soda. In the early 1970s the brand promoted placing an open box in the refrigerator to absorb odours, giving a mature, slow-growing product a new and recurring use. The revival of vinyl records since the late 2000s shows the limits of PLC determinism from another angle: a product form widely assumed to be in terminal decline found new demand among collectors and younger buyers, supported by record labels and retailers that kept supply alive. Both cases support the Dhalla and Yuspeh critique. Whether a product is "mature" or "declining" depends partly on how managers define its market and whether they continue to invest. For forecasting genuinely new categories, managers are better served by diffusion models such as the Bass model (Section 10.2) than by a generic curve.

RESEARCH TOOLKIT

Research uses. Model stage transitions as events (hazard models of takeoff and slowdown), estimate diffusion parameters by product form, or test whether marketing-mix elasticities differ across stages. *Sample hypothesis:* advertising elasticity will be higher in the growth stage than in maturity, whereas price elasticity will be higher in maturity. **Data.** Category sales histories, scanner data, and industry association statistics.

COMMON PITFALL

Reading the stage, then cutting the budget. Labelling a brand as "in decline" on the basis of a few flat quarters, then harvesting it, can turn a temporary slowdown into a permanent one. Diagnose *why* sales are flattening (saturation, substitution, or under-investment) before concluding that the product has entered a new stage.

28.7 Marketing Mix

THEORY CARD

Core claim. A marketing program is a combination of controllable decision variables that managers blend to reach target customers. The best-known classification is McCarthy's 4Ps: product, price, place, and promotion.
Origins. Marketing management. Borden (1964) popularised the "marketing mix", McCarthy (1960) condensed it into the 4Ps, and Booms and Bitner (1981) added three Ps for services. Critically reviewed by Constantinides (2006).
Level of analysis. Firm and offering.
Key constructs. Product, price, place (distribution), promotion (communication), people (participants), physical evidence, process.
Boundary conditions. An organizing framework rather than a causal theory. Most suited to consumer goods. Less complete for services, relationships, and digital platforms.
Related theories. Market orientation, Service-dominant logic, Product life cycle, Five forces (Section 31.4).

28.7.1 Origins and core logic

Borden (1964) described the marketing manager as a "mixer of ingredients", a phrase he credited to his Harvard colleague James Culliton, who blends a set of elements into a program suited to the firm's market. Borden's list was long, covering product planning, pricing, branding, channels of distribution, personal selling, advertising, promotions, packaging, display, servicing, physical handling, and fact finding and analysis. McCarthy (1960) condensed the idea into four categories, **product**, **price**, **place**, and **promotion**, a framework that became the organizing structure of introductory marketing textbooks.

Services researchers argued that four Ps were not enough. Because services are produced and consumed at the same time, the people who deliver them, the setting in which they are delivered, and the steps customers go through are part of the offering. Booms and Bitner (1981) therefore added **participants** (usually called people), **physical evidence**, and **process**, yielding the **7Ps**. The physical-evidence element connects to research on **servicescapes**, the designed physical environment of service delivery (Bitner 1992) (Table 28.9).

Table 28.9. The 4Ps and the services extension to 7Ps, illustrated for an airline.

Element	Core decision	Airline illustration	Origin
Product	What is offered: features, quality, brand, service level	Routes, cabin classes, baggage allowance	McCarthy (1960)
Price	What customers pay and how	Fare classes, dynamic pricing, fees	McCarthy (1960)

Table 28.9. The 4Ps and the services extension to 7Ps, illustrated for an airline. *(continued)*

Element	Core decision	Airline illustration	Origin
Place	How the offering reaches customers	Direct website and app, travel agents, alliances	McCarthy (1960)
Promotion	How the firm communicates	Advertising, loyalty-programme offers	McCarthy (1960)
People	Who delivers and co-produces the service	Gate agents, cabin crew, other passengers	Booms & Bitner (1981)
Physical evidence	Tangible cues to quality	Lounge design, cabin interior, uniforms	Booms & Bitner (1981)
Process	How the service is delivered	Check-in, boarding sequence, rebooking	Booms & Bitner (1981)

28.7.2 Evidence and critiques

As a classification scheme, the marketing mix is not "tested" in the way a causal theory is. Its empirical legacy lies instead in research on how the elements work: price and advertising elasticities, distribution effects, and the interactions among them. The critiques concern whether the classification directs managerial attention well.

DEBATE AND EVIDENCE

Is the 4P framework obsolete? Constantinides (2006) reviewed critiques of the mix across consumer, relationship, services, retail, and industrial marketing, and e-commerce. A common theme is that the 4Ps are **internally oriented**: they list what the firm controls rather than what customers experience, they treat the customer as a passive target, and they say little about relationships, interaction, or co-creation. Relationship marketing scholars argued that retention and trust do not fit in any P (see commitment-trust theory (Section 9.6) (R. M. Morgan and Hunt 1994)). S-D logic goes further and questions the idea that value is something the firm "mixes" at all (Vargo and Lusch 2004). Defenders reply that the 4Ps were never intended as a theory of value. They are a checklist that ensures no major decision is forgotten, and their simplicity is the source of their longevity. Numerous alternative lists have been proposed, but none has displaced the original.

Digital markets stretch each element. *Product* increasingly includes software that changes after purchase. *Price* is set by algorithms that respond to demand in real time and by subscription and freemium structures that separate the price of access from the price of use. *Place* is often a platform owned by someone else, such as an app store or marketplace, which shifts bargaining power toward the intermediary. And *promotion* now includes content created by customers themselves, which the firm can encourage but not control. None of these changes makes the mix irrelevant. They do make the boundary between controllable and uncontrollable variables less clear, which is precisely the concern raised by critics who argue that marketing is a process of interaction rather than a set of levers.

28.7.3 Applications in business

BUSINESS CASE

Designing a coherent mix. The practical power of the marketing mix lies in **consistency** among elements. A premium price must be supported by product quality, selective distribution, and communication that justifies the price. Starbucks, under Howard Schultz, articulated its stores as a "third place" between home and work. In 7P terms, that positioning was delivered as much through physical evidence (store design), people (baristas), and process (how orders are taken and drinks prepared) as through product and price. A common failure is an inconsistent mix, such as a luxury product sold through discount channels or a low-cost service that promises personal attention it cannot deliver. The latter is also Gap 4 in the gaps model.

RESEARCH TOOLKIT

Research uses. Marketing-mix models (time-series regressions of sales on price, advertising, distribution, and promotion) estimate element-level elasticities and support budget allocation. *Sample hypothesis:* the effect of advertising on sales will be stronger when distribution breadth is high (complementarity among mix elements). **Designs.** Store-week scanner data, geo-experiments, and hierarchical Bayesian models that pool elasticities across markets.

COMMON PITFALL

Optimising each P separately. Separate budget owners for pricing, promotion, and distribution often optimise their own lever at the expense of the whole. Interactions among elements are the point of a *mix*. Model them explicitly, and evaluate the program from the customer's perspective as well as the firm's.

28.8 Chapter Summary

KEY TAKEAWAYS

- **Market orientation**, the generation, dissemination, and use of market intelligence, is positively associated with performance, mainly through innovation, quality, and loyalty (Kohli and Jaworski 1990, Narver and Slater 1990, Kirca, Jayachandran, and Bearden 2005). A proactive orientation guards against the trap of serving only current customers (Narver, Slater, and MacLachlan 2004).
- **Service-dominant logic** reframes all exchange as service and locates value in use, co-created by customers (Vargo and Lusch 2004, 2016). It is a powerful lens for servitization and outcome-based business models, but it needs mid-range theories to generate testable predictions.
- The **gaps model** traces the customer's quality gap to four internal gaps in knowledge, standards, delivery, and communication (Parasuraman, Zeithaml, and Berry 1985). SERVQUAL's difference scores are contested, and performance-only measures often predict better (Cronin and Taylor 1992).
- **Customer equity** values customers as assets. CLV is convex in retention, so retention

improvements often matter more than margin or acquisition-cost improvements (Blattberg and Deighton 1996, S. Gupta, Lehmann, and Stuart 2004, Rust, Lemon, and Zeithaml 2004).
- The **hierarchy of effects** remains a useful set of measurable constructs, but evidence does not support a fixed sequence from cognition to affect to behaviour (Lavidge and Steiner 1961, Vakratsas and Ambler 1999).
- The **product life cycle** describes how markets evolve, but stage boundaries are hard to identify and the curve is partly shaped by managers' own decisions (T. Levitt 1965, Dhalla and Yuspeh 1976).
- The **marketing mix** organises controllable decisions. The 7Ps extend it to services, and critics argue that it remains firm-centric (Borden 1964, Booms and Bitner 1981, Constantinides 2006).

Table 28.10 presents the comparison of the chapter's theories.

Table 28.10. Comparison of the chapter's theories.

Theory	Unit of analysis	Core mechanism	Main critique	Typical business use
Market orientation	Business unit	Intelligence generation, dissemination, responsiveness	Common-method bias, can favour current customers	Organizational audits, cross-functional routines
Service-dominant logic	Service ecosystem	Value co-creation in use	Broad and hard to falsify	Servitization, outcome-based contracts
Gaps model of service quality	Firm and customer	Expectations versus perceptions, four provider gaps	Difference-score psychometrics	Service diagnostics and improvement
Customer equity	Customer and firm	Discounted margins weighted by retention	Sensitive to retention estimates	Subscription metrics, acquisition budgets
Hierarchy of effects	Consumer	Sequential cognitive, affective, conative stages	Little support for fixed order	Campaign objectives and funnel tracking
Product life cycle	Product class	Diffusion, saturation, competition	Self-fulfilling and descriptive	Stage-based planning, portfolio reviews
Marketing mix	Offering	Consistent blend of controllable variables	Internally oriented	Program design and budget allocation

28.9 Discussion Questions

1. Market orientation and disruptive innovation seem to give opposite advice about listening to customers. Reconcile the two using the distinction between responsive and proactive market orientation, and propose a measure that would let you test your reconciliation.
2. Service-dominant logic has been described as a perspective rather than a theory. Derive two falsifiable hypotheses from it for an equipment manufacturer considering a shift to outcome-based contracts.
3. A hotel chain's SERVQUAL survey shows the largest gap in responsiveness. Using the gaps model, list the internal sources that could produce this gap and the data you would collect to distinguish among them.
4. Using the CLV function in this chapter, find the retention rate at which a subscription business

with a monthly margin of 6 and an acquisition cost of 40 breaks even. How sensitive is your answer to the discount rate?

5. Design a field experiment that tests whether the hierarchy of effects holds for a low-involvement grocery product. What pattern of results would support Vakratsas and Ambler's critique?
6. A consumer-goods firm wants to cut support for a brand it believes has entered the decline stage. What evidence would you request before agreeing, and how does the Dhalla and Yuspeh critique inform your request?

28.10 Further Reading

- Kohli and Jaworski (1990): the defining statement of market orientation as a behavioural construct, with propositions that still organise the field.
- Kirca, Jayachandran, and Bearden (2005): the meta-analytic synthesis of what market orientation does and does not predict.
- Vargo and Lusch (2004): the article that launched service-dominant logic. Read it with Vargo and Lusch (2016) to see how the framework evolved.
- Parasuraman, Zeithaml, and Berry (1985): the gaps model. Pair it with Cronin and Taylor (1992) for the measurement debate.
- S. Gupta, Lehmann, and Stuart (2004): a clear demonstration that customer metrics can value a firm, and a template for CLV analysis.
- Vakratsas and Ambler (1999): the essential review of how advertising works, and why the hierarchy is too simple.
- Dhalla and Yuspeh (1976): a short, forceful critique that every user of the product life cycle should read.
- Constantinides (2006): a comprehensive review of critiques of the marketing mix.

Part VI

Effects, Phenomena, and Frameworks

Chapter 29

Named Effects in Behavioral Science

A software company adds a "Pro" plan that almost no one buys, and sales of its "Business" plan rise. A furniture retailer sells flat-packed products that customers must assemble themselves, and customers seem to like them more for the trouble. A soft-drink brand that has advertised the same slogan for years worries that its campaign has "worn out", while a challenger worries that no one has heard of it at all. A board appoints its first female chief executive just as the firm enters a crisis. Behavioural scientists have a name for each of these patterns, the **decoy effect**, the **IKEA effect**, **mere exposure** and **wear-out**, the **glass cliff**, and much of applied behavioural science consists of recognising such patterns and knowing when they will and will not appear.

This chapter is a field guide to named effects that recur in management, marketing, and organisational research. A named effect is not the same thing as a theory. An effect is a **reliable empirical regularity**: a pattern linking a manipulation or condition to an outcome. A theory explains *why* the regularity occurs and therefore predicts *when* it will occur. Many effects in this chapter have several candidate explanations, and part of the craft of using them is connecting each effect to the mechanisms developed elsewhere in the book: processing fluency (Section 23.8), prospect theory (Chapter 24), signaling (Section 8.2), and social identity (Section 22.2). Social facilitation (Section 22.6), a close relative of several effects here, is covered with social influence in Social facilitation (Section 22.6) (Table 29.1).

Table 29.1. Catalogue of named effects covered in this chapter.

Effect	One-line definition	Key source	Business use	Evidence strength
Mere exposure	Repeated exposure to a stimulus increases liking for it	Zajonc (1968), Bornstein (1989)	Ad frequency, brand familiarity	Strong (meta-analytic)
Warm-glow heuristic	Positive stimuli feel more familiar, even without prior exposure	Monin (2003)	Brand familiarity judgments, recognition research	Moderate (focused programme)
Goldilocks (inverted-U)	Liking peaks at moderate novelty or complexity	Berlyne (1970), McCall & McGhee (1977)	Product and ad design, innovation newness	Moderate, shape depends on measure
Coolidge	Renewed sexual interest with a novel partner (non-human animals)	Wilson, Kuehn & Beach (1963), Dewsbury (1981)	Analogy for variety seeking and novelty in product lines	Strong in animals, analogical for consumers
Bandwagon / snob / Veblen	Demand rises or falls with others' consumption or with price itself	Leibenstein (1950)	Luxury pricing, scarcity, network goods	Strong theoretical basis, mixed measurement

Table 29.1. Catalogue of named effects covered in this chapter. *(continued)*

Effect	One-line definition	Key source	Business use	Evidence strength
Glass cliff	Women and minorities are disproportionately appointed to precarious leadership roles	Ryan & Haslam (2005)	Succession and board decisions	Moderate, context-dependent
IKEA effect	Labour invested in creating a product increases its valuation	Norton, Mochon & Ariely (2012)	Self-assembly, co-creation, customisation	Moderate
Decoy (attraction)	Adding a dominated option increases choice of the option that dominates it	Huber, Payne & Puto (1982)	Pricing tiers and product line design	Contested, strongest with numeric attributes
Anchoring	Estimates are biased toward an initially considered number	Tversky & Kahneman (1974)	Reference prices, negotiation, forecasting	Strong (multi-lab replication)
Beauty in averageness	Averaged or prototypical stimuli are judged attractive	Langlois & Roggman (1990)	Product design typicality	Strong for faces, moderate for products
Pratfall	A minor blunder makes a highly competent person more likeable	Aronson, Willerman & Floyd (1966)	Brand candour, leader communication	Limited, moderated by observer and context

The evidence-strength column is a summary judgment, not a formal rating. It reflects the size and consistency of the literature, the existence of meta-analyses or multi-laboratory replications, and known disputes about robustness. Section 29.12 returns to the question of how much confidence to place in named effects.

29.1 Mere Exposure Effect

THEORY CARD

Core claim. Repeated exposure to a stimulus, without any reinforcement, increases liking for it.

Origins. Social psychology, Zajonc (1968), meta-analysed by Bornstein (1989) and, with a focus on the shape of the exposure-liking curve, by Montoya et al. (2017).

Level of analysis. Individual perceiver.

Key constructs. Exposure frequency, stimulus complexity, exposure duration, recognition, familiarity, liking, boredom (tedium).

Boundary conditions. Strongest for initially neutral or unfamiliar stimuli, brief exposures, and complex stimuli. Weak or reversed for initially disliked stimuli and after many exposures, when tedium sets in.

Related theories. Processing fluency (Section 23.8), Warm-glow heuristic, Goldilocks effect, Elaboration likelihood model (Section 26.1).

29.1.1 Core logic and evidence

Zajonc (1968) proposed that "mere repeated exposure" of an individual to a stimulus is a sufficient condition for enhancing his or her attitude toward it. He combined correlational evidence, words that appear more frequently in natural language tend to have more positive meanings, with experiments in which participants saw nonsense words, Chinese-like characters, or photographs of faces between zero and 25 times and then rated how much they liked them or how "good" their meaning was. Ratings rose with frequency of exposure. The finding was striking because it required no reward, no persuasive argument, and no conscious evaluation: frequency alone was enough.

Later work showed that the effect does not require conscious recognition. Kunst-Wilson and Zajonc (1980) presented irregular polygons so briefly that participants could not recognise them above chance later. Nonetheless, when asked which of two polygons they *liked* more, participants tended to prefer the ones they had seen before. Bornstein (1989) reviewed the first two decades of research and concluded that the effect is robust across stimuli and settings. His meta-analysis also identified moderators that remain central to applied work: the effect is stronger for brief exposures and for stimuli that are not consciously recognised, stronger for complex than for simple stimuli, and tends to level off after a moderate number of exposures, with very long exposure sequences producing smaller effects.

29.1.2 Mechanism

Two broad explanations dominate. The **fluency** account holds that prior exposure makes a stimulus easier to perceive and process, and that this ease is experienced as a mildly positive feeling that is attributed to the stimulus itself (Reber, Schwarz, and Winkielman 2004). The account explains why the effect is stronger when people do not recognise the stimulus: if people realise that a stimulus feels easy *because* they have seen it before, they discount the feeling. The **two-factor** account adds a second process. Exposure produces positive habituation (reduced uncertainty and increasing fluency) but also **tedium**: boredom that accumulates with repetition. Liking is the net of the two, so it first rises and then falls. Cacioppo and Petty (1979) documented the pattern for persuasive messages: agreement with a message rose from one to three presentations and then fell at five, while counter-arguments followed the opposite pattern. Figure 29.1 illustrates the two-factor logic.

Table 29.2 summarises the moderators most relevant to practice.

Table 29.2. Moderators of the mere exposure effect (synthesised from Bornstein, 1989, and the two-factor literature).

Moderator	Direction of effect	Likely mechanism	Managerial implication
Number of exposures	Rises, then levels off or declines	Habituation versus tedium	Plan frequency caps, monitor wear-out
Stimulus complexity	Larger effect for complex stimuli	Complex stimuli take longer to become boring	Richer creative executions sustain repetition
Exposure duration / awareness	Larger for brief, unrecognised exposures	Fluency is misattributed when its source is not noticed	Background placements can build familiarity
Initial valence	Weak or reversed for initially disliked stimuli	Repetition can intensify an existing reaction	Do not repeat a disliked execution
Variation across exposures	Varied executions delay wear-out	Reduces tedium while preserving the core stimulus	Rotate creative around a constant brand cue

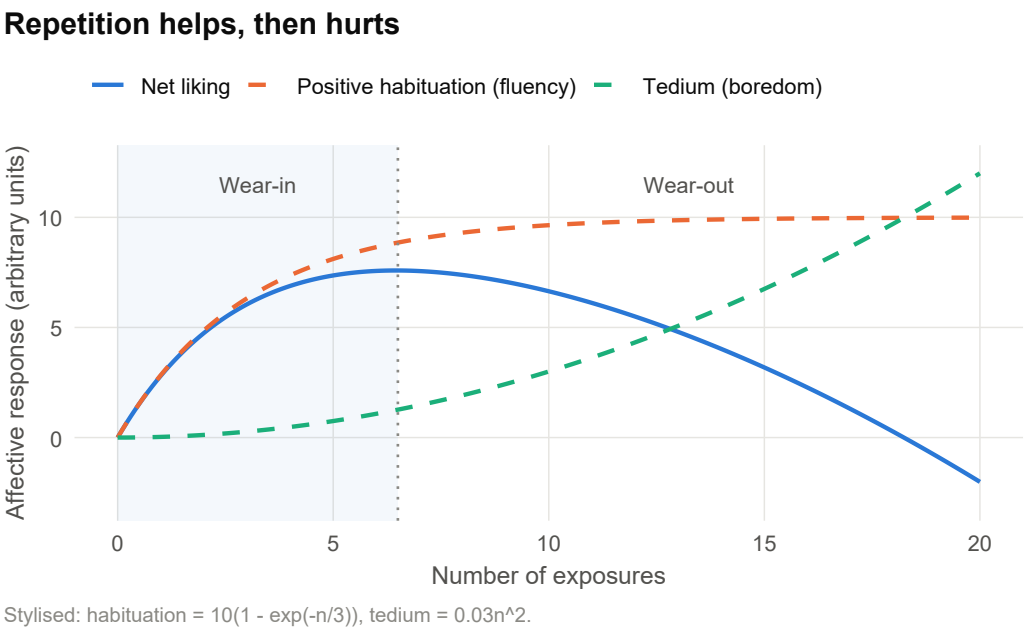


Figure 29.1. Stylised two-factor model of repetition. Positive habituation rises quickly and saturates, tedium accumulates with repetition. Net liking first rises (wear-in) and then declines (wear-out). Curves are illustrative functions, not estimates.

BUSINESS CASE

Brand familiarity and ad repetition: wear-in and wear-out. Media planners work with the two-factor curve every day, even when they do not call it that. A new brand needs **wear-in**: enough exposures for consumers to recognise the brand and to process its logo and packaging fluently. An established campaign risks **wear-out**: after many exposures, the same execution bores or irritates, and consumers begin to counter-argue. Practitioners respond in three ways that follow directly from the moderators above. They set **frequency caps** in digital campaigns to limit how often one person sees an advertisement. They **rotate creative executions** that vary the surface while holding constant the brand cues that should become familiar, and they invest in **distinctive brand assets**, colours, shapes, sonic logos, that accumulate familiarity across many different executions. The mere exposure effect also cautions against a naive reading of brand-tracking data: familiarity can raise liking without any change in beliefs about the product, so rising preference is not necessarily evidence that the message was understood.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* Liking for a new package design will increase over the first several exposures and decline thereafter, with the peak occurring later for more complex designs.
- *H2.* Varying advertising executions while holding the brand cue constant will delay wear-out relative to repeating a single execution.

Designs. Repeated-exposure experiments with incidental exposure tasks. Field data linking advertising frequency to brand metrics with exposure-level variation. Response-latency mea-

sures of fluency.
Measures. Liking ratings, implicit preference measures, recognition, processing speed, brand-tracking familiarity and favourability.

29.2 Warm-Glow Heuristic

THEORY CARD

Core claim. The positive valence of a stimulus increases its perceived familiarity, even when the person has never encountered it before. Familiarity breeds liking, but liking also breeds a feeling of familiarity.
Origins. Social cognition, Monin (2003).
Level of analysis. Individual perceiver.
Key constructs. Stimulus valence (attractiveness, positivity), perceived familiarity, recognition judgments, misattribution of affect.
Boundary conditions. Most pronounced when people have little diagnostic memory information and must rely on feelings to judge whether something is familiar.
Related theories. Mere exposure, Discrepancy attribution hypothesis (Section 23.9), Feelings-as-information theory (Section 23.7), Processing fluency (Section 23.8).

29.2.1 Core logic and evidence

The mere exposure effect runs from familiarity to liking. Monin (2003) asked whether the arrow can also run in the opposite direction. In a series of studies he found that attractive faces were judged more familiar than less attractive faces, and positive words were more likely than neutral or negative words to be mistakenly "recognised" as having appeared earlier, even when none of the stimuli had been shown before. He called this the **warm-glow heuristic**: people experience a positive affective glow and use it, among other cues, to infer that a stimulus is familiar. The account is a form of **misattribution**: a feeling that arises from one source (the stimulus's positivity) is attributed to another (past exposure).

The warm-glow heuristic completes a loop. Familiar things feel easy to process and are liked. Liked things feel familiar. Each process can reinforce the other, which helps explain why well-liked brands are often over-estimated in familiarity and why it is difficult to disentangle familiarity from liking in survey measures.

COMMON PITFALL

Confusing the warm-glow heuristic with warm-glow giving. The phrase "warm glow" is used for two unrelated ideas. Monin (2003)'s **warm-glow heuristic** is a cognitive effect in which positive stimuli feel familiar. Andreoni (1990)'s **warm-glow giving** is an economic model of charitable behaviour in which donors derive utility from the act of giving itself, not only from the total amount of the public good provided: a form of **impure altruism**. Andreoni's model explains, for example, why government contributions to a public good do not fully crowd out private donations. Citing one when the argument requires the other is a common error in student and practitioner writing. Table 29.3 distinguishes them.

Table 29.3. Two different "warm glows".

Feature	Warm-glow heuristic	Warm-glow giving
Discipline	Social cognition / psychology	Public economics
Key source	Monin (2003)	Andreoni (1990)
Phenomenon	Positive stimuli are judged more familiar	People enjoy the act of giving itself
Mechanism	Misattribution of positive affect to familiarity	Utility from one's own contribution (impure altruism)
Typical outcome	Familiarity and recognition judgments	Donations, crowding-out of private giving
Business application	Interpreting brand-familiarity and recognition data	Designing donation appeals and cause marketing

BUSINESS CASE

Measuring brand awareness. Brand-tracking surveys often ask respondents whether they have heard of a brand (aided awareness) and how they feel about it. The warm-glow heuristic implies that the two answers are not independent: respondents who like a brand's name, logo, or category may report familiarity with it even without prior exposure. One practical safeguard is to include **foil brands**, plausible but fictitious names, in awareness questions. The rate of "recognition" of foils estimates the false-alarm rate, and positive-sounding foils are likely to attract more false recognition than neutral ones. Comparing real brands against valence-matched foils yields a cleaner awareness estimate.

29.3 Goldilocks Effect

THEORY CARD

Core claim. Many affective and attentional responses follow an **inverted-U**: people prefer stimuli that are moderately novel, complex, or discrepant from what they know, not too familiar, not too strange, but "just right".
Origins. Experimental aesthetics (Berlyne 1970, 1971), developmental psychology (R. B. McCall and McGhee 1977, Kidd, Piantadosi, and Aslin 2012), the "just right" label was popularised in consumer behaviour by Berger (2016).
Level of analysis. Individual perceiver.
Key constructs. Novelty, complexity, arousal potential, schema discrepancy, familiarity, optimal level of stimulation.
Boundary conditions. The location of the peak depends on the perceiver's expertise and prior exposure. Some measures (such as typicality-based preference) show monotonic rather than inverted-U patterns.
Related theories. Mere exposure, Beauty in averageness, Optimal distinctiveness theory (Section 22.3), Hedonic adaptation (Section 25.2).

29.3.1 Origins and core logic

The idea that hedonic value peaks at intermediate stimulation is old, it is often represented as the **Wundt curve**, but its modern form owes most to Daniel Berlyne's psychobiological aesthetics. Berlyne (1971) argued that stimulus properties such as novelty, complexity, surprisingness,

and ambiguity jointly determine a stimulus's **arousal potential**, and that hedonic value is highest at moderate arousal potential. Berlyne (1970) examined how novelty and complexity interact with exposure: repeated exposure tended to make simpler patterns less pleasing while more complex patterns retained or gained appeal, because complex patterns take longer to become fully familiar.

Developmental psychologists reached a similar conclusion about attention. The **discrepancy hypothesis** holds that infants attend most to stimuli that are moderately discrepant from their existing schemas. Stimuli that match a schema exactly are uninteresting, and stimuli that are wildly discrepant cannot be assimilated (R. B. McCall and McGhee 1977). Kidd, Piantadosi, and Aslin (2012) provided a modern test using an information-theoretic measure of how predictable each event in a visual sequence was. Infants were least likely to look away from sequences of intermediate predictability, and more likely to look away when events were either highly predictable or highly surprising: a "Goldilocks effect" in attention.

In consumer research, Meyers-Levy and Tybout (1989) found an analogous pattern for product evaluation: products that were **moderately incongruent** with their category schema were evaluated more favourably than products that were fully congruent or extremely incongruent, because moderate incongruity can be resolved and the resolution is itself rewarding. Berger (2016) summarises the general lesson for social influence: things that are too novel feel unfamiliar and uneasy. Things that are too familiar feel boring. Optimal appeal lies in between (Figure 29.2).

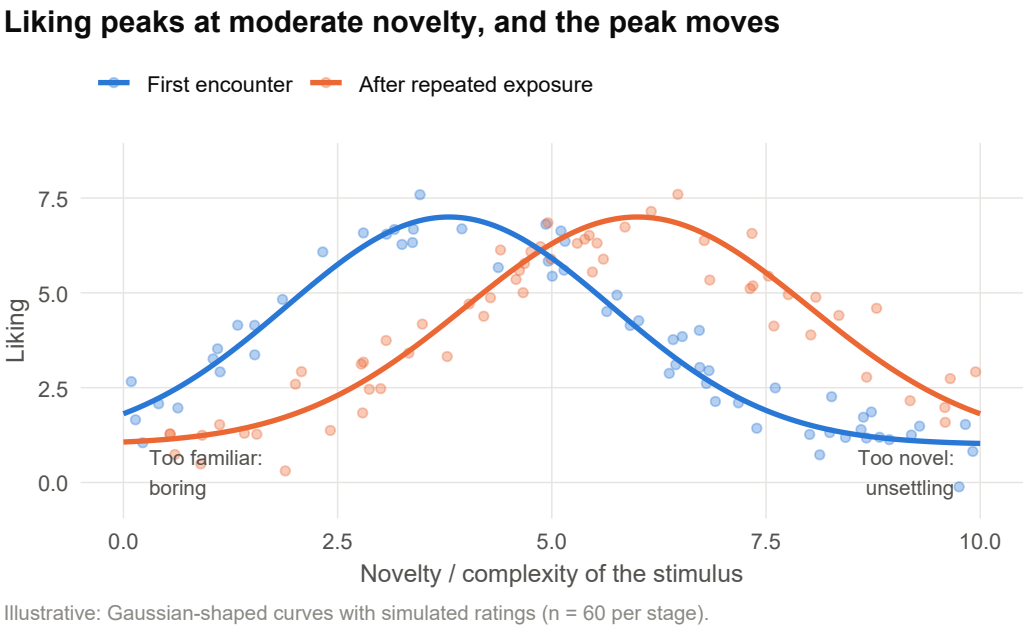


Figure 29.2. Illustrative inverted-U between novelty and liking. With repeated exposure, simple stimuli become boring and the preferred level of novelty shifts upward. Curves are stylised functions, points are simulated ratings.

DEBATE AND EVIDENCE

Inverted-U or monotonic typicality? Berlyne's framework predicts that moderately novel or complex stimuli are preferred. Yet studies by Martindale and colleagues found that **typicality**, how representative a stimulus is of its category, often predicted aesthetic preference

better than novelty or complexity, a result they described as anomalous for Berlyne's theory (Martindale, Moore, and West 1988, Martindale, Moore, and Borkum 1990). The finding links directly to the beauty-in-averageness effect. One reconciliation is that typicality and novelty are not opposite ends of one dimension: a design can be both recognisably typical and fresh in its details. Hekkert, Snelders, and Wieringen (2003) found that typicality and novelty were *both* positively related to preference for industrial designs once their negative correlation with each other was taken into account: the "most advanced, yet acceptable" principle. For researchers, the lesson is to measure typicality and novelty separately rather than assuming they are mirror images.

BUSINESS CASE

Product design and innovation newness. Firms launching new products face a Goldilocks trade-off. A design that is too close to existing products fails to attract attention or justify a premium. A design that is too radical can confuse consumers and depress adoption until the market catches up. Automotive design is a frequently discussed case: manufacturers typically evolve the appearance of established models gradually, preserving recognisable cues while refreshing details, and reserve radical departures for concept cars that test consumer reactions. The inverted-U also implies that the optimal level of novelty differs by segment. Experts and enthusiasts, who have seen many variants, tend to prefer higher novelty than novices, so a product line may include a mainstream variant for the broad market and a more adventurous variant for enthusiasts.

29.4 Coolidge Effect

THEORY CARD

Core claim. In many mammalian species, a male that has reached sexual satiety with one partner shows renewed sexual behaviour when a novel receptive partner is introduced.
Origins. Comparative psychology, J. R. Wilson, Kuehn, and Beach (1963), reviewed by Dewsbury (1981), the name refers to an anecdote about US President Calvin Coolidge reported in R. N. Johnson (1972).
Level of analysis. Individual organism.
Key constructs. Sexual satiety, stimulus novelty, recovery of responding.
Boundary conditions. Documented primarily in non-human animals. The extent and form of the effect vary across species and between sexes. Applications to human consumption are **analogies**, not direct applications.
Related theories. Goldilocks effect, Hedonic adaptation (Section 25.2), Mere exposure.

The name comes from a story that has been retold many times. As reported by R. N. Johnson (1972):

> "… The President and Mrs. Coolidge were being shown (separately) around an experimental government farm. When [Mrs. Coolidge] came to the chicken yard she noticed that a rooster was mating very frequently. She asked the attendant how often that happened and was told, 'Dozens of times each day.' Mrs. Coolidge said, 'Tell that to the President when he comes by.' Upon being told, the President asked, 'Same hen every time?' The reply was, 'Oh, no, Mr. President, a different hen every time.' President: 'Tell

that to Mrs. Coolidge.' "

The anecdote is almost certainly apocryphal, but the phenomenon is well documented. J. R. Wilson, Kuehn, and Beach (1963) showed that male rats that had stopped copulating with one female resumed sexual activity when a new female was introduced. Dewsbury (1981) reviewed evidence across many species, showing that novelty effects on copulatory behaviour are widespread but vary in form. Effects in females have been studied less and appear in some species but are generally weaker or different in form.

29.4.1 The business analogue: satiation and variety seeking

The underlying principle, that responding to a stimulus declines with repetition and recovers when the stimulus changes, has close parallels in consumer behaviour that do not depend on any claim about sexuality. **Sensory-specific satiety** is the decline in the pleasantness of a food as it is eaten, relative to foods that have not been eaten. B. J. Rolls et al. (1981) showed that the decline is specific to the eaten food, which helps explain why people eat more when offered variety. **Variety seeking**, switching among options for the sake of change, has been studied extensively in marketing (McAlister and Pessemier 1982). Ratner, Kahn, and Kahneman (1999) found that people sometimes choose less-preferred items for the sake of variety, and that this can reduce moment-to-moment enjoyment even when people remember the varied sequence favourably.

BUSINESS CASE

Novelty in product lines. Food and beverage companies, quick-service restaurants, and consumer-goods firms manage satiation through rotating flavours, limited-time offers, and seasonal menus. These tactics address a real problem: consumers who are satiated on one variant may switch brands rather than stop buying the category. Offering variety *within* the brand captures that switching. Media and entertainment platforms do the same through recommendation algorithms that balance familiar content with new content. The research also warns of costs: excessive variety increases complexity and inventory costs, and, as Ratner, Kahn, and Kahneman (1999) suggest, variety chosen for its own sake does not always improve experienced satisfaction.

29.5 Bandwagon, Snob, and Veblen Effects

THEORY CARD

Core claim. An individual's demand for a good can depend on how many others consume it and on its price as such. Under a **bandwagon effect**, demand rises because others consume the good. Under a **snob effect**, demand falls because others consume it. Under a **Veblen effect**, demand rises with the price itself, because a high price displays wealth.
Origins. Economics, Leibenstein (1950), building on Veblen (1899), extended in Bagwell and Bernheim (1996).
Level of analysis. Individual consumer and market demand.
Key constructs. Functional versus non-functional demand, external consumption effects, conspicuous price, exclusivity, status.
Boundary conditions. Requires that consumption be observable or that its price be known to relevant others. Strongest for status goods, fashion, and goods whose value depends on

network size.
Related theories. Costly signaling (Section 8.4), Identity signaling (Section 8.5), Countersignaling (Section 8.6), Optimal distinctiveness (Section 22.3).

29.5.1 Core logic

Standard demand theory assumes that a consumer's demand for a good depends on its price, the consumer's income, and the prices of other goods, but not on what other people buy. Leibenstein (1950) relaxed that assumption. He distinguished **functional** demand, which depends on the qualities inherent in the good, from **non-functional** demand arising from external effects on utility. Among the latter, he identified three (Table 29.4):

Table 29.4. Leibenstein's (1950) three external effects on demand.

Effect	Demand depends on	Direction	Consequence for market demand	Business example
Bandwagon	Number of others consuming	Positive: more adopters, more demand	Demand is more price-elastic than functional demand	Fashion trends, social platforms, best-seller lists
Snob	Number of others consuming	Negative: more adopters, less demand	Demand is less price-elastic than functional demand	Limited editions, niche luxury, exclusive clubs
Veblen	Price itself (conspicuous price)	Positive: higher price, more demand over some range	Demand curve can slope upward over a range	Prestige watches, luxury handbags, premium spirits

The snob effect connects to the identity-signaling literature reviewed in Chapter 8: people avoid goods that have become too common because the goods no longer signal distinction (Berger and Heath 2007). The Veblen effect connects to costly signaling: a good whose price is high and known can function as a credible display of wealth (Veblen 1899, Bagwell and Bernheim 1996). When many people adopt a status good, early adopters may abandon it, and marketers respond by cultivating an *illusion of distinction*: emphasising the ways in which offerings differ, even when they are similar at the core. Figure 29.3 sketches the implications for demand curves.

BUSINESS CASE

Luxury scarcity versus mass adoption. Luxury houses manage the snob and Veblen effects deliberately. They restrict distribution to owned boutiques and selected retailers, limit production of iconic items, avoid frequent discounting, and sometimes allocate the most desirable products to established clients rather than selling them to anyone who walks in. Each practice preserves exclusivity and keeps price conspicuous. The tension with growth is real: expanding into accessible "masstige" lines, outlet stores, or heavy discounting can attract new customers while diluting the brand's value to existing ones. The controversy in 2018, when Burberry disclosed that it had been destroying unsold stock to protect its brand and then announced it would end the practice, illustrates the costs of defending exclusivity in an era of sustainability scrutiny. By contrast, products with strong **bandwagon** or network effects (Section 17.6), social platforms, messaging apps, marketplaces, benefit from rapid mass adoption, and their managers subsidise early users to reach critical mass. Diagnosing which effect dominates in a category is therefore a strategic decision with opposite implications for pricing

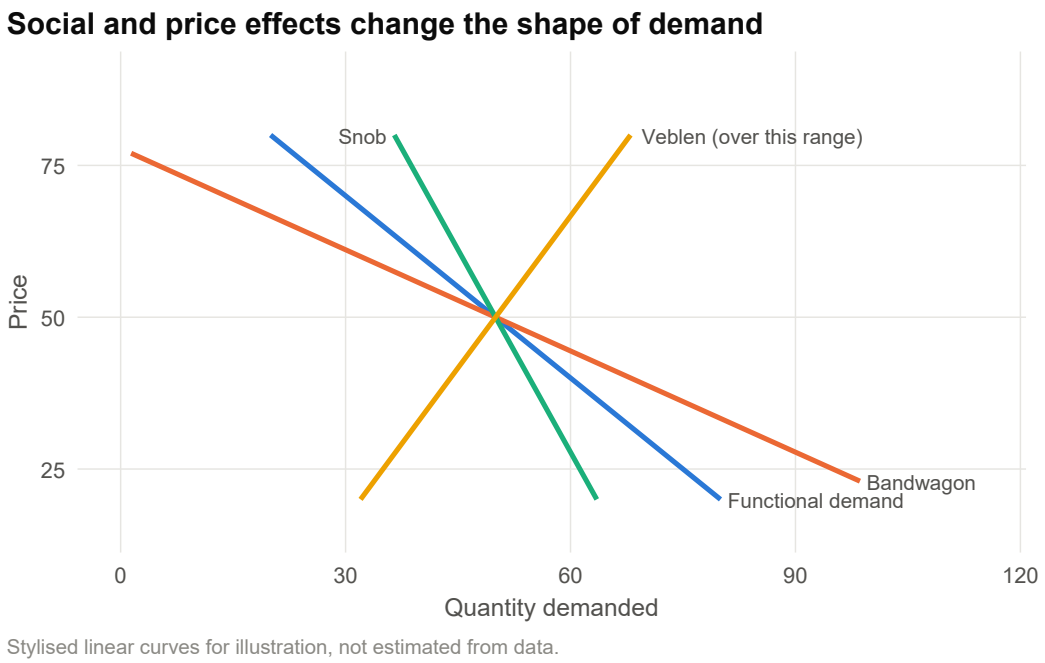


Figure 29.3. Stylised market demand curves. Relative to functional demand, a bandwagon effect makes demand more price-elastic, a snob effect makes it less elastic, and a Veblen effect can produce an upward-sloping segment over a range of prices. All curves pass through a common point for comparison.

and distribution.

COMMON PITFALL

Mistaking a quality inference for a Veblen effect. Demand that rises with price is not necessarily a Veblen effect. Consumers may infer *quality* from price when quality is hard to observe, which is an information effect (see Signaling theory (Section 8.2)), not a desire to display price to others. Distinguishing the two requires variation in observability: a Veblen effect should be stronger for publicly consumed goods, whereas price-quality inference should operate for private goods as well.

29.6 Glass Cliff Effect

THEORY CARD

Core claim. Women, and, in later research, members of racial and ethnic minority groups, are more likely than white men to be appointed to leadership positions that are **precarious**: roles in organisations that are performing poorly or are in crisis, where the risk of failure and criticism is high.

Origins. Organisational and social psychology, M. K. Ryan and Haslam (2005), reviewed in M. K. Ryan et al. (2016) and meta-analysed in Morgenroth et al. (2020).

Level of analysis. Individual appointment decisions within organisations, board and organi-

sational level.
Key constructs. Precariousness (prior performance, crisis), leader gender and ethnicity, leadership stereotypes ("think manager-think male", "think crisis-think female"), signalling change, scapegoating.
Boundary conditions. Depends on how precariousness is measured (share-price versus accounting performance), on the history of leadership in the organisation, and on the country and industry context.
Related theories. Upper echelons theory (Section 6.4), Social identity theory (Section 22.2), Signaling (Section 8.2), Agency theory (Section 6.1).

29.6.1 Origins and evidence

The glass cliff was identified as a response to a public claim. In 2003, a newspaper article argued that companies with more women on their boards performed worse. M. K. Ryan and Haslam (2005) examined the timing of board appointments in FTSE 100 companies and found the reverse causal story: companies that appointed women to their boards had typically experienced consistently poor share-price performance in the months *before* the appointment, whereas companies that appointed men had relatively stable performance. Women were not causing poor performance. They were more likely to be appointed when performance was already poor. The authors coined the term **glass cliff** to complement the familiar "glass ceiling": even when women break through to leadership, they may be placed in positions where failure is more likely.

Subsequent research, reviewed in M. K. Ryan et al. (2016), used both archival and experimental designs. In experiments, participants, including business and law students and managers, choose between equally qualified male and female candidates for a leadership role or high-risk assignment that is described as either going well or going badly. Female candidates are more often chosen when the situation is described as precarious. M. K. Ryan et al. (2011) found that the stereotype of the ideal manager shifts in crisis: people associate the characteristics needed to manage a failing company with stereotypically female traits more than they do for a successful company ("think crisis-think female"). Bruckmüller and Branscombe (2010) found that preference for a female leader in crisis depended on the organisation's leadership history: when a company had been led by men, a crisis increased the preference for a woman, in part because appointing a woman signalled change.

The pattern is not limited to gender. Studying chief executives of Fortune 500 firms, A. Cook and Glass (2014) found that women and racial and ethnic minorities were more likely than white men to be promoted to CEO of weakly performing firms, and that when firm performance declined during their tenure, they were more likely to be replaced by white men: a pattern the authors described as a "saviour effect" (Figure 29.4).

Table 29.5 presents the explanations for the glass cliff.

Table 29.5. Explanations for the glass cliff.

Explanation	Core argument	Representative evidence
Think crisis-think female	Stereotypically female traits are seen as fitting crisis management	Ryan et al. (2011)
Signalling change	Appointing a non-traditional leader signals that the organisation is changing course	Bruckmuller & Branscombe (2010)
Scapegoating	Leaders appointed in crisis can be blamed if recovery fails	Discussed in Ryan et al. (2016)

Table 29.5. Explanations for the glass cliff. *(continued)*

Explanation	Core argument	Representative evidence
Limited alternatives	Candidates with fewer opportunities may accept riskier roles	Discussed in Ryan et al. (2016)
In-group protection	Decision makers shield members of their own group from risky roles	Discussed in Ryan et al. (2016)
Saviour effect	Minority leaders are replaced by white men when performance declines	Cook & Glass (2014)

DEBATE AND EVIDENCE

How general is the glass cliff? Archival evidence is sensitive to measurement choices. Using US data and several performance measures, S. M. Adams, Gupta, and Leeth (2009) did not find that female executives were systematically appointed to firms with poorer prior performance, suggesting that the original pattern might depend on the performance measure, the market, or the period. Experimental evidence is more consistent, but experiments abstract from real selection processes. A meta-analysis by Morgenroth et al. (2020) found evidence for the glass cliff while documenting substantial variation across study designs, samples, and contexts. A balanced reading is that the glass cliff is a real tendency whose strength depends on context: in particular, on the organisation's leadership history and on how precariousness is defined and observed by decision makers.

BUSINESS CASE

Leadership appointments in crisis firms. Consider an *illustrative scenario*: a retailer's board must replace a chief executive after a year of falling sales and a data breach. The board's shortlist contains an internal female executive and an external male candidate. Glass cliff research does not say which candidate to choose. It identifies risks in how the decision is made and implemented. First, boards should ask whether the appointment is driven by a desire to *signal* change rather than by an assessment of fit, because signalling motives can coexist with under-investment in the new leader's success. Second, boards should equip leaders appointed to turnarounds with the resources a turnaround requires, authority to restructure, board support, time, and set performance expectations that reflect the starting position rather than the industry average. Third, evaluation should be benchmarked against the counterfactual trajectory of the firm, so that a leader who slows a decline is not judged a failure. These practices protect the firm as well as the leader: an under-supported turnaround is costly regardless of who leads it.

RESEARCH TOOLKIT

Sample hypotheses.

- *H1.* The probability that a newly appointed CEO is a woman increases with the firm's negative performance in the preceding year.
- *H2.* This relationship is stronger for firms with a history of exclusively male leadership (Bruckmüller and Branscombe 2010).
- *H3.* Conditional on performance during tenure, women and minority CEOs face a higher hazard of dismissal than white male CEOs (A. Cook and Glass 2014).

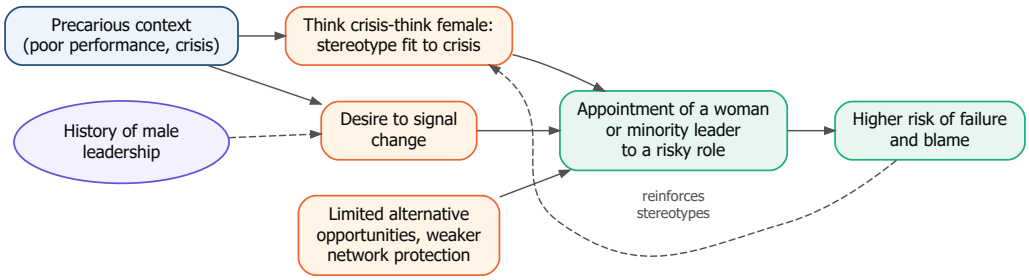


Figure 29.4. Mechanisms proposed for the glass cliff. Organisational precariousness activates beliefs about what a crisis requires and a desire to signal change. These shape appointments of women and minorities to risky roles, whose higher failure risk can feed back into stereotypes about their leadership.

Designs. Archival panels of executive and board appointments with lagged performance measures. Event-history models of tenure. Vignette experiments manipulating precariousness and leadership history.
Measures. Prior share-price and accounting performance, analyst downgrades, crisis events, leader demographics, tenure and exit.

29.7 IKEA Effect

THEORY CARD

Core claim. People value products that they have created or assembled themselves more than equivalent products created by others: labour leads to love.
Origins. Consumer psychology, Norton, Mochon, and Ariely (2012), related to the "I designed it myself" effect in mass customisation (Franke, Schreier, and Kaiser 2010) and to effort justification (Aronson and Mills 1959).
Level of analysis. Individual consumer.
Key constructs. Self-created product, labour or effort, task completion, feelings of competence and accomplishment, willingness to pay.
Boundary conditions. The effect depends on **successful completion**. It disappears when people fail to complete the task or when their creation is destroyed, and very frustrating tasks may reduce rather than increase valuation.
Related theories. Endowment effect (Section 23.4), Self-determination theory (Section 19.6), Decoy effect (another lever in product-line design).

29.7.1 Core logic and evidence

Norton, Mochon, and Ariely (2012) named the effect after the Swedish furniture retailer whose flat-pack products require customers to assemble them. In their experiments, participants who assembled a simple storage box were willing to pay more for it than participants who inspected an identical, pre-assembled box. Participants who folded origami valued their own amateur creations far above how other people valued them, and close to the value they placed on origami made by experts, and they expected others to share their valuation. Crucially, the effect depended on completion: when participants built and then destroyed their creations, or did not complete them, the

increase in valuation disappeared.

The effect has three candidate mechanisms that are not mutually exclusive. The first is **effort justification**: people value outcomes for which they have suffered, as in the classic finding that a severe initiation increased liking for a group (Aronson and Mills 1959). The second is **competence**: successfully completing a task signals to oneself that one is capable, and the product becomes a symbol of that competence. The third is **ownership and self-extension**: a self-made product is more closely linked to the self, which connects the IKEA effect to the endowment effect (Section 23.4). Franke, Schreier, and Kaiser (2010) found a related pattern in mass customisation: consumers were willing to pay more for products they had designed themselves, beyond what the better fit to their preferences could explain, and this "I designed it myself" effect was associated with a feeling of accomplishment (Figure 29.5).

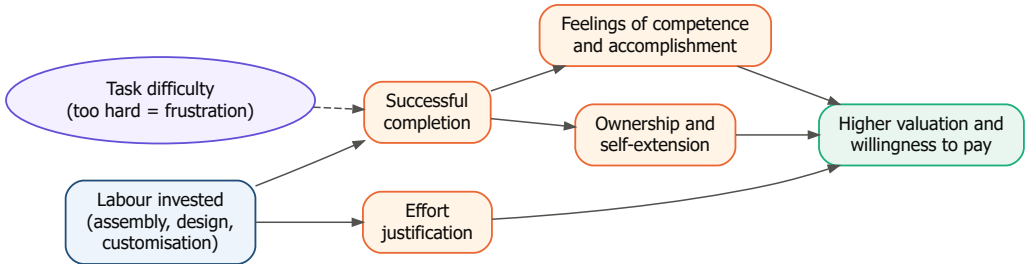


Figure 29.5. Mechanism of the IKEA effect. Labour that culminates in successful completion produces feelings of competence and ownership, which raise the creator's valuation. Failure to complete eliminates the effect (Norton, Mochon, & Ariely, 2012).

BUSINESS CASE

Co-creation and self-assembly products. The IKEA effect is visible in many business models that ask customers to do part of the work: flat-pack furniture, meal kits that require cooking, build-your-own toys, and online configurators for shoes, computers, and cars. The research suggests three design principles. First, **calibrate effort**: the task must be demanding enough to feel like an accomplishment but easy enough to complete. Instructions, pre-drilled parts, and modular design reduce the risk of failure that would eliminate the effect. Second, **make completion visible**: a finished product, a shareable design, or a personalised element ensures that the customer's contribution is salient. Third, **beware the creator's bias**: just as participants expected others to value their origami, customers and employees overvalue their own creations. In product development, this means that teams can become attached to features they built. In co-creation programmes, it means that participants' enthusiasm is not a reliable guide to the market's.

29.8 Decoy Effect

THEORY CARD

Core claim. Adding an option that is clearly inferior to one existing option (the **target**) but not to another (the **competitor**) increases the share of choices going to the target. The effect is also called the **attraction effect** or **asymmetric dominance effect**.

Origins. Consumer research, J. Huber, Payne, and Puto (1982), extended by Simonson (1989) and Simonson and Tversky (1992), robustness debated in a 2014 *Journal of Marketing Research* exchange (Frederick, Lee, and Baskin 2014, J. Huber, Payne, and Puto 2014, Yang and Lynn 2014).
Level of analysis. Individual choice, market shares.
Key constructs. Target, competitor, decoy, asymmetric dominance, regularity, compromise effect, justification.
Boundary conditions. Strongest when options are described by a small number of numeric attributes, when the dominance relation is easy to see, and when people expect to justify their choice. Weak or absent with perceptual or holistic stimuli.
Related theories. Screening theory (Section 8.3), Heuristics and biases (Section 23.2), Anchoring, Prospect theory (Chapter 24).

29.8.1 Core logic

Most models of rational choice imply **regularity**: adding a new option to a choice set cannot increase the probability of choosing any option already in the set. The new option can only take share away. J. Huber, Payne, and Puto (1982) showed that this principle can be violated. When they added an option that was dominated by one alternative (worse on both attributes) but not by the other, the share of the dominating alternative increased. Figure 29.6 illustrates the geometry: the decoy sits in the region that the target dominates but the competitor does not.

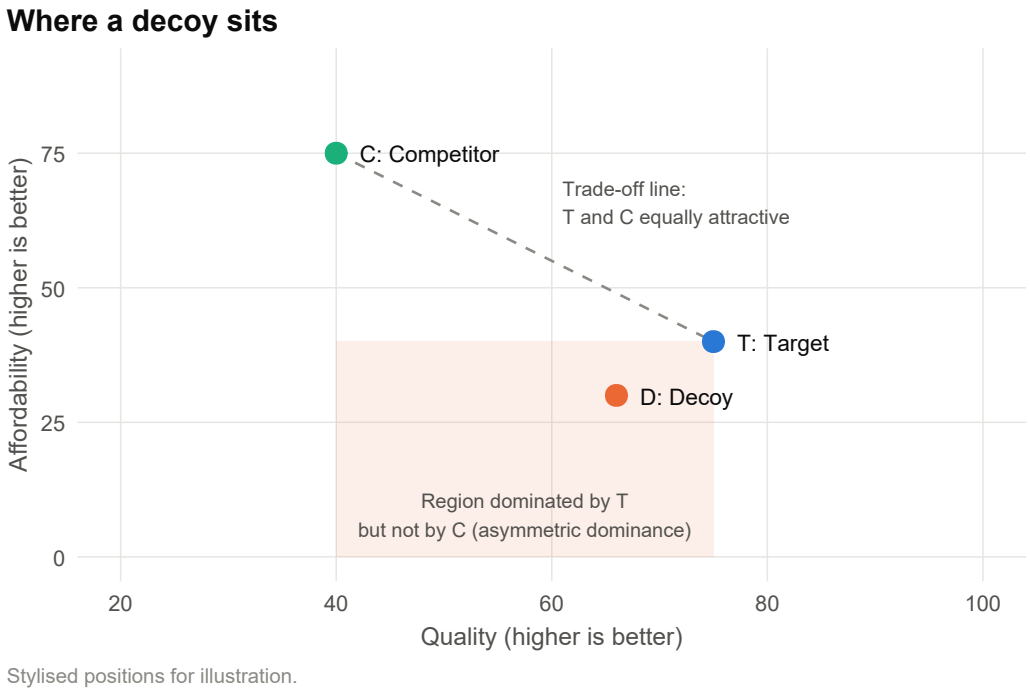


Figure 29.6. Stylised attribute space for the decoy effect. The decoy (D) is worse than the target (T) on both quality and affordability but is not dominated by the competitor (C). Its presence makes T look better by comparison. Positions are illustrative.

Several context effects are related. The **compromise effect** describes the tendency for an option

to gain share when it becomes the intermediate option in a set, because extreme options are less attractive than moderate ones: an aspect of **extremeness aversion** (Simonson and Tversky 1992). Simonson (1989) proposed that both attraction and compromise effects reflect **choice based on reasons**: a dominated decoy gives the chooser an easy justification for the target ("it is clearly better than D"), and the effects are stronger when people expect to be evaluated on their choice. Table 29.6 summarises the main variants.

Table 29.6. Context effects in multi-attribute choice.

Context effect	Position of the added option	Predicted shift	Proposed mechanism	Key source
Attraction (asymmetric dominance)	Dominated by the target only	Toward the target	Dominance provides a reason, relative advantage	Huber, Payne & Puto (1982)
Compromise	Makes the target the middle option	Toward the middle option	Extremeness aversion, ease of justification	Simonson (1989), Simonson & Tversky (1992)
Similarity (substitution)	Very similar to one existing option	Away from the similar option	New option draws share from close substitutes	Discussed in Huber, Payne & Puto (1982)

DEBATE AND EVIDENCE

How robust is the attraction effect? Frederick, Lee, and Baskin (2014) reported a series of studies in which the attraction effect was frequently absent when options were described perceptually or qualitatively, for example, with pictures rather than numeric tables, and in some studies the effect reversed. They argued that the effect is much narrower than the literature implied. Yang and Lynn (2014) likewise reported difficulty obtaining the effect with realistic stimuli. In reply, J. Huber, Payne, and Puto (2014) accepted that the effect has limits but argued that it is robust where its conditions are met: when decision makers recognise the dominance relation, when attributes are commensurable, and when the decoy is plausible rather than absurd. The exchange is a model of productive debate. It moved the field from asking *whether* the effect exists to asking *when* it occurs, and it cautions practitioners against assuming that a decoy that worked in a numeric laboratory task will work in a visually rich storefront.

BUSINESS CASE

Pricing tiers and decoys. Subscription and software pricing pages commonly present three or four plans. Some plans may exist primarily to shape choice among the others. Ariely (2008) popularised the example of a magazine subscription offer that included a print-only option priced the same as a combined print-and-web option. Few would choose print-only, but its presence made the combined offer look like a bargain, and removing it shifted choices toward the cheaper web-only option in his classroom demonstrations. Restaurants use a similar logic when an expensive wine on the menu makes the second-most-expensive bottle look reasonable, and cinemas when a medium popcorn is priced close to the large. The decoy literature suggests guidelines for practice: decoys work best when the comparison with the target is easy to see (shared attributes, numeric presentation, adjacent placement), when customers are comparing rather than choosing by gut feeling, and when the decoy is credible. It also suggests caution. Customers who recognise a decoy as manipulative may lose trust (see the

persuasion knowledge model (Section 26.2)), and a decoy that customers actually choose is simply a poorly priced product.

TRY IT YOURSELF

Find a pricing page for a subscription service with at least three tiers. For each pair of tiers, ask whether one dominates the other on price and features. Is any tier asymmetrically dominated? Is any tier positioned as a compromise? Predict how choice shares would change if that tier were removed.

29.9 Anchoring Effect

THEORY CARD

Core claim. Numeric estimates are biased toward a value that was considered beforehand, the **anchor**, even when the anchor is obviously irrelevant.
Origins. Judgment and decision making, Tversky and Kahneman (1974), extended to valuation by Ariely, Loewenstein, and Prelec (2003), reviewed by Furnham and Boo (2011).
Level of analysis. Individual judgment, dyadic negotiation.
Key constructs. Anchor (externally provided or self-generated), adjustment, selective accessibility, plausibility of the anchor.
Boundary conditions. Robust across domains and expertise levels. Effects are larger when the target quantity is uncertain and the anchor is plausible, and can be reduced by generating reasons why the anchor is inappropriate.
Related theories. Heuristics and biases (Section 23.2), Framing (Section 26.6), Decoy effect, Prospect theory (Chapter 24).

29.9.1 Core logic and evidence

In the best-known demonstration, Tversky and Kahneman (1974) spun a wheel of fortune that was rigged to stop at either 10 or 65 in front of participants, who then estimated the percentage of African countries in the United Nations. The arbitrary number influenced the estimates: median estimates were 25 among participants who saw 10 and 45 among those who saw 65 (Figure 29.7). Tversky and Kahneman attributed the effect to **insufficient adjustment**: people start from the anchor and adjust toward a plausible value, but stop too soon.

Anchoring extends to valuation. Ariely, Loewenstein, and Prelec (2003) asked participants to write down the last two digits of their social security number and to state whether they would pay that amount, in dollars, for various products such as wine and computer accessories. Participants then stated their maximum willingness to pay. Valuations were strongly correlated with the arbitrary digits. Yet the *relative* valuations of different products were sensible: participants valued items that should be worth more at higher prices. The authors described this pattern as **coherent arbitrariness**: preferences are internally consistent but anchored to arbitrary starting points.

Expertise does not eliminate anchoring. Northcraft and Neale (1987) gave real estate agents and amateurs identical information about a house, varying only the listing price. Both groups' appraisals were influenced by the listing price, and the agents tended to deny that it had influenced

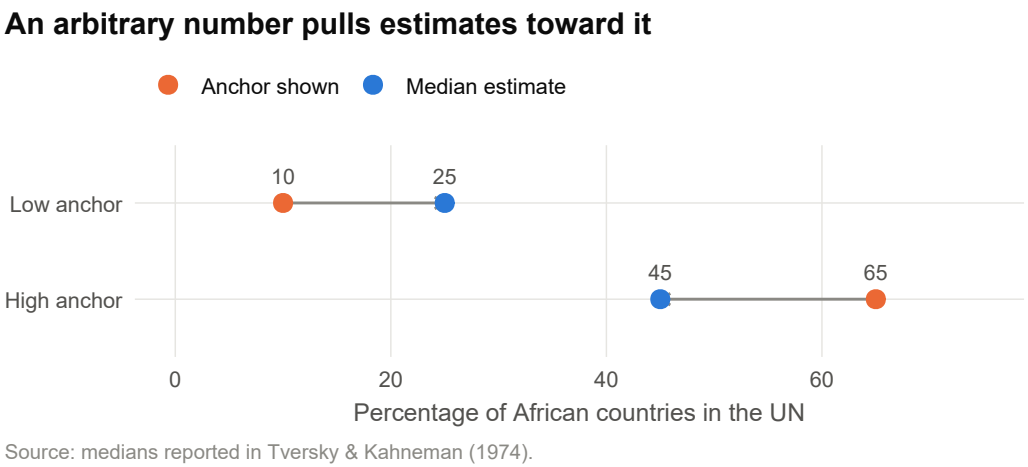


Figure 29.7. Anchoring with an arbitrary number. Median estimates of the percentage of African countries in the United Nations after participants saw a wheel of fortune stop at 10 or 65, as reported by Tversky & Kahneman (1974).

them. In negotiations, **first offers** act as anchors: Galinsky and Mussweiler (2001) found that the party making the first offer tended to obtain better outcomes, and that the effect was reduced when the recipient focused on information inconsistent with the anchor, such as the offerer's alternatives.

29.9.2 Mechanism

Two mechanisms have been proposed, and they appear to apply to different kinds of anchors. The **adjustment** account fits **self-generated** anchors: values people produce themselves as starting points, such as the known boiling point of water when estimating it at altitude. Epley and Gilovich (2006) found that people adjust from such anchors until they reach a plausible value and then stop, producing estimates biased toward the anchor. The **selective accessibility** account fits **externally provided** anchors. Strack and Mussweiler (1997) argued that considering whether the target quantity is higher or lower than the anchor makes anchor-consistent information more accessible, and that this information then shapes the final estimate. Anchoring is also among the most replicable effects in psychology: it replicated robustly in the multi-laboratory Many Labs project (R. A. Klein et al. 2014) (Table 29.7).

Table 29.7. Anchoring paradigms and their business counterparts.

Paradigm	Anchor type	Typical finding	Business setting	Source
Wheel of fortune	Arbitrary, externally provided	Estimates move toward the random number	Survey and forecasting design	Tversky & Kahneman (1974)
Social security digits	Arbitrary, self-relevant	Willingness to pay tracks the digits (coherent arbitrariness)	Reference prices, price recommendations	Ariely, Loewenstein & Prelec (2003)
Listing price	Informative but biased	Expert appraisals follow the listing price	Real estate, procurement	Northcraft & Neale (1987)
First offer	Strategic, from counterpart	First mover obtains better outcomes	Negotiation, salary setting	Galinsky & Mussweiler (2001)

Table 29.7. Anchoring paradigms and their business counterparts. *(continued)*

Paradigm	Anchor type	Typical finding	Business setting	Source
Self-generated	Known value adjusted from	Adjustment stops too early	Budgeting from last year's figures	Epley & Gilovich (2006)

BUSINESS CASE

Reference prices and negotiation. Retailers display "was" prices, manufacturers publish suggested retail prices, and software firms list an annual price beside a discounted monthly equivalent. Each provides an anchor against which the actual price is judged. Regulators in many jurisdictions restrict fictitious reference prices precisely because anchors influence perceived value. In business-to-business negotiation, the anchoring literature recommends that well-prepared parties consider making the first offer, and that recipients of an aggressive first offer counter its influence by focusing on their own alternatives and target rather than on the counterpart's number (Galinsky and Mussweiler 2001). Inside organisations, last year's budget is a self-generated anchor. Zero-based budgeting is, in part, an attempt to remove it.

COMMON PITFALL

Treating every reference point as an anchor. A displayed price can affect judgments because it is an anchor, because it is *informative* (a high list price may truthfully signal quality), or because it changes the reference point against which gains and losses are evaluated (Prospect theory (Chapter 24)). Demonstrating anchoring requires showing that an uninformative or randomly assigned number shifts judgments.

29.10 Beauty in Averageness

THEORY CARD

Core claim. Stimuli that are close to the average or prototype of their category , faces, designs, musical performances, tend to be judged attractive.
Origins. Perception and evolutionary psychology, Langlois and Roggman (1990), extended to non-face categories by Halberstadt and Rhodes (2003) and to products by Veryzer, Jr. and Hutchinson (1998) and Landwehr, Labroo, and Herrmann (2011).
Level of analysis. Individual perceiver.
Key constructs. Averageness, prototypicality (typicality), symmetry, processing fluency, aesthetic preference.
Boundary conditions. Averageness is attractive but not maximally attractive: some non-average features can be preferred, and in some categories distinctiveness adds value. The effect is weaker when typicality is confounded with boredom.
Related theories. Goldilocks effect, Processing fluency (Section 23.8), Mere exposure, Wisdom of the crowd (Section 30.5).

29.10.1 Core logic and evidence

Langlois and Roggman (1990) created composite faces by digitally averaging photographs of individuals. Composite faces were judged more attractive than most of the individual faces from which they were created, and attractiveness tended to increase as more faces were included in the composite (see also Strzalko and Kaszycka (1992) on how attractiveness assessments vary within and between raters). The pattern is not specific to faces. Halberstadt and Rhodes (2003) found that computer-averaged images of birds, fish, and automobiles were also judged attractive. In music, Repp (1997) found that a performance constructed by averaging the expressive timing and dynamics of many pianists' performances was judged aesthetically pleasing: in some respects more so than most individual performances. In experimental aesthetics, Martindale and colleagues found that typicality was a strong predictor of preference for many kinds of stimuli (Martindale, Moore, and West 1988, Martindale, Moore, and Borkum 1990).

The effect has clear business relevance. Veryzer, Jr. and Hutchinson (1998) found that prototypicality and unity increased aesthetic responses to new product designs. Landwehr, Labroo, and Herrmann (2011) showed that a measure of **design fluency**, capturing how typical and how complex a car's design was, improved forecasts of automobile sales beyond conventional predictors, suggesting that the ease with which consumers process a design is commercially consequential.

29.10.2 Explanations

Three families of explanation have been proposed, summarised in Table 29.8.

Table 29.8. Three explanations for beauty in averageness.

Explanation	Core argument	Key sources	Prediction that distinguishes it
Evolutionary (good genes)	Average and symmetric faces signal developmental stability and genetic diversity	Thornhill & Gangestad (1993), Grammer & Thornhill (1994), Langlois & Roggman (1990), Rhodes (2006)	Effect should be strongest for faces and mates
Processing fluency	Prototypes are easy to process, and ease is experienced as pleasure	Winkielman et al. (2006), Reber (2011), Reber, Schwarz & Winkielman (2004), Landwehr et al. (2011)	Effect should extend to any category with a prototype, including novel ones
Wisdom of crowds (error cancellation)	Averaging cancels idiosyncratic deviations, moving the result closer to an optimum	Surowiecki (2005), Halberstadt & Rhodes (2003), Repp (1997)	Averages of skilled performances should beat most individuals

The **evolutionary** account holds that averageness and symmetry in faces are cues to health and genetic quality, so preferences for them would have been favoured by selection (Thornhill and Gangestad 1993, Grammer and Thornhill 1994, Rhodes 2006). The **fluency** account holds that prototypes are processed easily, and that the resulting fluency is experienced as positive affect. Winkielman et al. (2006) found that prototypical patterns were classified faster and liked more, consistent with fluency, and Reber (2011) develops the argument that fluency underlies culturally shared taste. Because the fluency account predicts the effect for any category with a prototype, including categories with no evolutionary significance, such as cars or dot patterns, findings such as those of Halberstadt and Rhodes (2003) are often read as favouring it. The **wisdom-of-crowds** account draws on the statistical point that averaging independent estimates cancels their errors (Surowiecki 2005): an

averaged face or performance smooths out idiosyncratic deviations, moving the result toward an optimum (Halberstadt and Rhodes 2003, Repp 1997).

BUSINESS CASE

Typicality in product design. The beauty-in-averageness effect explains why many successful designs look "right" at first sight: they are close to the category prototype. It also explains a frequent tension between designers, who value originality, and market research, which often favours familiar forms. The Goldilocks effect and the "most advanced, yet acceptable" principle suggest a resolution: keep the overall form typical, the silhouette of a car, the layout of a website, the shape of a bottle, and innovate in details that distinguish the brand. Landwehr, Labroo, and Herrmann (2011)'s results suggest that design properties such as typicality and complexity can be measured during development and used to forecast sales, allowing firms to treat design not purely as a matter of taste but as a measurable predictor of market performance.

29.11 Pratfall Effect

THEORY CARD

Core claim. A minor blunder can make a highly competent person more likeable, whereas the same blunder makes a person of average competence less likeable.
Origins. Social psychology, Aronson, Willerman, and Floyd (1966), qualified by R. Helmreich, Aronson, and LeFan (1970).
Level of analysis. Individual perceiver evaluating a target person (and, by extension, a brand).
Key constructs. Competence, blunder (pratfall), likeability, perceiver self-esteem, blemishing.
Boundary conditions. Requires that the target's competence already be established and that the blunder be minor and unrelated to core competence. Effects depend on characteristics of the perceiver.
Related theories. Persuasion knowledge model (Section 26.2), Social comparison theory (Section 22.1), Signaling theory (Section 8.2).

Aronson, Willerman, and Floyd (1966) had participants listen to a tape recording of a student who was supposedly trying out for a college quiz team. In one version, the candidate answered almost all questions correctly and described an impressive record. In another, he performed at an average level. In half of the recordings, the candidate clumsily spilled coffee over himself near the end. The blunder *increased* the attractiveness of the superior candidate and *decreased* the attractiveness of the average candidate. The interpretation was that a minor flaw makes a highly competent person seem more human and approachable, whereas for a person of average ability it simply adds to an impression of mediocrity.

The effect is less general than its popularity suggests. R. Helmreich, Aronson, and LeFan (1970) found that the effect of a pratfall depended on the self-esteem of the observer, so that the blundering superior person was not uniformly preferred. Related research in consumer behaviour points to a broader principle. Ein-Gar, Shiv, and Tormala (2012) documented a **blemishing effect**: adding a small piece of negative information to an otherwise positive description of a product can *increase*

evaluations, especially when people process information with little effort, because the weak negative information makes the positive information stand out by contrast.

BUSINESS CASE

Brand candour and leader communication. Brands sometimes admit a minor weakness to make their strengths more credible. Avis's long-running campaign built around being "No. 2" in car rentals, created by the agency Doyle Dane Bernbach in the 1960s, is a widely cited example of turning an admitted limitation into a reason to believe the brand's promise to "try harder". In product reviews, research on blemishing suggests that a mostly positive review profile with a few mild criticisms can be more persuasive than uniformly glowing reviews. For leaders, the pratfall effect suggests that acknowledging a minor error can increase warmth, but only when competence is already established and the error does not touch the core of the role. A chief financial officer who misspeaks a statistic in a speech may seem human. One who misstates earnings does not.

29.12 Using Named Effects Responsibly

Named effects are attractive because they are memorable, specific, and easy to apply. The same features make them easy to misuse. Three cautions follow from the evidence reviewed in this chapter.

First, **effects vary in robustness**. Anchoring replicated robustly across many laboratories (R. A. Klein et al. 2014), and mere exposure rests on a large meta-analytic base (Bornstein 1989, Montoya et al. 2017). By contrast, the attraction effect is sensitive to how options are presented (Frederick, Lee, and Baskin 2014), the glass cliff depends on context (Morgenroth et al. 2020), and the pratfall effect depends on the observer (R. Helmreich, Aronson, and LeFan 1970). The catalogue at the start of this chapter rates evidence strength for this reason. See the discussion of the replication crisis (Section 3.7) for general principles.

Second, **effects need mechanisms**. An effect tells us *that* something happens. A mechanism tells us *when* it will happen. The mere exposure effect is useful to a media planner only together with the two-factor account that predicts wear-out. The decoy effect is useful to a pricing manager only with the boundary conditions that determine whether customers will notice dominance. Linking each effect to a theory , fluency, prospect theory, signaling, social identity, is what turns a list of curiosities into a toolkit.

Third, **effects interact and sometimes conflict**. Familiarity breeds liking (mere exposure), but too much familiarity breeds boredom (Goldilocks, Coolidge). Typical designs are attractive (beauty in averageness), but distinctiveness signals status (snob effect). The resolution is rarely to choose one effect over another. It is to specify the conditions, the stage of exposure, the consumer segment, the observability of consumption, under which each dominates.

29.13 Chapter Summary

KEY TAKEAWAYS

- A **named effect** is an empirical regularity. Its value depends on the mechanism that explains it and the boundary conditions that limit it.
- **Mere exposure** increases liking, but repetition eventually produces tedium. The two-factor curve underlies wear-in and wear-out in advertising (Zajonc 1968, Bornstein 1989, Cacioppo and Petty 1979).
- The **warm-glow heuristic** runs the other way, liking increases perceived familiarity (Monin 2003), and must not be confused with Andreoni's warm-glow giving (Andreoni 1990).
- **Goldilocks** (inverted-U) patterns describe preference for moderate novelty and complexity (Berlyne 1970, R. B. McCall and McGhee 1977, Kidd, Piantadosi, and Aslin 2012), while **beauty in averageness** shows that typical stimuli are attractive (Langlois and Roggman 1990). The two are reconciled by treating typicality and novelty as separate dimensions (Hekkert, Snelders, and Wieringen 2003).
- **Bandwagon, snob, and Veblen effects** make demand depend on others' consumption and on price itself (Leibenstein 1950), with opposite implications for scarcity and mass-market strategies.
- The **glass cliff** describes the disproportionate appointment of women and minorities to precarious leadership roles (M. K. Ryan and Haslam 2005, M. K. Ryan et al. 2016, A. Cook and Glass 2014).
- The **IKEA effect**, **decoy effect**, **anchoring**, and **pratfall effect** offer practical levers for product design, pricing, negotiation, and communication: each with documented limits (Norton, Mochon, and Ariely 2012, J. Huber, Payne, and Puto 1982, Frederick, Lee, and Baskin 2014, Tversky and Kahneman 1974, Aronson, Willerman, and Floyd 1966).

Table 29.9 presents the comparison of the chapter's effects.

Table 29.9. Comparison of the chapter's effects.

Effect	Primary mechanism	Key moderator	Main business lever	Main risk of misuse
Mere exposure	Fluency, habituation vs. tedium	Number of exposures, complexity	Frequency planning	Ignoring wear-out
Warm-glow heuristic	Misattribution of affect to familiarity	Diagnosticity of memory	Awareness measurement	Confusing with warm-glow giving
Goldilocks	Optimal arousal, schema discrepancy	Expertise, prior exposure	Calibrating novelty	Assuming one peak for all segments
Coolidge	Satiation and recovery with novelty	Species, domain	Variety in product lines	Over-extending an animal finding
Bandwagon / snob / Veblen	Social and price externalities in demand	Observability of consumption	Scarcity vs. mass adoption	Confusing Veblen with price-quality inference
Glass cliff	Crisis stereotypes, signalling change	Leadership history, performance measure	Succession governance	Over-generalising across contexts
IKEA effect	Competence, effort justification, ownership	Task completion	Co-creation and self-assembly	Creator's overvaluation

Table 29.9. Comparison of the chapter's effects. *(continued)*

Effect	Primary mechanism	Key moderator	Main business lever	Main risk of misuse
Decoy	Reasons for choice, relative advantage	Numeric vs. perceptual presentation	Pricing tiers	Assuming lab robustness in the field
Anchoring	Adjustment, selective accessibility	Uncertainty, anchor plausibility	Reference prices, first offers	Confusing anchors with information
Beauty in averageness	Fluency, evolutionary cues, error cancellation	Category, distinctiveness value	Typical design forms	Ignoring the value of distinctiveness
Pratfall	Humanising a competent target	Established competence, observer	Brand candour	Generalising from a single study

29.14 Discussion Questions

1. The mere exposure effect and the Goldilocks effect make apparently opposite predictions about the effect of familiarity. Specify the conditions under which each should dominate, and design an advertising experiment to test them.
2. A brand-tracking survey shows that a brand's familiarity and favourability have both risen. Using the warm-glow heuristic and mere exposure, explain why the causal direction is ambiguous. How would you design a study to separate the two?
3. A luxury fashion brand is considering launching a lower-priced line. Use Leibenstein's framework to predict the effects on existing and new customers. What evidence would tell you whether the snob or the Veblen effect is more important for this brand?
4. Archival and experimental studies of the glass cliff do not always agree. What are the strengths and weaknesses of each design, and what data would provide the most convincing test?
5. Following the 2014 debate about the attraction effect, what conditions would you check before recommending a decoy pricing tier to a client? How would you test the recommendation in the field?
6. Choose two effects from this chapter that could be explained by processing fluency. Explain how fluency accounts for each, and identify one finding that a fluency account cannot easily explain.

29.15 Further Reading

- Zajonc (1968): the original monograph on mere exposure, combining correlational and experimental evidence.
- Bornstein (1989): the classic meta-analysis that identifies the moderators of the exposure effect.
- Leibenstein (1950): a short, clear treatment of bandwagon, snob, and Veblen effects that remains the starting point for social effects in demand.
- M. K. Ryan et al. (2016): a review of a decade of glass cliff research, its explanations, and its impact.
- Norton, Mochon, and Ariely (2012): the experiments that named the IKEA effect and identified completion as a boundary condition.

- Frederick, Lee, and Baskin (2014): a careful challenge to the robustness of the attraction effect, best read together with J. Huber, Payne, and Puto (2014).
- Ariely, Loewenstein, and Prelec (2003): an elegant demonstration that preferences can be coherent yet arbitrary.
- Langlois and Roggman (1990): the composite-face studies that launched research on beauty in averageness.

Chapter 30

Social Phenomena

Why does a challenger brand gain more from saying "we are just like the market leader" than the leader would gain from saying the reverse? Why can the average guess of a few hundred amateurs beat the judgement of a seasoned expert, and why does that advantage evaporate once the amateurs start talking to one another? Why do reporters from rival outlets end up writing the same story, why do a handful of novels, songs, and scientists capture most of the attention in their fields, and why does a shopper linger longer in a store that smells faintly of vanilla?

None of these questions is answered by a single grand theory. Each concerns a **social phenomenon**: a robust, recurring empirical regularity in how people judge, choose, cooperate, and influence one another. Philosophers of science distinguish *phenomena*, which are stable features of the world that theories are built to explain, from the *data* through which we detect them (Bogen and Woodward 1988). A phenomenon such as the bystander effect or the Matthew effect is therefore not a theory in itself. It is a pattern that several theories compete to explain and that managers ignore at their peril. Knowing the phenomenon tells us *what* tends to happen. Knowing the mechanisms behind it tells us *when* it will happen and *how* to design around it.

This chapter surveys twelve such phenomena, grouped by theme: **similarity and attraction** (similarity asymmetry, homophily, assortative mating, the social magnet). **Collective judgement** (the wisdom of crowds, groupthink, pack journalism). **Cumulative advantage** (the Matthew effect). **Collective effort and action** (social loafing, the bystander effect, the tragedy of the commons), and the link between **scent and emotion**. For each, we state the regularity, explain its mechanism, review the evidence and its limits, and show how it bears on business decisions (Table 30.1).

Table 30.1. Chapter at a glance: twelve social phenomena.

Phenomenon	Core regularity	Level of analysis	Key source
Similarity asymmetry	The less prominent object is judged more similar to the prominent one than vice versa	Individual cognition	Tversky (1977)
Homophily	Contact between similar people occurs at a higher rate than among dissimilar people	Dyad / network	McPherson et al. (2001)
Assortative mating	Partners resemble each other more than chance would predict	Dyad / population	Schwartz & Mare (2005)
Social magnet	Others' behaviour both attracts (imitation) and repels (differentiation)	Individual / group	Berger (2016)
Wisdom of the crowd	Aggregated independent judgements outperform most individuals	Group / market	Galton (1907), Surowiecki (2005)
Groupthink	Cohesive, insulated groups suppress dissent and decide badly	Small group	Janis (1972, 1982)

Table 30.1. Chapter at a glance: twelve social phenomena. *(continued)*

Phenomenon	Core regularity	Level of analysis	Key source
Pack journalism	Competing news outlets converge on the same stories and sources	Organization / field	Crouse (1973), Boczkowski (2010)
Matthew effect	Early advantage begets further advantage	Individual / field	Merton (1968)
Social loafing	People exert less effort when their contribution is pooled	Group	Latane et al. (1979)
Bystander effect	Helping becomes less likely as the number of witnesses rises	Group / situation	Darley & Latane (1968)
Tragedy of the commons	Individually rational use depletes shared resources	Community / institution	Hardin (1968), Ostrom (1990)
Scent and emotion	Odours evoke unusually emotional memories and shape evaluations	Individual	Rolls (2004), Willander & Larsson (2007)

DEFINITION

Definition 30.1. Social phenomenon. A stable, replicable regularity in social behaviour or judgement that holds across settings and is detectable with different kinds of data. Phenomena are the *explananda* of social theory: theories earn their keep by explaining why a phenomenon occurs and by predicting the conditions under which it strengthens, weakens, or reverses (Bogen and Woodward 1988).

30.1 Similarity Asymmetry

THEORY CARD

Core claim. Similarity judgements are directional. When people judge how similar a less prominent object (a variant) is to a more prominent one (a prototype), they report more similarity than when the comparison runs the other way.
Origins. Cognitive psychology. Amos Tversky's feature-based **contrast model** of similarity (Tversky 1977), later applied widely to brands and advertising.
Level of analysis. Individual cognition. Applied to pairs of products, brands, countries, or people.
Key constructs. Common features, distinctive features, feature salience, subject and referent of a comparison, prototypicality.
Boundary conditions. The asymmetry requires that the two objects differ in salience or prominence and that the judgement has a clear direction ("Is A like B?"). It shrinks when objects are equally familiar or when the task is non-directional ("How similar are A and B?").
Related theories. Processing fluency (Section 23.8), Beauty in averageness (Section 29.10), Anchoring effect (Section 29.9), Social comparison (Section 22.1).

30.1.1 Core logic

Before 1977, most psychological models represented objects as points in a geometric space, where similarity is a decreasing function of distance. Distance is symmetric: the distance from Lisbon to Madrid equals the distance from Madrid to Lisbon. Tversky (1977) showed that human similarity judgements violate this symmetry. People agree that "North Korea is similar to China" more read-

ily than “China is similar to North Korea”, and that a son resembles his father more than a father resembles his son. In each case the less prominent object is the *subject* of the comparison and the more prominent object is the *referent*.

Tversky replaced geometry with features. Each object is described by a set of features, and similarity is a weighted contrast between what the objects share and what distinguishes them:

$$S(a,b) = \theta f(A \cap B) - \alpha f(A - B) - \beta f(B - A), \qquad \theta, \alpha, \beta \geq 0,$$

where A and B are the feature sets of objects a and b, $f(\cdot)$ measures the salience of a set of features, $A \cap B$ are the common features, $A - B$ are features of a not shared by b, and $B - A$ the reverse. In a directional judgement (“a is like b”), attention focuses on the subject a, so its distinctive features weigh more heavily than those of the referent: $\alpha > \beta$. Subtracting the two directional judgements gives

$$S(a,b) - S(b,a) = (\alpha - \beta)\left[f(B - A) - f(A - B)\right].$$

When the referent b is the richer, more salient object, the prototype, the famous brand, the large country, $f(B - A) > f(A - B)$, and the variant is judged more similar to the prototype than the prototype is to the variant. Figure 30.1 plots this prediction.

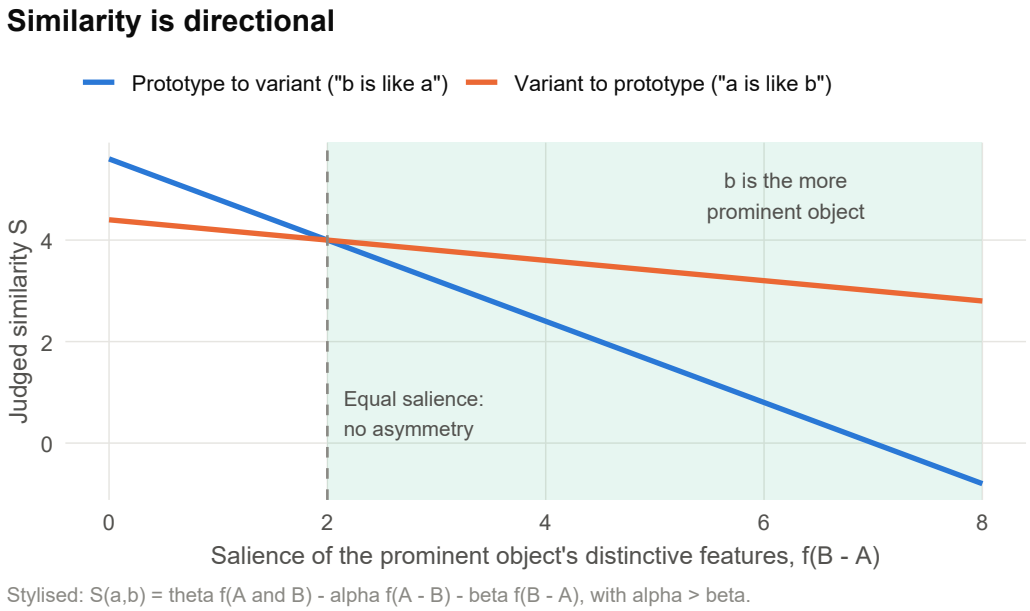


Figure 30.1. Stylised predictions of Tversky’s contrast model. As the prominent object’s distinctive features become more salient, the variant-to-prototype judgement (’a is like b’) stays high while the prototype-to-variant judgement falls. Parameters are illustrative (theta = 1, alpha = 0.8, beta = 0.2).

The contrast model generates several propositions that carry directly into marketing:

- **P1 (Directionality).** Judged similarity is higher when the less prominent object is the subject and the more prominent object is the referent.

- **P2 (Task dependence).** When people judge *difference* rather than similarity, attention shifts toward distinctive features, so the same pair can appear both more similar and more different than another pair (Tversky 1977).

30.1.2 Applications in business

Brands are feature sets that differ sharply in salience. A market leader is rich in associations. A new entrant or a private label is sparse. The contrast model therefore implies that a comparison framed with the challenger as subject ("Brand X is just like the leader") will produce more perceived similarity, and more borrowed equity, than the reverse framing, a direct implication of the contrast model (Tversky 1977). The same logic warns leaders against initiating comparisons with challengers: naming a weaker rival makes it the referent of the comparison and can lend it associations it could not otherwise claim.

For **brand extensions**, the parent brand is typically the prototype and the extension the variant. Consumers evaluate extensions more favourably when they perceive a good fit with the parent (D. A. Aaker and Keller 1990), and the contrast model suggests why: an extension inherits similarity from the parent more readily than the parent is redefined by the extension. Managers should nonetheless treat the asymmetry as a tendency rather than a guarantee: an extension with highly salient negative features can still feed back to the parent.

BUSINESS CASE

Illustrative scenario: framing a comparative advertisement. A regional coffee chain wants to position its espresso against a dominant national chain. Two scripts are tested. Script A says "Our espresso is as rich as [Leader]'s". Script B says "[Leader]'s espresso is as rich as ours". The contrast model predicts that Script A will produce higher similarity ratings because the challenger is the subject and the leader's many salient features are the referent. A sensible test uses a between-subjects design with similarity, quality, and purchase-intention measures, plus a manipulation check on brand familiarity: the asymmetry should be larger among consumers who know the leader well and the challenger poorly.

COMMON PITFALL

Assuming similarity is symmetric in perceptual maps. Multidimensional scaling and many embedding methods force symmetric distances. If the underlying judgements are directional, such maps can misstate how close a challenger is to a leader in consumers' minds. Collect directional judgements when brand prominence differs, and test for asymmetry before averaging the two directions.

30.2 Homophily

THEORY CARD

Core claim. Contact between similar people occurs at a higher rate than among dissimilar people, so personal networks are homogeneous with respect to many socio-demographic, behavioural, and intrapersonal characteristics.

Origins. Sociology, the term was popularised by Lazarsfeld and Merton in the 1950s, the

definitive review is McPherson, Smith-Lovin, and Cook (2001). In psychology, the parallel *similarity-attraction* paradigm is associated with Byrne (1971).
Level of analysis. Dyad and network.
Key constructs. Status homophily, value homophily, baseline homophily, inbreeding homophily, opportunity structure, tie formation and dissolution.
Boundary conditions. Homophily is strongest on race and ethnicity, followed by age, religion, education, occupation, and gender. It is weaker for ties that are instrumental, short-lived, or formed in diverse foci of activity (McPherson, Smith-Lovin, and Cook 2001).
Related theories. Social identity theory (Section 22.2), Weak ties (Section 9.3), Structural holes (Section 9.4), Social capital (Section 9.5).

30.2.1 Core logic and propositions

The proverb "birds of a feather flock together" names one of the most robust regularities in social science. McPherson, Smith-Lovin, and Cook (2001) distinguish **status homophily**, based on ascribed characteristics (race, ethnicity, sex, age) and acquired ones (education, occupation, religion), from **value homophily**, based on attitudes, beliefs, and values. They also distinguish two sources of the pattern. **Baseline homophily** arises simply because the pool of potential contacts is itself homogeneous: a school, a neighbourhood, or an industry offers mostly similar others. **Inbreeding homophily** is the residual similarity beyond what the opportunity structure would produce, reflecting choice, comfort, and selective tie dissolution (Table 30.2).

Table 30.2. Forms and sources of homophily (organized following McPherson, Smith-Lovin, and Cook, 2001).

Dimension	Example characteristics	Main sources	Business relevance
Status homophily (ascribed)	Race, ethnicity, sex, age	Segregated neighbourhoods, schools, workplaces	Diversity of referral pools and hiring networks
Status homophily (acquired)	Education, occupation, religion	Sorting into institutions and occupations	Board interlocks, alumni networks, B2B relationships
Value homophily	Attitudes, beliefs, tastes	Selection on similarity plus mutual influence	Echo chambers, brand communities, recommendation systems
Baseline homophily	Any attribute	Composition of the opportunity pool (foci)	Changing who meets whom: office design, events, platforms
Inbreeding homophily	Any attribute	Preference, comfort, selective tie dissolution	Bias in 'cultural fit' hiring and partner choice

Three propositions follow. Networks are more homophilous the more homogeneous the foci in which people meet (**opportunity**). Holding opportunity constant, people form ties with similar others at higher rates because similarity eases communication and validates their views (**choice**) (Byrne 1971), and ties between dissimilar people decay faster (**dissolution**). The main consequence is that homophily localises information: what circulates within a homogeneous network is redundant, which is why novel information tends to arrive through weak ties (Section 9.3) that bridge dissimilar groups.

Figure 30.2 summarises how opportunity and choice combine to produce homogeneous networks, and why the result matters.

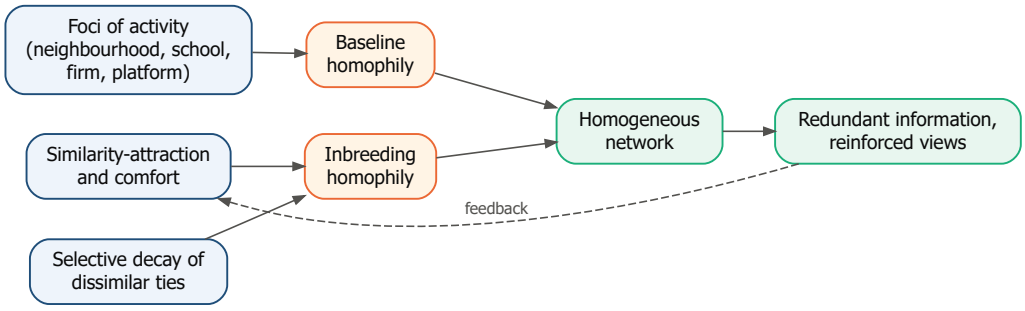


Figure 30.2. Mechanism of homophily: opportunity structures generate baseline homophily, while choice and selective tie dissolution add inbreeding homophily. Homogeneous networks then circulate redundant information and reinforce shared views.

30.2.2 Evidence and extensions

Homophily is easy to observe and hard to explain causally. If similar people are connected *and* behave alike, is that because they chose one another (selection) or because they influenced one another after connecting (contagion)? Shalizi and Thomas (2011) show that, in observational network data, the two are generically confounded: without strong assumptions or experimental variation, one cannot separate them. Aral, Muchnik, and Sundararajan (2009) illustrate the stakes. In a large instant-messaging network, they found that conventional methods substantially overestimated peer influence in the adoption of a mobile service, because much of the apparent contagion reflected homophily-driven diffusion. Longitudinal studies of email networks show that homophily emerges from both preference and the structure of shared contexts, with triadic closure amplifying small initial biases (Kossinets and Watts 2009).

BUSINESS CASE

Hiring as cultural matching. Rivera (2012) studied hiring in elite investment banks, law firms, and consulting firms. Evaluators frequently assessed candidates on shared leisure pursuits, experiences, and self-presentation: a search for "fit" that operated as a form of homophily. Because these cultural markers correlate with social class, matching on fit can reproduce socio-economic advantage beyond what credentials alone would predict. The managerial implication is not to abandon judgements of fit but to define fit in terms of job-relevant behaviours, structure interviews, and audit whether evaluators' notion of fit tracks performance.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Referral-based hiring will produce more demographically homogeneous teams than job-board hiring, controlling for applicant pools. *H2.* The estimated effect of peer adoption on a customer's adoption will fall once homophily on observed and latent traits is modelled.

Designs. Randomised exposure to peers (field experiments on platforms). Matched-sample estimation (Aral, Muchnik, and Sundararajan 2009). Longitudinal network models that separate selection from influence (stochastic actor-oriented models). Natural experiments that change foci (office moves, dorm assignment).

Measures. Homophily indices relative to group composition (to separate baseline from inbreeding homophily). Tie formation and decay rates. Attribute similarity in embeddings of text or purchases.

30.3 Assortative Mating

THEORY CARD

Core claim. Couples resemble each other on many traits, age, ethnicity, education, religion, values, even language style, more than random pairing would produce.
Origins. Population biology and demography, sociological studies of homogamy, the most cited modern analysis of educational homogamy in the United States is C. R. Schwartz and Mare (2005). Assortative mating is homophily applied to the most consequential tie people form.
Level of analysis. Dyad (the couple) and population (the distribution of pairings).
Key constructs. Homogamy versus heterogamy, marriage market, preferences versus opportunities, linguistic synchrony.
Boundary conditions. Assortment depends on who meets whom (schools, workplaces, platforms) and on preferences. It varies across cohorts, countries, and traits.
Related theories. Homophily, Social exchange theory (Section 9.1), Social magnet.

People tend to marry others of similar age, nationality, ethnic background, religion, and education. Like homophily more generally, this pattern reflects both **preferences** (attraction to similar others) and **opportunities** (people meet partners in age-graded, education-graded settings). C. R. Schwartz and Mare (2005) analysed U.S. census and survey data spanning more than six decades and found that educational homogamy declined between 1940 and about 1960 and then rose substantially through 2003. The increase was especially pronounced at the extremes: university graduates became more likely to marry one another, and the least-educated became less likely to marry up.

These trends matter beyond the family. Because couples pool income, rising educational assortment can widen household inequality even if individual earnings inequality were unchanged: a mechanism that links the micro-level choice of a partner to macro-level distributions of wealth, consumption, and children's opportunities.

Similarity extends to subtle behaviours. **Language style matching** (LSM) is the degree to which two people use function words, pronouns, articles, prepositions, conjunctions, at similar rates. Ireland et al. (2011) reported two studies. In speed-dating conversations, higher LSM predicted mutual romantic interest, that is, whether both partners wanted to see each other again. In a study of dating couples' everyday instant messages, higher LSM predicted whether the couple was still together about three months later. Because function words are produced largely without conscious control, LSM appears to index psychological attunement rather than strategic mimicry. It is a measure of verbal synchrony, not proof that matching *causes* relationship success.

BUSINESS CASE

Matching markets and recommendation design. Online dating platforms, mentoring pro-

grammes, and co-founder matching services all face the same design choice: should the algorithm match on similarity, on complementarity, or on stated preferences? Assortative mating research suggests that similarity on values and background predicts mutual interest, but it also shows that similarity-based matching concentrates advantage and reduces exposure to difference. Platforms can make the trade-off explicit, for example by mixing high-similarity recommendations with a controlled share of "exploration" matches and measuring outcomes for each.

TRY IT YOURSELF

Take a sample of email or chat exchanges between members of two project teams (with consent). Compute the share of function words in each message (pronouns, articles, prepositions) and an LSM score for each pair. Does LSM in the first week predict which pairs are still collaborating a month later? What alternative explanations, shared task, shared seniority, would you need to rule out?

30.4 Social Magnet

THEORY CARD

Core claim. Other people act like a magnet on our choices: sometimes they attract (we imitate and conform), sometimes they repel (we diverge to stay distinct), and often they do both at once.
Origins. Consumer and social psychology. Synthesised in Berger (2016), building on research on imitation (Chartrand and Bargh 1999), identity signaling (Section 8.5) (Berger and Heath 2007), and optimal distinctiveness (Section 22.3) (Brewer 1991).
Level of analysis. Individual choice within social groups.
Key constructs. Imitation, conformity, divergence, identity relevance, in-group and out-group, optimal distinctiveness.
Boundary conditions. Attraction dominates when others are similar and the domain is not identity-relevant. Repulsion dominates when the others are a dissimilar out-group or when the domain signals identity.
Related theories. Optimal distinctiveness (Section 22.3), Identity signaling (Section 8.5), Conformity (Section 22.4), Snob effect (Section 29.5), Social norms (Section 22.5).

Berger (2016) uses the image of a magnet to capture a duality that much theory treats separately. On one side, people imitate others, often without awareness: in conversation they mirror posture and mannerisms, a pattern Chartrand and Bargh (1999) called the **chameleon effect**, and mimicry smooths interactions. On the other side, people avoid choices that would make them look like someone they are not. Berger and Heath (2007) show that consumers abandon tastes once dissimilar out-groups adopt them, especially in identity-relevant domains such as music, clothing, and cars.

The two forces combine in the "same but different" pattern. Optimal distinctiveness theory holds that people seek a balance between belonging and uniqueness (Brewer 1991). Consumers therefore often choose the *category* the group chooses but a *variant* that distinguishes them: the same brand of sneaker in a less common colour, the same popular restaurant but an unusual dish (Berger 2016). Figure 30.3 organizes these forces.

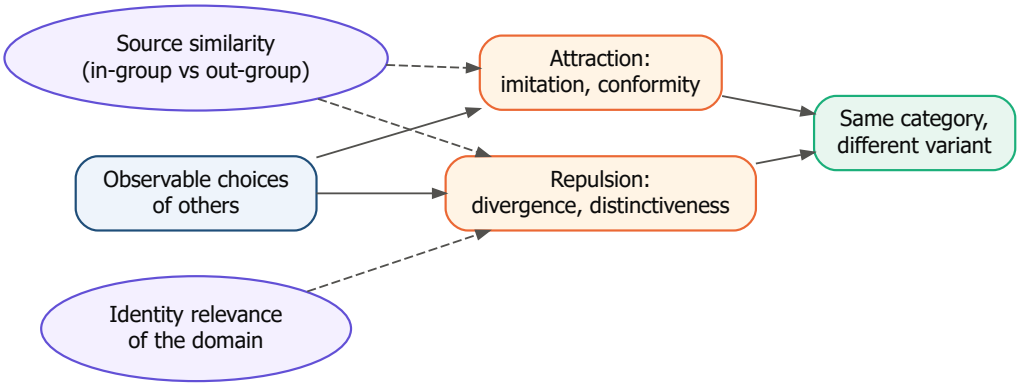


Figure 30.3. The social magnet: others' choices pull people toward imitation when the source is similar and the domain is functional, and push people toward differentiation when the source is a dissimilar out-group or the domain signals identity. The joint result is often 'same category, different variant'.

BUSINESS CASE

Positioning between belonging and distinctiveness. Product lines that offer customization within a popular platform, configurable colours on a best-selling shoe, personalised engraving on a mass-market phone case, exploit both poles of the magnet. The platform supplies social proof and belonging. The variant supplies distinctiveness. Segments high in need for uniqueness, or whose identity is threatened by an out-group's adoption of the brand, require more visible differentiation. Segments seeking reassurance require less.

30.5 Wisdom of the Crowd

THEORY CARD

Core claim. Under the right conditions, the aggregate of many independent judgements is more accurate than the typical individual judgement, and often more accurate than the best individual.

Origins. Statistics and social science, Galton's analysis of an ox-weighing competition (Galton 1907), popularised by Surowiecki (2005), formalised in the forecasting literature (Clemen 1989, Larrick and Soll 2006) and in models of diversity (Hong and Page 2004, Page 2007).

Level of analysis. Group, organization, or market.

Key constructs. Independence, diversity of information, decentralisation, aggregation rule (mean, median, market price), bracketing, correlated error.

Boundary conditions. Errors must be at least partly independent and not share a common bias, individuals must have some relevant information, a sensible aggregation rule must exist. Social influence, herding, and common information sources erode the effect (Lorenz et al. 2011).

Related theories. Pack journalism, Groupthink, Conformity (Section 22.4), Heuristics and biases (Section 23.2).

30.5.1 Origins

In 1906 Francis Galton examined the tickets from a weight-judging competition at a livestock fair in Plymouth, England, where visitors paid to guess the dressed weight of an ox. He reported that the middlemost (median) estimate of roughly eight hundred entrants was within about one per cent of the true weight, and interpreted the result as evidence for the trustworthiness of a democratic judgement: *vox populi* (Galton 1907). A century later, Surowiecki (2005) synthesised research and cases showing that crowds can outperform experts in forecasting, estimation, and problem solving, and articulated the conditions under which they do so.

30.5.2 Core logic

The statistical heart of the phenomenon is simple. Write each person's estimate as the truth plus an error. When errors are independent and centred on zero, averaging cancels them: the error of the mean shrinks with the square root of the crowd size. Larrick and Soll (2006) call this the **averaging principle** and show why it is so powerful: whenever individual estimates *bracket* the truth, some above, some below, the average is guaranteed to beat the average individual. Yet people systematically misappreciate the principle. Asked how to combine two opinions, many prefer to "chase the expert", picking the judge they believe to be better, even when simple averaging would outperform that strategy.

Page (2007) expresses the same logic as an identity, the **diversity prediction theorem**: for squared error,

$$\underbrace{(\bar{x}-\theta)^2}_{\text{crowd error}} = \underbrace{\frac{1}{n}\sum_{i=1}^{n}(x_i-\theta)^2}_{\text{average individual error}} - \underbrace{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2}_{\text{prediction diversity}},$$

where x_i are individual estimates, $\bar{x}$ is their mean, and θ is the truth. A crowd is accurate when its members are individually informed *or* when their errors are diverse. The theorem is an accounting identity, not an empirical claim, but it clarifies what managers can change: individual accuracy (through expertise and information) and diversity (through independence and heterogeneous perspectives).

The weakness of crowds lies in **correlated error**. If everyone reads the same analyst report, shares the same incentive, or hears the first speaker in a meeting, part of their error is common and no amount of averaging removes it. With pairwise error correlation ρ and individual error standard deviation σ, the error of the crowd mean has standard deviation $\sigma\sqrt{\rho + (1-\rho)/n}$, which approaches the floor $\sigma\sqrt{\rho}$ as the crowd grows. Figure 30.4 simulates this result.

30.5.3 Evidence and extensions

The averaging advantage is among the best-replicated findings in forecasting. A review of the forecast-combination literature concluded that simple combinations of forecasts, often an unweighted mean, frequently match or beat more elaborate weighting schemes and the best individual method (Clemen 1989). Laboratory evidence shows the same pattern for individual judges (Larrick and Soll 2006).

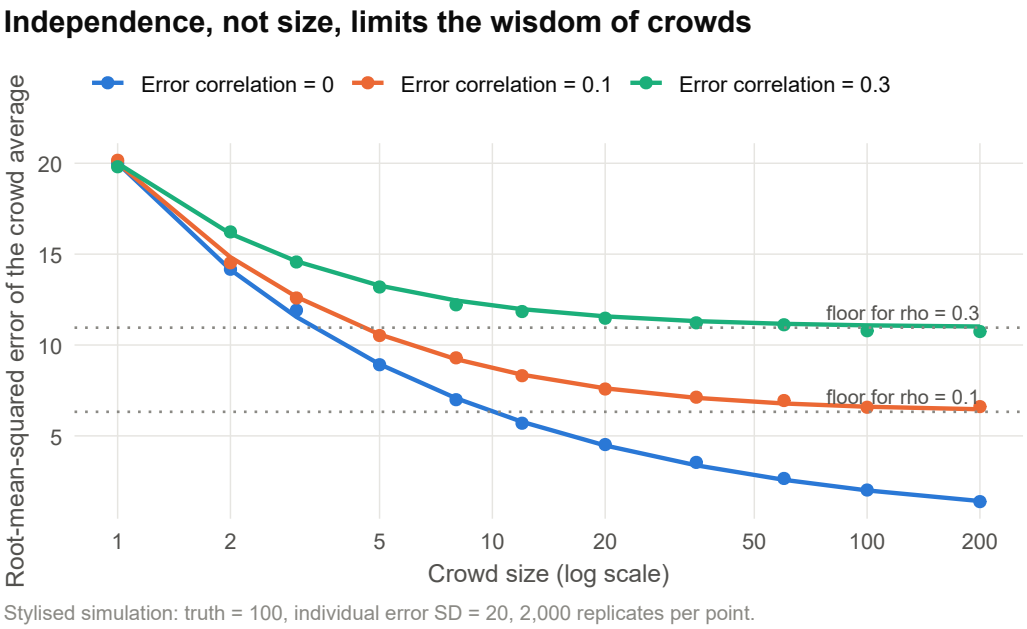


Figure 30.4. Illustrative simulation of crowd accuracy. With independent errors, the error of the crowd's average falls steadily with crowd size. With even modestly correlated errors, accuracy plateaus: beyond a few dozen members, adding people barely helps. Points are simulated, lines show the analytical expectation.

The conditions, however, are demanding. Surowiecki (2005) identifies four: **diversity** of opinion, **independence** of judgement, **decentralisation** of knowledge, and a mechanism for **aggregation**. Table 30.3 maps each condition to the way it fails and to design remedies.

Table 30.3. Conditions for crowd wisdom, failure modes, and remedies (conditions after Surowiecki, 2005).

Condition	What it requires	How it fails	Design remedy
Diversity	People hold different information or models	Homogeneous teams, common training and data	Recruit forecasters from different functions and regions
Independence	Judgements are formed before seeing others'	Anchoring on early or senior voices, herding	Collect estimates privately before discussion
Decentralisation	People draw on local, specialised knowledge	Only headquarters' view is solicited	Ask those closest to customers and operations
Aggregation	A rule turns many judgements into one	Loudest voice wins, no formal rule	Use the mean or median, or a market price

Social influence is the most common way crowds lose their wisdom. In an experiment in which participants answered factual estimation questions and, in some conditions, saw others' answers between rounds, Lorenz et al. (2011) found that social information narrowed the diversity of estimates without improving collective accuracy, while making participants more confident. A crowd that converges looks more authoritative precisely when it has become less informative. The same mechanism appears in creative tasks: in networked idea-generation, more densely connected participants were exposed to redundant inspirations and produced less innovative ideas than participants in sparser networks (Stephen, Zubcsek, and Goldenberg 2016).

Diversity of *perspectives* can matter as much as diversity of *errors*. Using an agent-based model of

problem solving, Hong and Page (2004) showed that a randomly selected group of diverse problem solvers can outperform a group of the individually best solvers, because the best solvers tend to think alike and get stuck on the same local optima.

DEBATE AND EVIDENCE

Does diversity trump ability? The Hong-Page result has been widely cited in arguments for diverse teams. Critics argue that the mathematical result depends on specific modelling assumptions and has been over-generalised to claims about demographic diversity, which the model does not address (A. Thompson 2014). A balanced reading is that cognitive diversity, different information, heuristics, and representations, can improve collective problem solving when individuals are competent and errors are not shared. Whether a particular demographic mix delivers such cognitive diversity is an empirical question.

30.5.4 Prediction markets

A **prediction market** aggregates beliefs through prices. Traders buy and sell contracts whose payoff depends on a future event, for example, a contract that pays one dollar if a product ships by a given date, and the market price can be read as a crowd probability. Wolfers and Zitzewitz (2004) review the evidence from political, sports, and corporate markets and conclude that market prices are typically reasonable, if imperfect, forecasts, often at least as accurate as polls and expert opinion. Markets add two features to simple averaging: they weight traders by their willingness to bet, and they continuously incorporate new information. They also inherit the crowd's weaknesses: thin participation, shared biases, and potential manipulation.

BUSINESS CASE

Forecasting with internal markets and simple averaging. Several large firms have run internal prediction markets on questions such as project completion dates, demand, and product quality. Cowgill and Zitzewitz (2015) analysed markets at Google, Ford, and a third firm and found that they were informative relative to available alternatives yet showed systematic biases, including optimism. The practical lesson is that a firm does not need a market to capture most of the benefit. A disciplined protocol, collect independent, private estimates from a diverse set of informed employees, aggregate them with a mean or median, and only then discuss, often captures the bulk of the averaging gain (Larrick and Soll 2006, Clemen 1989). Markets are most valuable when information is widely dispersed, changes over time, and is unlikely to surface through the hierarchy.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Forecast accuracy of an aggregated crowd will decline when members observe earlier forecasts before submitting their own. *H2.* The benefit of adding forecasters will be smaller when forecasters share a common information source.
Designs. Lab and online estimation tasks with randomised social information (Lorenz et al. 2011). Field data from corporate forecasting tournaments. Archival comparisons of consensus and individual analyst forecasts.
Measures. Mean squared error of aggregated and individual forecasts, prediction diversity, error correlation across forecasters, bracketing rate.

COMMON PITFALL

Confusing consensus with wisdom. Agreement is not evidence of accuracy. A crowd whose members talked first, read the same report, or deferred to the same leader may agree strongly and still be wrong. Always ask whether the judgements being aggregated were formed independently.

30.6 Groupthink

THEORY CARD

Core claim. Highly cohesive groups operating under structural faults and stress develop a "concurrence-seeking tendency" that suppresses dissent and critical appraisal, producing defective decisions.
Origins. Political and social psychology. Janis (1972), revised in Janis (1982), through case studies of U.S. foreign-policy decisions.
Level of analysis. Small decision-making group.
Key constructs. Cohesiveness, insulation, directive leadership, homogeneity, concurrence seeking, symptoms of groupthink, defective decision making.
Boundary conditions. Janis argued that cohesiveness produces groupthink only in combination with structural faults (insulation, lack of impartial leadership and of methodical procedures, homogeneity) and a provocative situational context (external threat, recent failure).
Related theories. Wisdom of the crowd, Conformity (Section 22.4), Spiral of silence (Section 22.8), Upper echelons theory (Section 6.4).

Irving Janis analysed decisions such as the Bay of Pigs invasion and the escalation of the Vietnam War, contrasting them with better-managed decisions such as the handling of the Cuban missile crisis (Janis 1972, 1982). He argued that the fiascoes shared a pattern: capable, loyal members of a cohesive group came to value agreement over accuracy. Figure 30.5 summarises the model, and Table 30.4 lists the eight symptoms Janis described.

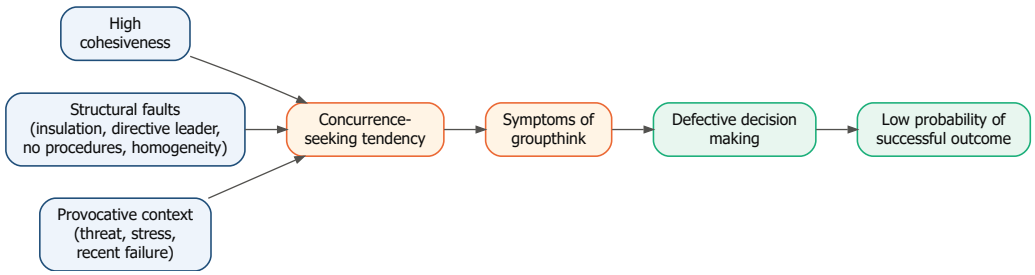


Figure 30.5. Janis's groupthink model: cohesiveness, structural faults, and a provocative context produce concurrence seeking, which manifests in symptoms of groupthink and leads to defective decision making. Redrawn from the logic in Janis (1982).

Table 30.4. The eight symptoms of groupthink (Janis, 1982) with managerial illustrations.

Symptom type	Symptom	What it looks like in a business meeting
Overestimation of the group	Illusion of invulnerability	Excessive optimism about a launch or acquisition. Risks waved away
Overestimation of the group	Belief in the group's inherent morality	Ethical concerns dismissed because 'we are the good guys'
Closed-mindedness	Collective rationalisation	Warning signals explained away rather than investigated
Closed-mindedness	Stereotyped views of out-groups	Competitors or regulators assumed to be weak or foolish
Pressure toward uniformity	Self-censorship	Members keep doubts to themselves
Pressure toward uniformity	Illusion of unanimity	Silence interpreted as agreement
Pressure toward uniformity	Direct pressure on dissenters	Sceptics labelled 'not team players'
Pressure toward uniformity	Self-appointed mindguards	Members filter out unwelcome information before it reaches the leader

DEBATE AND EVIDENCE

How well does the model hold up? Groupthink is among the most popular ideas in management education, yet its empirical support is mixed. Reviewing a quarter century of research, Esser (1998) found that case studies generally supported the model, whereas laboratory tests produced inconsistent results, particularly for the central role Janis assigned to cohesiveness. Many scholars now treat groupthink as a useful description of a family of failure modes, insulation, directive leadership, and suppressed dissent, rather than as a tightly specified causal theory.

BUSINESS CASE

Designing dissent into decisions. Janis's prescriptions remain practical: leaders should withhold their preferences at the outset, assign a devil's advocate, invite outside experts, split the group into subgroups that work independently, and hold a "second-chance" meeting before committing (Janis 1982). A related technique is the **pre-mortem**: before a decision is finalised, team members imagine that the project has failed and write down the reasons why, which legitimises dissent and surfaces risks that would otherwise be self-censored (G. Klein 2007). Both approaches restore the independence that the wisdom-of-crowds logic requires.

30.7 Pack Journalism

THEORY CARD

Core claim. Journalists from competing outlets tend to cover the same stories, use the same sources, and adopt the same framing, producing homogeneous news despite nominal competition.

Origins. Journalism studies. The term was popularised by Timothy Crouse's account of re-

porters travelling with the 1972 U.S. presidential campaign (Crouse 1973). Systematic evidence on imitation in online news comes from Boczkowski (2010).
Level of analysis. Organization (newsroom) and organizational field (news industry).
Key constructs. Competitive monitoring, shared sourcing, uncertainty about newsworthiness, imitation, homogenisation.
Boundary conditions. Strongest when journalists are physically or digitally co-present, when news value is uncertain, when the cost of missing a story others have exceeds the reward for a unique story, and when monitoring competitors is cheap.
Related theories. Wisdom of the crowd, Agenda setting (Section 26.5), Institutional theory (Section 7.2), Conformity (Section 22.4).

Crouse described the campaign press corps as a closed community: reporters travelled together, ate together, received the same briefings, and checked one another's leads, so that coverage converged on a shared interpretation of the campaign (Crouse 1973). The phenomenon did not disappear when news moved online. In some respects it intensified. Studying leading Argentine news sites, Boczkowski (2010) found that journalists monitored competitors' output continuously and that the resulting imitation increased the similarity of news content, even as the volume of available information grew. Abundance of information, paradoxically, produced homogeneity of news.

Pack journalism is an organizational instance of **herding**. Models of informational cascades show that when actors observe predecessors' choices but not their private information, it can be rational to follow the predecessors and ignore one's own signal. Once a cascade starts, later choices add no new information and the cascade can be wrong (Bikhchandani, Hirshleifer, and Welch 1992, Banerjee 1992). Reputational incentives reinforce the pattern: an editor is rarely punished for missing a story everyone missed, but is punished for missing a story everyone else ran. Figure 30.6 shows how these forces combine.

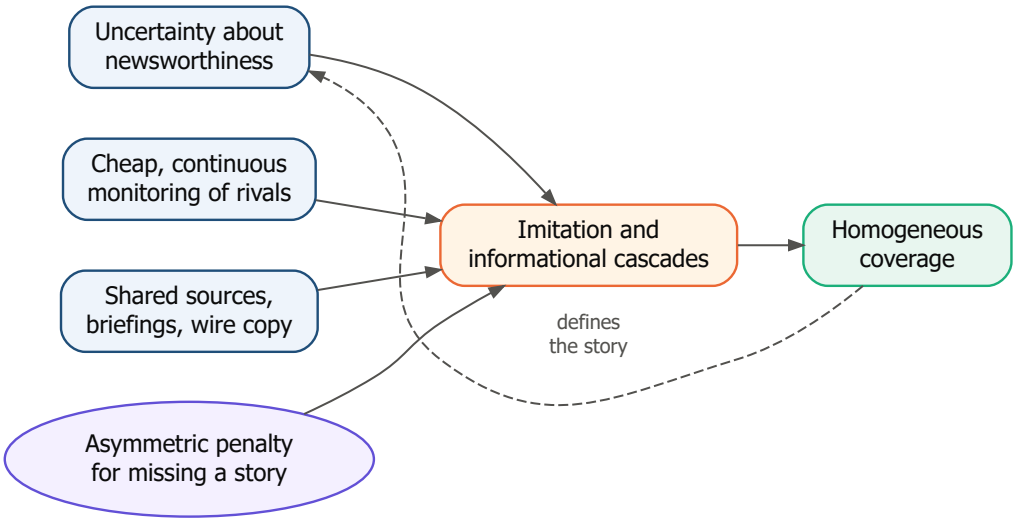


Figure 30.6. Mechanism of pack journalism. Uncertainty about newsworthiness, cheap monitoring of competitors, shared sources, and asymmetric reputational penalties lead outlets to imitate one another, producing homogeneous coverage that in turn confirms what counts as 'the story'.

BUSINESS CASE

Media monitoring and herd coverage. For communication and investor-relations teams, pack journalism has two implications. First, coverage is lumpy: once a few influential outlets frame a story, a product recall, a CEO departure, others are likely to follow the same frame, so the first hours of a crisis matter disproportionately. Second, the volume of coverage is a poor proxy for the amount of independent scrutiny. Fifty articles that rewrite one wire story contain one source of information. Media-monitoring dashboards should therefore track the number of *original* sources and frames, not just mentions, and analysts who use news-based sentiment as a signal should discount duplicated content.

30.8 Matthew Effect

THEORY CARD

Core claim. Advantage accumulates: those who receive early recognition, resources, or attention gain disproportionate access to future recognition, resources, and attention, independently of their underlying merit.
Origins. Sociology of science. Merton (1968b), drawing on interviews with Nobel laureates. Later generalised as **cumulative advantage** (DiPrete and Eirich 2006) and as **preferential attachment** in network science (Barabási and Albert 1999).
Level of analysis. Individual careers, products, and firms within a field.
Key constructs. Status, visibility, cumulative advantage, preferential attachment, path dependence, inequality and unpredictability of outcomes.
Boundary conditions. Strongest when quality is hard to observe, attention is scarce, and choices are visible to others. Weaker when quality can be assessed directly or when evaluation is blind.
Related theories. Signaling theory (Section 8.2), Social magnet, Diffusion of innovations (Section 10.1), Resource-based theory (Section 4.1).

30.8.1 Core logic

Merton (1968b) named the effect after the verse in the Gospel of Matthew that promises more to those who already have. He observed that eminent scientists receive disproportionate credit for collaborative or simultaneous discoveries, while lesser-known scientists receive less credit for comparable contributions. Merton saw two sides to the phenomenon: it is unfair to the less known, but it also helps the scientific community allocate scarce attention, because the name of an eminent scientist signals that a paper is worth reading.

The mechanism generalises well beyond science. Wherever success makes future success more likely, through visibility, reputation, resources, or network position, small early differences, including random ones, can grow into large and persistent inequalities (DiPrete and Eirich 2006). In networks, new nodes attach preferentially to already well-connected nodes, producing the highly skewed degree distributions found in citation networks and the web (Barabási and Albert 1999). In labour and entertainment markets, Rosen (1981) showed that when a performer's output can be consumed by many at little extra cost, small differences in talent translate into very large differences in earnings: the economics of superstars.

Figure 30.7 simulates cumulative advantage in a market of twenty products of *identical* quality. When consumers choose independently, market shares are nearly equal. When each consumer is more likely to choose a product in proportion to its prior popularity, a few products capture large shares, and which products win differs from one simulated world to the next.

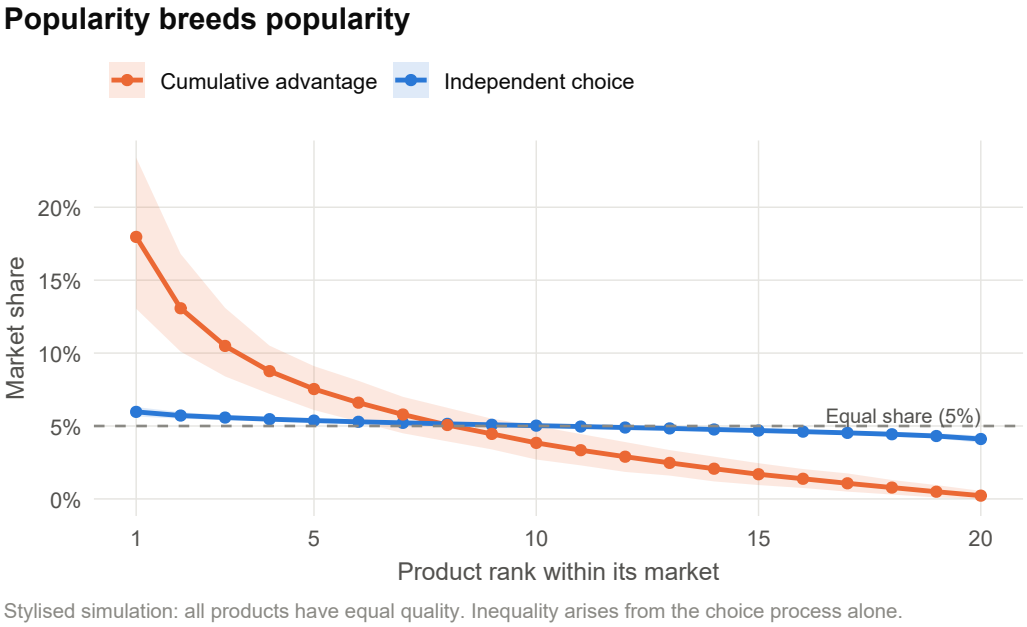


Figure 30.7. Illustrative simulation of cumulative advantage. Twenty products of identical quality compete for 2,000 sequential consumers in 300 simulated markets. Under independent choice, shares are nearly equal. Under cumulative advantage (choice probability proportional to prior adoptions + 1), the top-ranked product captures a far larger share. Lines show the mean share at each rank. Bands show the 10th-90th percentiles across markets.

30.8.2 Evidence and extensions

The most striking experimental demonstration comes from Salganik, Dodds, and Watts (2006), who created an online music market in which participants could listen to and download songs by unknown bands. Participants were randomly assigned either to an independent condition, in which they saw no information about others' choices, or to one of several "worlds" in which they saw download counts. Social influence increased both the **inequality** of success, popular songs became more popular, and its **unpredictability**: the same song could succeed in one world and fail in another. Quality mattered, since the best songs rarely did very badly and the worst rarely did very well, but it left a wide range of possible outcomes.

DEBATE AND EVIDENCE

Effect or fable? Observational evidence of a Matthew effect is confounded: high-status actors may simply produce better work. P. Azoulay, Stuart, and Wang (2014) addressed this by studying scientists who received a prestigious appointment, comparing citations to articles published *before* the appointment with carefully matched control articles. They found a status effect, but one considerably smaller than naive comparisons imply and concentrated where quality is most uncertain: for example, for more recent articles and for scientists whose repu-

tations were less established. The lesson is that cumulative advantage is real but bounded by the availability of direct information about quality.

BUSINESS CASE

Star systems in creative industries. Publishing, film, music, and gaming are markets with uncertain quality, abundant products, and visible popularity: ideal conditions for cumulative advantage. Firms respond in two ways. Some *ride* the dynamic, concentrating marketing spend on early winners, ranking products by popularity, and signing established stars whose prior success attracts attention (Rosen 1981). Others *counteract* it, because a portfolio managed purely on early popularity under-invests in good products that started unlucky. The Salganik et al. findings imply that early sales are partly noise: a publisher or platform can reserve "discovery" slots, randomise initial exposure for new releases, or evaluate titles on early engagement *per exposure* rather than on raw popularity (Salganik, Dodds, and Watts 2006).

30.9 Social Loafing

THEORY CARD

Core claim. People exert less effort when working collectively, with their inputs pooled, than when working individually or coactively.
Origins. Social psychology. Anticipated by Ringelmann's rope-pulling studies and named by Latané, Williams, and Harkins (1979). Integrated by the **collective effort model** in a meta-analysis by Karau and Williams (1993).
Level of analysis. Individual within a group.
Key constructs. Identifiability of individual contributions, evaluation potential, task meaningfulness, dispensability, group size, expectancy and instrumentality.
Boundary conditions. Reduced or eliminated when contributions are identifiable and evaluated, when the task is meaningful or personally involving, when members value the group, and when each contribution is unique.
Related theories. Expectancy theory (Section 19.2), Social facilitation (Section 22.6), Agency theory (Section 6.1), Tragedy of the commons.

Latané, Williams, and Harkins (1979) asked participants to clap and shout as loudly as they could, alone or in groups. Sound per person fell as group size rose. Part of the decline reflected coordination losses, but when participants merely *believed* they were shouting in a group, "pseudo-groups" in which they were in fact alone, effort still fell, revealing a motivational loss. The authors called it **social loafing** and traced it to the reduced identifiability of individual output.

Karau and Williams (1993) meta-analysed the subsequent literature and found social loafing to be robust across tasks and populations. They explained it with the **collective effort model**, an extension of expectancy theory (Section 19.2): people work hard to the extent that they expect their effort to matter for an outcome they value. Pooling weakens both links: effort is less clearly tied to the group result, and the group result is less clearly tied to individual rewards. The model predicts, and the meta-analysis supports, a set of moderators summarised in Table 30.5. It also predicts **social compensation**: people sometimes work *harder* collectively when the task matters to them and they expect co-workers to perform poorly.

Table 30.5. Moderators of social loafing and corresponding design levers (moderators drawn from Karau and Williams, 1993).

Moderator (collective effort model)	Effect on loafing	Team-design lever
Identifiability and evaluation of individual input	Reduces	Make individual contributions visible (named owners, contribution logs)
Task meaningfulness / personal involvement	Reduces	Connect the task to customers and outcomes members care about
Uniqueness of each member's contribution	Reduces	Assign distinct roles rather than redundant ones
Valued group (cohesion, friends, in-group)	Reduces	Invest in team identity, keep membership stable
Group size	Increases with size	Keep working units small, split large teams
Expected poor performance of co-workers	Can reverse (social compensation) on meaningful tasks	Avoid relying on compensation, it breeds resentment
Collectivist cultural orientation	Reduces	Adapt incentive design to cultural context

Social loafing is the micro-level counterpart of the free-rider problem in the economics of collective action (Section 17.2) (Olson 1965): in both, the link between private effort and private reward is diluted when the output is shared.

BUSINESS CASE

Designing teams against loafing. Consider a cross-functional product team of twelve that is evaluated only on the launch outcome. The collective effort model predicts loafing, especially among members whose contributions are hard to see. Remedies follow directly from Table 30.5: break the team into smaller pods with named owners for each deliverable. Make interim contributions visible in shared trackers. Tie part of recognition to individual or pod-level milestones, and articulate why the product matters to users. Each lever strengthens either the link from effort to performance or the link from performance to valued outcomes. Peer evaluations can help, but they work best when they assess specific behaviours rather than general impressions.

30.10 Bystander Effect

THEORY CARD

Core claim. The presence of other witnesses reduces the likelihood that any individual will help a person in need, and slows help when it does come.
Origins. Social psychology, prompted by public reaction to the 1964 murder of Kitty Genovese and tested experimentally by Darley and Latané (1968), elaborated in Latané and Darley (1970), meta-analysed by Fischer et al. (2011).
Level of analysis. Individual within a situation involving multiple witnesses.
Key constructs. Diffusion of responsibility, pluralistic ignorance, audience inhibition, the five-step intervention model, danger of the emergency.
Boundary conditions. Attenuated when the emergency is clearly dangerous, when a perpe-

trator is present, and when the costs of intervention are physical (Fischer et al. 2011). Also weakened when bystanders know one another or when responsibility is explicitly assigned. **Related theories.** Social loafing, Conformity (Section 22.4), Social norms (Section 22.5), Spiral of silence (Section 22.8).

In the classic experiment, students conversing over an intercom heard another participant apparently suffer a seizure (Darley and Latané 1968). Participants who believed they were the only listener were far more likely to seek help, and to do so quickly, than those who believed several others were also listening. The authors attributed the difference to **diffusion of responsibility**: when many could help, each feels less personally obliged.

Latané and Darley (1970) framed helping as a sequence of decisions, each of which other bystanders can derail (Figure 30.8). A bystander must notice the event, interpret it as an emergency, accept personal responsibility, decide how to help, and implement the decision. Others' presence can impede the second step through **pluralistic ignorance** (everyone looks calm, so it must not be serious), the third through diffusion of responsibility, and the fifth through **audience inhibition** (fear of embarrassment).

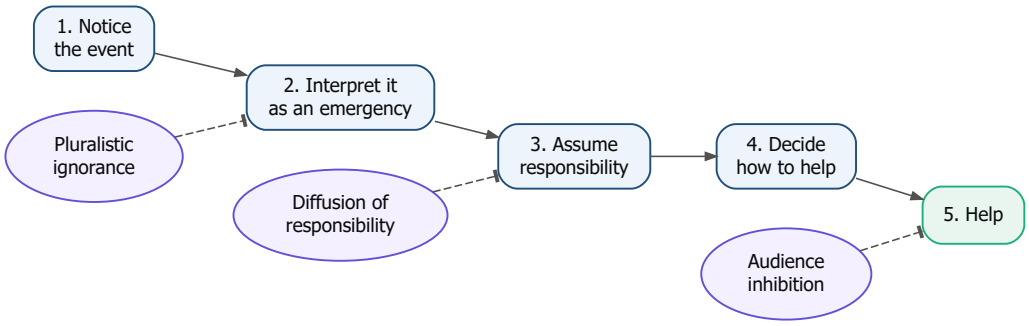


Figure 30.8. Latane and Darley's five-step model of bystander intervention, with the points at which the presence of other bystanders inhibits helping. Redrawn from the logic in Latane and Darley (1970).

DEBATE AND EVIDENCE

The parable of the 38 witnesses. The Genovese case, as popularly told, 38 neighbours watching a murder and doing nothing, inspired the research programme but is not an accurate account of the event. Manning, Levine, and Collins (2007) show that the number of witnesses, what they could see, and whether anyone called the police were misreported, and argue that the story distorted psychology's understanding of helping by framing groups only as inhibitors. The experimental phenomenon stands on its own evidence, and the meta-analysis by Fischer et al. (2011) confirms a reliable bystander effect, but with an important qualification: in clearly dangerous emergencies the effect is attenuated, plausibly because danger is recognised faster and other bystanders become potential sources of physical support.

BUSINESS CASE

Diffusion of responsibility in organizations. Many organizational failures look like bystander effects: a defect everyone on a shared channel saw but no one reported, a compliance

concern raised in a large meeting and then dropped, a customer complaint copied to ten people and answered by none. The remedies mirror the five-step model. Make anomalies salient (step 1-2: clear alerting thresholds). Assign responsibility explicitly rather than to "the team" (step 3: a named incident owner or on-call rotation). Provide scripts for what to do (step 4: escalation runbooks), and reduce the social cost of speaking up (step 5: non-punitive reporting channels). Incident-response practices that designate a single incident commander are, in effect, an institutional antidote to diffusion of responsibility.

30.11 Tragedy of the Commons

THEORY CARD

Core claim. When a resource is open to all and use by one reduces what is available to others, individually rational use can deplete the resource to everyone's detriment, unless institutions govern access and use.
Origins. Ecology and economics, Hardin (1968), the institutional response is Ostrom (1990), which documented long-enduring self-governed commons and earned Ostrom a share of the 2009 Nobel Memorial Prize in Economic Sciences.
Level of analysis. Community, organization, or institution governing a shared resource.
Key constructs. Common-pool resource (subtractable, hard to exclude), open access, externality, free riding, monitoring, graduated sanctions, collective-choice arrangements.
Boundary conditions. Tragedy is likely under open access with anonymous, numerous users and no communication. It is not inevitable where users can communicate, monitor, and craft enforceable rules.
Related theories. Social loafing, Institutional theory (Section 7.2), Transaction cost economics (Section 6.3), Stakeholder theory (Section 7.4).

30.11.1 Core logic

Hardin (1968) asked readers to imagine a pasture open to all herders. Each herder gains the full benefit of adding another animal but bears only a fraction of the cost of the resulting overgrazing, which is shared by all. Each therefore adds animals, and the pasture is ruined. Hardin concluded that the commons could be saved only by private property or by mutually agreed coercion such as regulation.

The logic is that of a social dilemma: the individually dominant strategy produces a collectively inferior outcome. It applies to fisheries, groundwater, the atmosphere, and, within firms, to shared budgets, shared computing resources, and shared attention (the meeting everyone can call, the inbox everyone can fill).

30.11.2 Ostrom's challenge

Ostrom (1990) questioned the claim that only markets or states could prevent tragedy. Drawing on case studies of communities that had managed common-pool resources for long periods, Swiss and Japanese mountain commons, irrigation institutions in Spain and the Philippines, and others, she showed that users themselves often devised durable rules. She also documented failures, and from

the comparison she distilled the **design principles** in Table 30.6 that characterise long-enduring institutions.

Table 30.6. Ostrom's design principles for long-enduring common-pool resource institutions (Ostrom, 1990), with corporate analogues.

#	Design principle	Meaning	Corporate analogue
1	Clearly defined boundaries	Who may use the resource, and the resource's limits, are clear	Who may draw on a shared cloud budget or data platform
2	Congruence with local conditions	Rules for use and contribution fit local circumstances	Usage quotas tailored to each team's workload
3	Collective-choice arrangements	Most users can take part in modifying the rules	Users sit on the governance board for the shared service
4	Monitoring	Monitors are accountable to users or are users	Transparent usage dashboards visible to all teams
5	Graduated sanctions	Violations meet sanctions that escalate with severity	Warnings, then throttling, then charge-backs
6	Conflict-resolution mechanisms	Rapid, low-cost arenas to resolve disputes	A standing forum to settle allocation disputes
7	Minimal recognition of rights to organize	External authorities do not undermine users' rules	Senior leadership respects locally agreed rules
8	Nested enterprises	Governance is layered for larger systems	Team, division, and enterprise rules that fit together

Figure 30.9 contrasts, in stylised form, the resource dynamics under open access, where effort expands as long as harvesting is profitable, with a governed regime that caps total use below the resource's regenerative capacity.

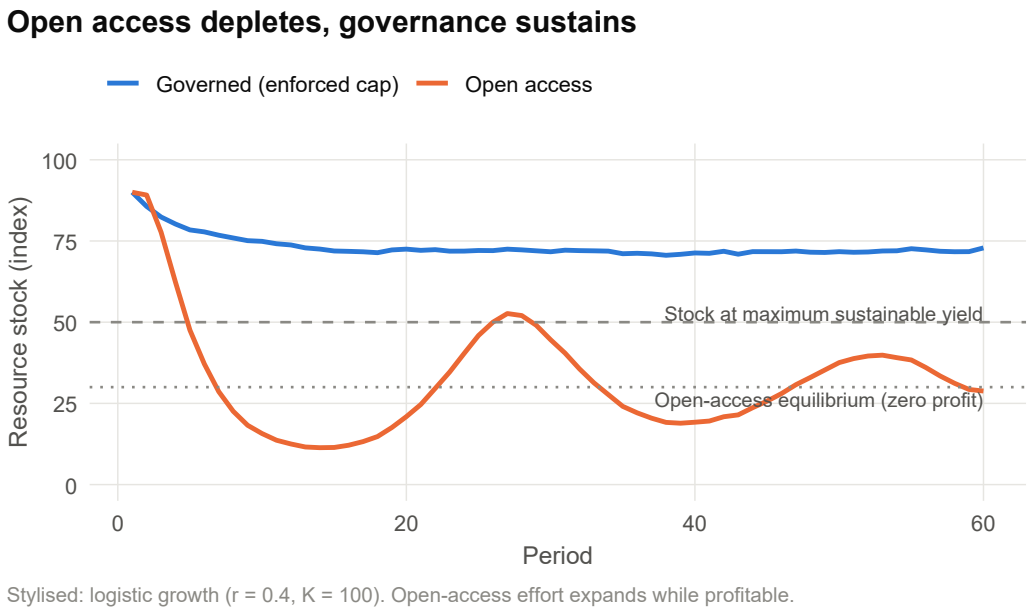


Figure 30.9. Stylised resource dynamics. Under open access, users expand effort while harvesting is profitable, driving the stock down toward the level at which profits vanish. Under a governed regime with a cap set below maximum sustainable yield and enforced by monitoring, the stock stabilises at a high level. Logistic growth with small random shocks.

BUSINESS CASE

Governing internal commons. Shared corporate resources, a central data engineering team, a cloud-computing budget, a pool of user-research participants who can be surveyed only so often, are common-pool resources. Without rules, each team draws on them as if they were free, leading to queues, overspending, and survey fatigue. Charging every use at full cost (privatisation) or centralising all allocation (regulation) are Hardin's two solutions. Ostrom's principles suggest a third: clear boundaries on who may use the resource, transparent monitoring of use, graduated responses to overuse, and a governance forum in which the user teams themselves set and revise the rules.

COMMON PITFALL

Treating every shared resource as doomed. Hardin's parable is often invoked as if tragedy were inevitable. Ostrom's evidence shows that outcomes depend on institutions. When analysing a commons problem, first describe the existing rules, formal and informal, for access, monitoring, and sanctioning, and ask which of the design principles are missing.

30.12 Scent and Emotion

THEORY CARD

Core claim. Odours are tightly linked to emotion and memory: they evoke emotionally intense autobiographical memories and colour evaluations of places and products, often without conscious awareness.
Origins. Neuroscience and cognitive psychology (E. T. Rolls 2004, Willander and Larsson 2007). Applied in marketing research on ambient scent and sensory marketing (Spangenberg, Crowley, and Henderson 1996, D. J. Mitchell, Kahn, and Knasko 1995, Krishna 2012).
Level of analysis. Individual perception, memory, and evaluation.
Key constructs. Olfactory processing, orbitofrontal cortex, odour-evoked autobiographical memory, ambient scent, scent congruence, sensory marketing.
Boundary conditions. Effects depend on the pleasantness and intensity of the scent, its congruence with the product or setting, and individual and cultural differences in odour associations.
Related theories. SOR framework (Section 31.5), Affect circumplex (Section 25.3), Feelings as information (Section 23.7), Mere exposure effect (Section 29.1).

30.12.1 Core logic

Smell occupies a distinctive place among the senses. The **orbitofrontal cortex** receives olfactory input together with taste, touch, and visual inputs and represents the reward value of stimuli. It is also central to emotion (E. T. Rolls 2004). This anatomical overlap helps explain why odours are experienced less as neutral information and more as pleasant or unpleasant, and why they so readily trigger feelings.

The link extends to memory. Willander and Larsson (2007) compared autobiographical memories cued by odours with memories cued by verbal and visual representations of the same items. Odour-cued memories were rated as more emotional and were accompanied by a stronger feeling of being

brought back in time. Smell, in short, is an unusually direct route to emotionally charged memories: the phenomenon often called the “Proust effect”.

30.12.2 Evidence in marketing

Retailers and service firms have used these properties for decades. Spangenberg, Crowley, and Henderson (1996) manipulated ambient scent in a simulated retail store and found that the presence of a pleasant scent improved evaluations of the store and its merchandise. The precise scent mattered less than whether a pleasant scent was present at all. D. J. Mitchell, Kahn, and Knasko (1995) showed that **congruence** matters for decision making: when the ambient odour fit the product category being considered, consumers spent more time processing, engaged in more holistic processing, and sought more variety. Krishna (2012) integrates such findings into a broader framework of **sensory marketing**, in which sensory inputs shape perception, which in turn affects emotion, cognition, memory, and behaviour (Figure 30.10).

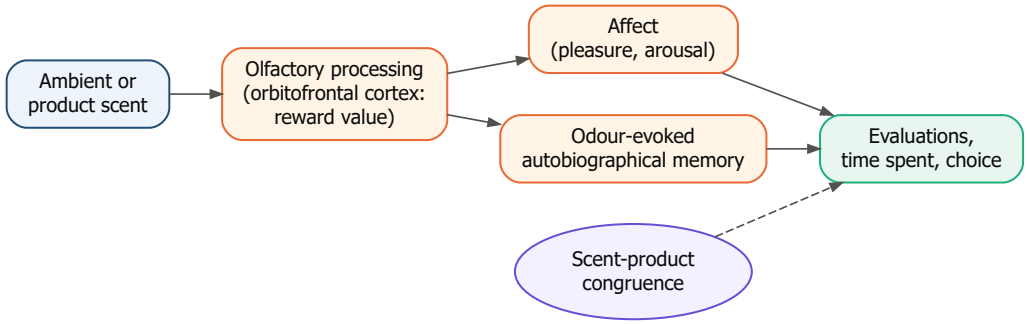


Figure 30.10. How ambient scent shapes consumer responses. Olfactory input is processed in regions that represent reward value and emotion, producing affect and evoking memories. Affect and memory then colour evaluations and behaviour. Congruence between scent and product moderates the effect.

BUSINESS CASE

Sensory branding in retail and hospitality. Hotels, airlines, and retailers have widely reported using signature scents in lobbies, cabins, and stores to create a recognisable and pleasant atmosphere. The research above suggests three design principles. First, *pleasantness and presence* do much of the work (Spangenberg, Crowley, and Henderson 1996), so a subtle, broadly liked scent is safer than an intense or polarising one. Second, *congruence* matters (D. J. Mitchell, Kahn, and Knasko 1995): a scent should fit the category and brand (fresh linen for bedding, not for a bakery). Third, because odours are strongly tied to memory (Willander and Larsson 2007), a consistent signature scent can become a retrieval cue for the brand itself, but only if it is used consistently over time and across locations. Firms should test effects with controlled store or time-slot experiments and monitor customers and employees who are sensitive to fragrances.

30.13 Bullwhip Effect

THEORY CARD

Core claim. Fluctuations in orders grow larger as they move upstream in a supply chain, from consumers to retailers, wholesalers, distributors, and manufacturers, even when end-consumer demand is stable.

Origins. System dynamics and operations management. Simulated by Forrester (1961), demonstrated experimentally in the beer distribution game (Sterman 1989), and named and analysed by H. L. Lee, Padmanabhan, and Whang (1997a) and H. L. Lee, Padmanabhan, and Whang (1997b).

Level of analysis. Supply chain (inter-organizational system).

Key constructs. Order variance amplification, demand signal processing, order batching, price fluctuation, rationing and shortage gaming, lead time, supply line, information sharing.

Boundary conditions. Strongest when members forecast from the orders they receive rather than from end demand, when lead times are long, and when promotions or allocation rules create incentives to distort orders. Production smoothing can offset it.

Related theories. Wisdom of the crowd, Bounded rationality (Section 23.1), Transaction cost economics (Section 6.3), Information asymmetry (Section 8.1).

30.13.1 Core logic

Each member of a supply chain sees only the orders placed by its immediate customer, not the demand of final consumers. When it forecasts from those orders, adjusts inventory targets, and adds its own ordering rules, it passes on a signal that is noisier than the one it received. Repeated at every stage, the noise compounds. Forrester (1961) showed with simulation models that delays and ordering policies could turn small changes in retail sales into large swings in factory production. H. L. Lee, Padmanabhan, and Whang (1997a) named the phenomenon the **bullwhip effect** and identified four causes. Crucially, they showed that each can arise from *rational* decisions by individual members, not only from mistakes (Table 30.7).

Table 30.7. The four causes of the bullwhip effect and their remedies (after Lee, Padmanabhan and Whang, 1997).

Cause	Mechanism	Typical remedy
Demand signal processing	Each stage updates forecasts and safety stocks from the orders it receives, so random changes look like trends	Share point-of-sale data, shorten lead times, vendor-managed inventory
Order batching	Fixed ordering costs and full-truckload economics lead to infrequent, lumpy orders	Lower transaction costs (electronic ordering), mixed-product shipments
Price fluctuation	Promotions and trade deals encourage forward buying, so orders no longer track consumption	Everyday low pricing, limits on forward buying
Rationing and shortage gaming	When supply is allocated in proportion to orders, buyers inflate orders and cancel later	Allocate on past sales, restrict cancellations, share capacity information

Figure 30.11 isolates the first cause. Consumer demand is stationary. Every stage uses the same reasonable rule: forecast demand with a moving average of recent incoming orders and order enough to bring inventory up to a target that covers the replenishment lead time. Nothing irrational hap-

pens, yet order variability grows at every step.

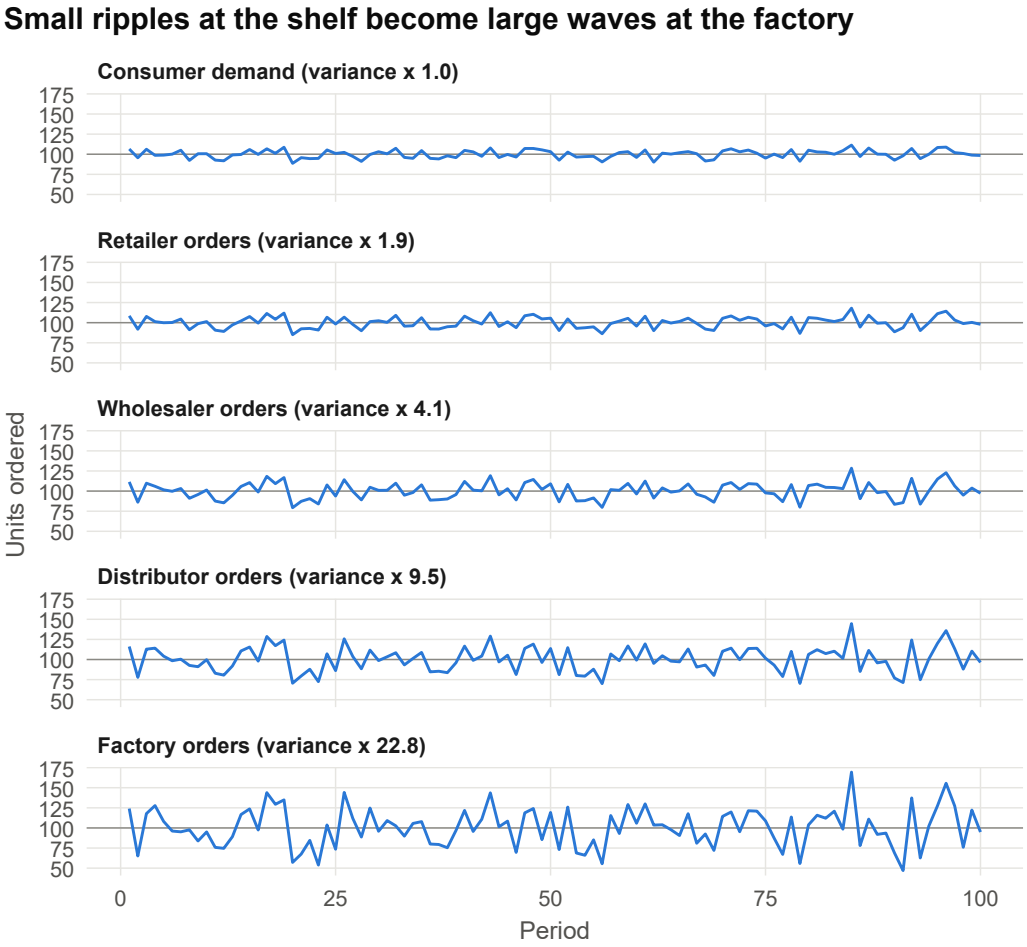


Figure 30.11. Illustrative simulation of demand signal processing. Consumer demand is stationary (mean 100, standard deviation 5). Each stage forecasts with a six-period moving average of the orders it receives and follows an order-up-to rule with a two-period lead time. Variance multiples relative to consumer demand are shown in the panel titles. Simulated data, not an actual supply chain.

The amplification in the figure is not an accident of the parameters. F. Chen et al. (2000) derived a lower bound for the increase in order variance at a single stage that uses a moving-average forecast: it grows with the lead time and falls with the number of observations in the forecast. Longer lead times and more reactive forecasts mean more amplification.

The other three causes add to this baseline, and two of them originate in marketing decisions. **Order batching** turns smooth consumption into periodic spikes because buyers accumulate demand until an order is worth placing. **Price fluctuation** is the marketing department's contribution: a trade promotion that offers a temporary discount invites retailers to buy far more than they will sell during the deal and then to stop ordering, so the manufacturer sees a surge followed by a drought that has nothing to do with consumption. **Rationing and shortage gaming** occurs when a product is scarce and the supplier allocates in proportion to orders. Buyers then order more than they need, and the phantom orders disappear as soon as supply recovers, leaving the supplier with ex-

cess capacity built for demand that never existed (H. L. Lee, Padmanabhan, and Whang 1997a). For managers, the implication is that supply-chain volatility is often self-inflicted, and that pricing and promotion policies should be evaluated for their upstream effects as well as for their effect on sales.

30.13.2 The beer game and the behavioural account

The best-known demonstration is the **beer distribution game**, a classroom simulation developed at MIT in which four players manage a retailer, a wholesaler, a distributor, and a factory. Sterman (1989) analysed the decisions of many teams. Consumer demand in the game changes only once, a single step increase early on, yet orders and inventories oscillate with growing amplitude up the chain. Sterman showed that players' decisions are well described by an anchoring-and-adjustment heuristic that gives too little weight to the **supply line**, the orders already placed but not yet received. Players reorder because inventory is low, forget that earlier orders are on the way, and then find themselves swamped. In debriefings, players often attribute the swings to erratic customer demand, even though demand changed only once. The lesson is that the structure of a system with delays can defeat intelligent people who see only part of it.

DEBATE AND EVIDENCE

Rational or behavioural, and how common? H. L. Lee, Padmanabhan, and Whang (1997a) show that the bullwhip effect can arise even when every member optimises, so remedies must change information and incentives. Sterman (1989) shows that bounded rationality adds a further, behavioural source, so training and decision support also matter. A second debate concerns prevalence. Using industry-level data, Cachon, Randall, and Schmidt (2007) found that amplification is common in wholesale industries but much less evident in many retail and manufacturing industries, where firms smooth production relative to demand. The bullwhip effect is a robust *tendency* of supply-chain structure, not a law that every firm exhibits.

BUSINESS CASE

Procter & Gamble and Pampers. In the case that gave the effect its name, H. L. Lee, Padmanabhan, and Whang (1997b) describe how Procter & Gamble's logistics executives examined orders for Pampers disposable diapers. Babies consume diapers at a steady rate, and retail sales varied only modestly. Distributors' orders to P&G, however, swung far more, and P&G's own orders to its suppliers swung more still. The pattern had no basis in consumption. It reflected forecasting from orders, batching, and the forward buying encouraged by trade promotions. The remedies discussed by Lee and colleagues map onto the four causes: sharing sell-through data with upstream partners, vendor-managed inventory and continuous replenishment, smaller and more frequent orders, and stable pricing. P&G's move toward everyday low pricing in the 1990s is often discussed as a response to the price-fluctuation cause.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Firms that receive point-of-sale data from downstream partners will show lower ratios of order variance to demand variance. *H2.* Trade-promotion intensity will increase upstream order variability more than retail sales variability. *H3.* Experimental participants who are shown the supply line will exhibit less amplification.

Measures. The bullwhip ratio (variance of orders or production divided by variance of demand), computed at the product, firm, or industry level.
Designs. Firm-level panel data on shipments and sales (Cachon, Randall, and Schmidt 2007), before-and-after studies of information-sharing programmes, and laboratory beer games that manipulate information and lead times (Sterman 1989).

30.14 Chapter Summary

KEY TAKEAWAYS

- Phenomena are robust regularities that theories explain. Knowing the **mechanism** behind a phenomenon tells managers when it will appear and how to design around it (Bogen and Woodward 1988).
- Similarity is **directional**: variants seem more similar to prototypes than the reverse, which favours challengers in comparisons (Tversky 1977).
- **Homophily** and **assortative mating** sort networks and couples by similarity through both opportunity and choice, with consequences for information flow and inequality (McPherson, Smith-Lovin, and Cook 2001, C. R. Schwartz and Mare 2005).
- Others act as a **social magnet**, attracting imitation and provoking differentiation (Berger 2016).
- Crowds are wise when judgements are **diverse, independent, and aggregated**. Social influence, shared sources (pack journalism), and cohesive insulation (groupthink) destroy that wisdom (Surowiecki 2005, Lorenz et al. 2011, Janis 1982).
- **Cumulative advantage** amplifies small early differences into large, partly arbitrary inequalities (Merton 1968b, Salganik, Dodds, and Watts 2006).
- **Social loafing**, the **bystander effect**, and the **tragedy of the commons** all arise when responsibility or reward is diffused. Identifiability, explicit assignment, and well-designed institutions counteract them (Karau and Williams 1993, Fischer et al. 2011, Ostrom 1990).
- **Scent** is closely tied to emotion and memory and can shape evaluations, especially when congruent with the product (Willander and Larsson 2007, Spangenberg, Crowley, and Henderson 1996).

Table 30.8 presents the comparison of the chapter's phenomena.

Table 30.8. Comparison of the chapter's phenomena.

Phenomenon	Core mechanism	Key moderator	Managerial lever
Similarity asymmetry	Focus on the subject's distinctive features	Difference in prominence	Frame comparisons with the challenger as subject
Homophily	Opportunity plus similarity-attraction	Composition of foci	Diversify meeting contexts and referral pools
Assortative mating	Preference and opportunity in matching markets	Educational and social sorting	Balance similarity and exploration in matching
Social magnet	Imitation versus identity-driven divergence	Identity relevance, out-group adoption	Offer variants within popular platforms
Wisdom of the crowd	Error cancellation through aggregation	Independence of errors	Collect private estimates, average them

Table 30.8. Comparison of the chapter's phenomena. *(continued)*

Phenomenon	Core mechanism	Key moderator	Managerial lever
Groupthink	Concurrence seeking	Insulation, directive leadership	Devil's advocate, pre-mortem
Pack journalism	Imitation and informational cascades	Monitoring cost, uncertainty	Track original sources, not mentions
Matthew effect	Cumulative advantage	Uncertainty about quality	Randomise exposure, judge per-exposure performance
Social loafing	Weakened effort-outcome link	Identifiability, meaningfulness	Small teams, named owners
Bystander effect	Diffusion of responsibility	Danger, explicit assignment	Assign a single accountable owner
Tragedy of the commons	Externalities under open access	Institutions for monitoring and sanctions	Apply Ostrom's design principles
Scent and emotion	Olfactory-emotional processing	Congruence, pleasantness	Subtle, congruent signature scents
Bullwhip effect	Local forecasting and ordering amplify demand signals	Lead time, batching, promotions	Share point-of-sale data, shorten lead times

30.15 Discussion Questions

1. Tversky's contrast model implies that the leader in a category should avoid comparative advertising that names a challenger. Design an experiment to test this implication and identify one boundary condition under which it might fail.
2. A platform finds that users who follow similar accounts adopt the same new features. Explain why this correlation cannot by itself establish peer influence, and propose a design that could separate homophily from contagion.
3. Your firm's quarterly demand forecast is produced in a meeting chaired by the head of sales. Using the wisdom-of-crowds conditions and the groupthink model, redesign the process. What would you measure to know whether the redesign worked?
4. Pack journalism and the Matthew effect both involve imitation. How do their mechanisms differ, and what does each imply for a start-up seeking media attention?
5. Social loafing, the bystander effect, and the tragedy of the commons have been described as variants of one underlying problem. What is that problem, and where do the three phenomena differ in mechanism and remedy?
6. A retailer plans to roll out a signature scent across 200 stores. Propose a rollout that doubles as a field experiment, specifying outcomes, controls, and potential confounds.

30.16 Further Reading

- Tversky (1977): the original statement of the contrast model. Still the best introduction to why similarity is not symmetric.
- McPherson, Smith-Lovin, and Cook (2001): the definitive review of homophily, its forms, and its sources.
- Surowiecki (2005): an accessible synthesis of when crowds are wise and when they are not. Pair it with Lorenz et al. (2011) for the experimental limits.

- Larrick and Soll (2006): a rigorous and practical analysis of why averaging works and why people underuse it.
- Salganik, Dodds, and Watts (2006): an elegant experiment showing how social influence creates inequality and unpredictability in cultural markets.
- Karau and Williams (1993): the meta-analysis and theoretical integration that turned social loafing into a design problem.
- Ostrom (1990): the classic empirical answer to Hardin. Essential for anyone who manages shared resources.
- Krishna (2012): an integrative review of sensory marketing, including scent.

Chapter 31

Models and Frameworks

A product team redesigns its onboarding flow and finds that new users remember less, not more, after a richer tutorial. A strategy team is asked whether to enter an industry with booming demand yet notoriously thin margins. An innovation manager must decide whether to fund the project that combines familiar technologies in a new way or the one that invents a component from scratch. A retailer debates whether slower background music is worth the licence fee. None of these decisions requires a new theory. Each requires a good **framework**: a structured way of laying out the parts of a problem, the relationships among them, and the questions to ask before acting.

Social scientists use the terms *theory*, *model*, and *framework* loosely, but the distinctions are useful. A **theory** explains *why* constructs are related and specifies the boundary conditions of that explanation (Whetten 1989). A **model** is a simplified representation of a specific system, often formal, such as a multi-store model of memory or a growth model of recombination. A **framework** is broader and more modest: it names the elements that matter and organizes inquiry, without always committing to a single causal mechanism. Frameworks earn their place in a theory book because they are where theory meets practice: managers rarely apply a theory directly, but they constantly apply frameworks built from theories.

This chapter presents five frameworks that recur across the book and in business practice. The **information processing model** describes how people encode, store, and retrieve information. The **Geneplore model** describes the cognitive cycle of creative generation and exploration, and **combinatorial creativity** scales that logic up to the recombination of knowledge in science, technology, and markets. **Porter's five forces**, with its companion the value net, structures the analysis of industries. The **stimulus-organism-response (SOR) framework** links environments to emotions and behaviour. The chapter ends with a capstone **mechanism toolkit** that maps the recurring causal mechanisms of this book onto the theories that use them (Table 31.1).

Table 31.1. Chapter at a glance: models and frameworks.

Framework	Core question	Level of analysis	Key source
Information processing (multi-store) model	How is information encoded, held, and retrieved?	Individual cognition	Atkinson & Shiffrin (1968)
Geneplore model	How do people generate and refine creative ideas?	Individual cognition	Finke, Ward & Smith (1992)
Combinatorial creativity	How do new ideas arise from recombining old ones?	Individual / team / economy	Weitzman (1998), Uzzi et al. (2013)

Table 31.1. Chapter at a glance: models and frameworks. *(continued)*

Framework	Core question	Level of analysis	Key source
Five forces and value net	What determines an industry's profitability?	Industry	Porter (1979, 2008), Brandenburger & Nalebuff (1996)
Stimulus-organism-response	How do environments shape emotion and behaviour?	Individual in a setting	Mehrabian & Russell (1974)
Mechanism toolkit	Which causal mechanisms recur across theories?	Cross-level	Hedstrom & Swedberg (1998)

31.1 Information Processing Model

THEORY CARD

Core claim. Human memory can be described as a flow of information through distinct stores, a sensory register, a limited-capacity short-term store, and a durable long-term store, regulated by control processes such as attention, rehearsal, and retrieval.
Origins. Cognitive psychology. Atkinson and Shiffrin (1968), building on the information-theoretic turn of the 1950s (G. A. Miller 1956). Later revised by the working-memory model (A. D. Baddeley and Hitch 1974) and challenged by the levels-of-processing framework (Craik and Lockhart 1972).
Level of analysis. Individual cognition.
Key constructs. Sensory register, short-term store (working memory), long-term store, capacity, duration, encoding, rehearsal, retrieval, control processes, chunking.
Boundary conditions. The stores are functional distinctions, not anatomical boxes. The model describes deliberate learning well. It captures less well implicit learning, emotion's influence on memory, and the depth at which information is processed.
Related theories. Processing fluency (Section 23.8), Elaboration likelihood model (Section 26.1), Bounded rationality (Section 23.1), Mere exposure effect (Section 29.1).

31.1.1 Origins and intellectual history

The cognitive revolution of the 1950s recast the mind as an information-processing system with limited capacity. In a celebrated paper, G. A. Miller (1956) observed that the span of immediate memory is limited to roughly seven, plus or minus two, items, but that an "item" can be a **chunk**: a letter, a word, or a familiar phrase. Capacity is therefore limited in chunks rather than in raw information, and expertise expands effective capacity by building larger chunks.

Atkinson and Shiffrin (1968) integrated such findings into a multi-store model. They distinguished the **structural features** of memory, permanent stores that do not vary across situations, from **control processes** that the person chooses to deploy, such as rehearsal, coding strategies, and retrieval search. Information from the environment first enters a **sensory register** that holds a rich but very brief trace. Sperling (1960) showed, using a partial-report technique, that the visual sensory trace holds more than people can report but fades within a fraction of a second. Attended information passes to the **short-term store**, which is small and loses information within seconds unless it is

rehearsed. Rehearsal both maintains information in the short-term store and transfers it to the **long-term store**, which is effectively unlimited and durable. Retrieval brings long-term information back into the short-term store, where it is used to guide responses. Figure 31.1 redraws this architecture.

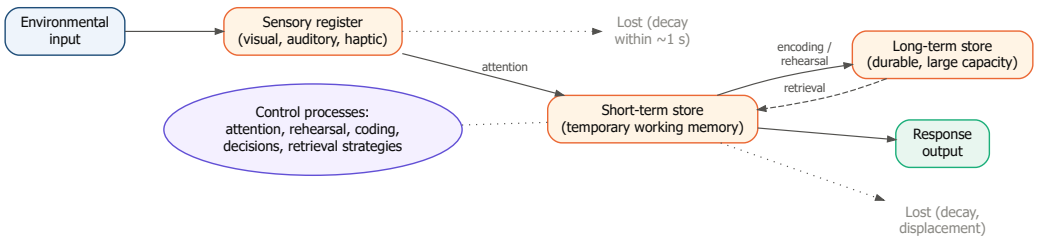


Figure 31.1. The multi-store model of memory. Environmental input passes through a sensory register to a limited short-term store, where control processes determine what is rehearsed, encoded into long-term memory, and retrieved. Information is lost from each store through decay or displacement. Redrawn from the architecture described by Atkinson and Shiffrin (1968).

Table 31.2 summarises the properties typically attributed to each store.

Table 31.2. Properties of the three stores in the multi-store model.

Store	Capacity	Duration	Main code	How information is lost	Key evidence
Sensory register	Large	Well under a second (visual)	Sensory (modality-specific)	Rapid decay, masking	Sperling (1960) partial report
Short-term store	About 7 +/- 2 chunks, about 4 when chunking is controlled	Seconds without rehearsal	Largely phonological / verbal	Decay and displacement by new items	Miller (1956), Cowan (2001)
Long-term store	Effectively unlimited	Minutes to a lifetime	Largely semantic	Interference, retrieval failure	Primacy effect, Glanzer & Cunitz (1966)

31.1.2 Core logic and propositions

The model generates several testable propositions:

- **P1 (Limited capacity).** The amount of information that can be held and manipulated at once is small. Cowan (2001) argues that, when rehearsal and chunking are prevented, the limit is closer to four chunks than to seven.
- **P2 (Rehearsal and transfer).** The more an item is rehearsed in the short-term store, the more likely it is to be transferred to long-term memory.
- **P3 (Dissociable stores).** Manipulations that affect one store (for example, a short filled delay that empties the short-term store) should leave the other store's contribution intact.

The classic test of P2 and P3 is the **serial-position curve**. When people recall a list of words in any order, recall is best for the first items (**primacy**) and the last items (**recency**), and worst in

the middle (Murdock 1962). The multi-store account attributes primacy to extra rehearsal of early items, which transfers them to long-term memory, and recency to items still held in the short-term store at the time of recall. Glanzer and Cunitz (1966) provided the critical dissociation: a filled delay of about thirty seconds before recall eliminated the recency effect but left primacy intact, whereas slowing the presentation rate improved recall of early and middle items but did not affect recency. Figure 31.2 reproduces the pattern in stylised form.

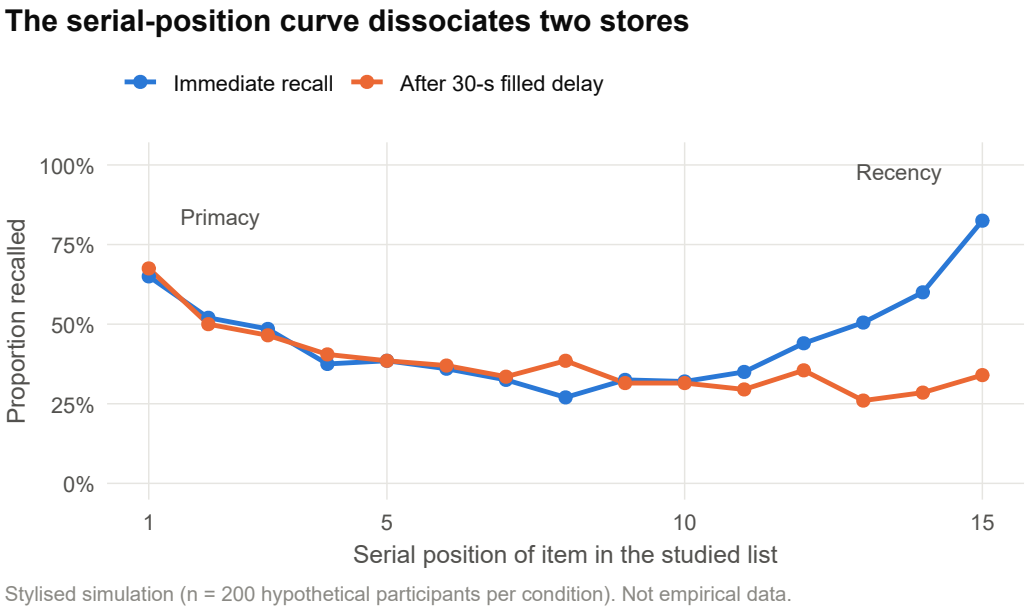


Figure 31.2. Stylised serial-position curves for immediate and delayed free recall of a 15-item list. Primacy (early items) reflects rehearsal-driven transfer to long-term memory. Recency (late items) reflects the short-term store and disappears after a filled delay. Simulated data for 200 hypothetical participants per condition. Pattern consistent with Murdock (1962) and Glanzer and Cunitz (1966).

31.1.3 Evidence and extensions

The multi-store model was enormously generative, and its successors refined rather than discarded it. Two refinements matter most.

Working memory. A. D. Baddeley and Hitch (1974) argued that the short-term store is not a single passive buffer but a **working memory** system used for reasoning, comprehension, and learning. Their model comprises a *central executive* that allocates attention, a *phonological loop* that holds speech-based information, and a *visuospatial sketchpad* for visual and spatial information. A. Baddeley (2000) later added an *episodic buffer* that integrates information from the subsystems and long-term memory into coherent episodes. The working-memory view explains why two verbal tasks interfere with each other more than a verbal and a visual task do.

Levels of processing. Craik and Lockhart (1972) challenged the idea that durability depends on which *store* holds information. They proposed instead that memory is a by-product of perceptual and conceptual analysis, and that retention depends on the **depth** of processing: meaning-based (semantic) analysis yields more durable memories than analysis of sound or appearance, and mere repetition at a shallow level adds little. The proposal shifted attention from boxes to processes.

DEBATE AND EVIDENCE

Stores or processes? Critics of the multi-store model point out that rehearsal is not one thing: maintenance rehearsal (repeating a phone number) keeps information active but does little for long-term retention, whereas elaborative rehearsal (relating information to what one knows) is highly effective (Craik and Lockhart 1972). Others note that the short-term store is better described as an activated portion of long-term memory with a small focus of attention (Cowan 2001). Most researchers now treat the multi-store model as a useful first approximation: right about capacity limits and the distinction between active and durable memory, incomplete about how encoding works.

31.1.4 Applications in business

The model is the cognitive foundation of much consumer and organizational research. Brand knowledge, for instance, is commonly conceptualised as an associative network in long-term memory, with brand awareness reflecting the strength of the brand node and brand image the associations linked to it (Keller 1993). Advertising aims to get information past attention into long-term memory and to make it retrievable at the moment of choice.

The model's most direct practical implication concerns **cognitive load**. Cognitive load theory holds that learning suffers when tasks impose demands that exceed working-memory capacity: for example, when novices must search for a solution rather than study a worked example (Sweller 1988). Mayer and Moreno (2003) distil the research into practical techniques such as *segmenting* material into learner-paced chunks, *pretraining* key concepts, and *weeding* out interesting but extraneous material.

BUSINESS CASE

Illustrative scenario: designing memorable onboarding. A software firm's first-week onboarding consists of a two-hour video tour of every feature. Usage data show that new customers adopt few features and that support tickets cluster around topics the video covered. The information processing model diagnoses the problem: the tour overloads working memory, the middle of the sequence suffers from the serial-position dip, and there is no retrieval practice. A redesign applies four principles: (1) *segment* the content into short task-based modules released as users need them (Mayer and Moreno 2003). (2) Put the most important actions at the start and end of each module. (3) Replace passive viewing with brief retrieval tasks, since practising recall improves long-term retention more than re-studying (Roediger and Karpicke 2006), and (4) *space* reminders over days rather than massing them in one session (Cepeda et al. 2006). The same logic applies to employee onboarding, compliance training, and sales enablement.

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Advertisements placed in the middle of a commercial break will be recalled less than those placed first or last. *H2.* Segmented onboarding will increase 30-day feature adoption relative to a single-session tutorial, with larger effects for more complex products.
Designs. Experiments manipulating serial position, presentation rate, and delay, A/B tests of onboarding flows, dual-task paradigms to measure load.

Measures. Free and cued recall, recognition, response times, self-reported mental effort, feature adoption and support-ticket rates as field outcomes.

COMMON PITFALL

Citing "seven plus or minus two" as a design rule. Miller's number described immediate memory span for chunks of varying size, and later work suggests a lower limit for unrehearsed, unchunked material (Cowan 2001). Menus and forms do not need exactly seven items. They need items that users can chunk meaningfully and that do not all have to be held in mind at once.

31.2 Geneplore Model

THEORY CARD

Core claim. Creative thinking alternates between a **generative** phase, in which people construct mental representations called preinventive structures, and an **exploratory** phase, in which they interpret, elaborate, and test those structures. The cycle repeats until a satisfactory creative product emerges.
Origins. Cognitive psychology of creativity. Introduced in *Creative Cognition* by Ronald Finke, Thomas Ward, and Steven Smith (Finke, Ward, and Smith 1992, see also Finke, Ward, and Smith 1996).
Level of analysis. Individual cognition, extendable to teams.
Key constructs. Generative processes, preinventive structures, preinventive properties, exploratory processes, product constraints.
Boundary conditions. Describes the cognitive processes of idea generation under normal conditions. It does not by itself predict which ideas will succeed socially or commercially.
Related theories. Combinatorial creativity, Exploration and exploitation (Section 5.3), Construal level theory (Section 20.3), Beauty in averageness (Section 29.10).

31.2.1 Core logic

The Geneplore model ("generate" plus "explore") treats creativity not as a mysterious gift but as the application of ordinary cognitive processes, retrieval, association, combination, and categorisation, in particular sequences (Finke, Ward, and Smith 1992). The key theoretical move is the **preinventive structure**: an incompletely specified idea, a visual form, a mental blend of two concepts, a novel combination of words, that has creative potential but whose meaning and use are not yet fixed. Generation produces such structures, exploration discovers what they could be. Constraints on the final product can be imposed at either phase. Figure 31.3 shows the cycle, and Table 31.3 lists the processes and structures the model names.

Table 31.3. Components of the Geneplore model (as described in Finke, Ward, and Smith, 1992).

Component	Examples
Generative processes	Retrieval of existing structures, association, mental synthesis, mental transformation, analogical transfer, categorical reduction
Preinventive structures	Visual patterns, object forms, mental blends, category exemplars, mental models, verbal combinations

Table 31.3. Components of the Geneplore model (as described in Finke, Ward, and Smith, 1992). *(continued)*

Component	Examples
Preinventive properties	Novelty, ambiguity, meaningfulness, emergence, incongruity, divergence
Exploratory processes	Attribute finding, conceptual interpretation, functional inference, contextual shifting, hypothesis testing, searching for limitations
Product constraints	Product type, category, features, functions, components, resources

A characteristic research paradigm illustrates the model. Participants are given a few simple parts, a sphere, a hook, a cylinder, and asked to combine them into an interesting form without knowing its purpose (generation), and then to interpret the form as a useful object in an assigned category such as furniture or tools (exploration). Separating the two phases allows researchers to study how ambiguity in the preinventive form and constraints in the interpretation phase each contribute to creativity (Finke, Ward, and Smith 1992).

31.2.2 Evidence and extensions

Two findings show that generation is shaped, and limited, by existing knowledge. First, imagination is **structured**. When Ward (1994) asked people to imagine creatures living on other planets, most invented creatures that preserved the core properties of Earth animals, such as bilateral symmetry and familiar sense organs and appendages. New ideas are built from the category knowledge people already have. Second, examples can cause **fixation**. Design engineers shown an example solution tended to reproduce its features, including flawed ones (Jansson and Smith 1991), and participants in creative generation tasks conformed to the features of examples even when told to produce something different (S. M. Smith, Ward, and Schumacher 1993).

These findings link the Geneplore model to older evolutionary accounts of creativity, in which ideas arise through variation and are then selectively retained (D. T. Campbell 1960), and forward to research on recombination in teams and organizations, discussed in the next section.

BUSINESS CASE

Illustrative scenario: an ideation workshop built on Geneplore. A consumer goods firm runs innovation workshops that typically yield incremental line extensions. Applying the model, facilitators split each session into two phases. In the *generation* phase, teams combine randomly drawn ingredients, formats, and usage occasions into rough concepts without evaluating them, which increases the ambiguity and novelty of preinventive structures. In the *exploration* phase, teams interpret each concept for a specific target customer, search for its limitations, and either focus it or send it back for regeneration. To reduce fixation, facilitators avoid showing competitor products before generation and introduce examples only during exploration (S. M. Smith, Ward, and Schumacher 1993, Jansson and Smith 1991).

COMMON PITFALL

Equating creativity with unconstrained brainstorming. The Geneplore model treats constraints as tools, not enemies: constraints imposed at the right phase can direct exploration and make ideas usable. The common error is to impose evaluative constraints during generation, which kills ambiguity, or to impose none at all, which yields ideas that cannot be

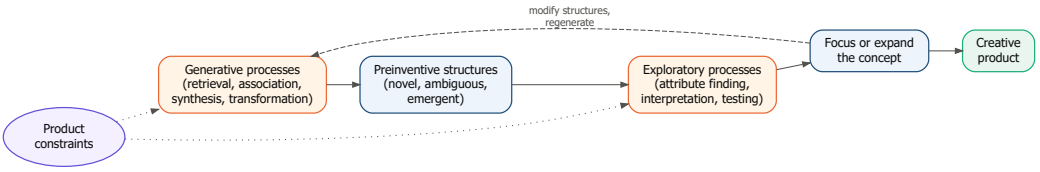


Figure 31.3. The Geneplore cycle. Generative processes produce preinventive structures with emergent properties. Exploratory processes interpret them. People then either focus on a promising concept or expand it, feeding back into a new round of generation. Product constraints can be imposed at any point. Redrawn from the model described in Finke, Ward, and Smith (1992).

implemented.

31.3 Combinatorial Creativity

THEORY CARD

Core claim. New ideas, technologies, and products are largely new combinations of existing components. The most valuable innovations tend to balance familiar combinations with a small dose of atypical ones.

Origins. Economics of innovation, from Schumpeter's "new combinations" (Schumpeter 1934) to recombinant growth theory (Weitzman 1998). Empirical work on patents (Fleming 2001, Youn et al. 2015), science (Uzzi et al. 2013), ideation networks (Stephen, Zubcsek, and Goldenberg 2016), and films (Wei 2020).

Level of analysis. Individual inventor, team, firm, and economy.

Key constructs. Components, combinations, novelty (atypicality), conventionality, recombinant uncertainty, knowledge brokering.

Boundary conditions. Strongest where knowledge is modular and can be recombined (technology, science, content). The value of novelty depends on the audience's capacity to make sense of it.

Related theories. Geneplore model, Absorptive capacity (Section 5.2), Structural holes (Section 9.4), Knowledge-based view (Section 4.3), Beauty in averageness (Section 29.10).

31.3.1 Core logic

Schumpeter (1934) defined economic development as the carrying out of "new combinations": of goods, methods of production, markets, sources of supply, and forms of organization. Weitzman (1998) formalised the idea in a model of **recombinant growth**, in which new ideas are produced by combining old ones. Because the number of possible combinations grows much faster than the number of ideas, the model implies that, in the long run, growth is limited not by a shortage of potential ideas but by the capacity to process them: to test, develop, and select among the explosion of possible combinations. Figure 31.4 illustrates the arithmetic.

If combinations are abundant, the strategic question becomes *which* combinations to try. Fleming (2001) studied this question with patent data, treating each patent's technology subclasses as its components. He found that experimenting with new components and new combinations lowered the *average* usefulness of inventions, as measured by later citations, but increased their *variability*:

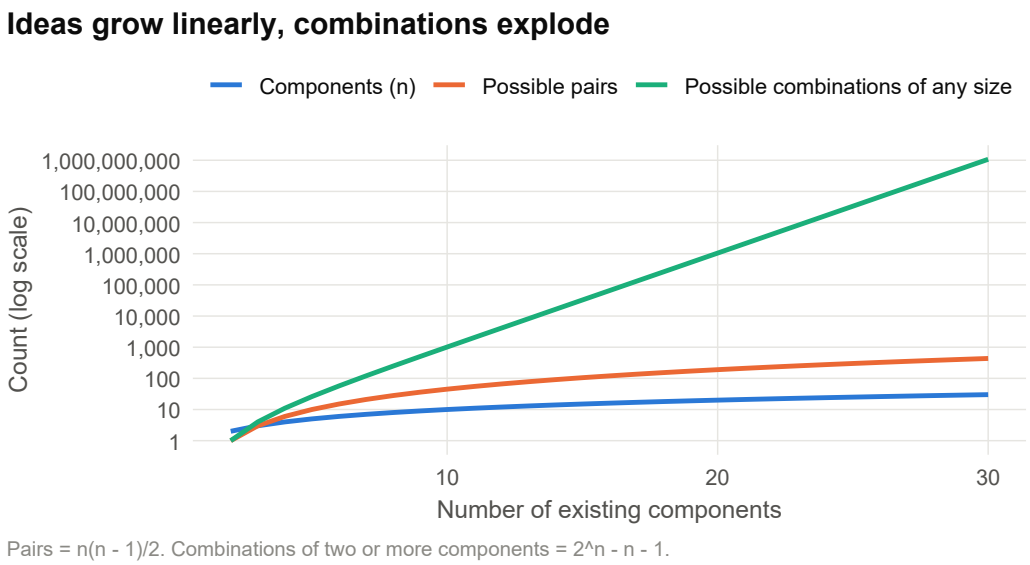


Figure 31.4. The combinatorial explosion behind recombinant growth. The number of possible pairs grows quadratically with the number of components, and the number of possible subsets of any size grows exponentially. Deterministic counts, not data.

producing more failures and also the occasional breakthrough. Search over familiar components, by contrast, is reliable but exhausts its potential. Recombination is thus a portfolio problem: novelty buys upside at the cost of predictability.

31.3.2 Evidence and extensions

Table 31.4 summarises influential empirical studies.

Table 31.4. Empirical studies of combinatorial creativity.

Study	Context	Unit of recombination	Key insight
Fleming (2001)	U.S. patents	Technology subclasses	Novel combinations lower mean usefulness but raise variance (breakthrough potential)
Uzzi et al. (2013)	Scientific papers	Pairs of journals cited together	High-impact papers pair highly conventional combinations with a tail of atypical ones
Youn et al. (2015)	U.S. patents	Technology codes	Invention is predominantly the combination of existing codes rather than the creation of new ones
Stephen et al. (2016)	Networked ideation tasks	Ideas seen from network neighbours	Lower network connectivity reduces redundant inspiration and raises innovativeness
Wei (2020)	Motion pictures	Film attributes in a similarity network	A film's position relative to prior films relates to its market performance
Hargadon & Sutton (1997)	Product design firm (IDEO)	Solutions from different industries	Brokering knowledge across industries is a systematic source of new combinations

Uzzi et al. (2013) analysed the reference lists of millions of scientific papers, scoring each pair of cited journals by how typical or atypical it was relative to chance. The papers most likely to become

highly cited were those that combined a high degree of **conventionality**, mostly familiar pairings that anchor the work in established knowledge, with a **tail of novelty**: a few highly unusual pairings. Papers that were purely conventional or purely novel were less likely to be hits, and teams were more likely than solo authors to insert novel combinations. Youn et al. (2015) found a similar combinatorial structure in U.S. patents: most invention consists of combining existing technology codes, often in new ways, rather than introducing wholly new codes. Stephen, Zubcsek, and Goldenberg (2016) show that the social structure of idea generation matters: in interdependent ideation tasks, denser networks exposed participants to redundant ideas and reduced the innovativeness of their output. And Wei (2020) constructs a similarity network among motion pictures, illustrating how the recombinant logic extends to cultural products, where a new film's relationship to existing films can be measured and related to its commercial performance.

Together, these studies suggest an **inverted-U** relationship between novelty and expected impact, combined with rising variance as novelty increases. Figure 31.5 illustrates this pattern with stylised data: average impact peaks at moderate novelty, while the upper tail of outcomes (the chance of an extreme hit) keeps rising further into the novel range.

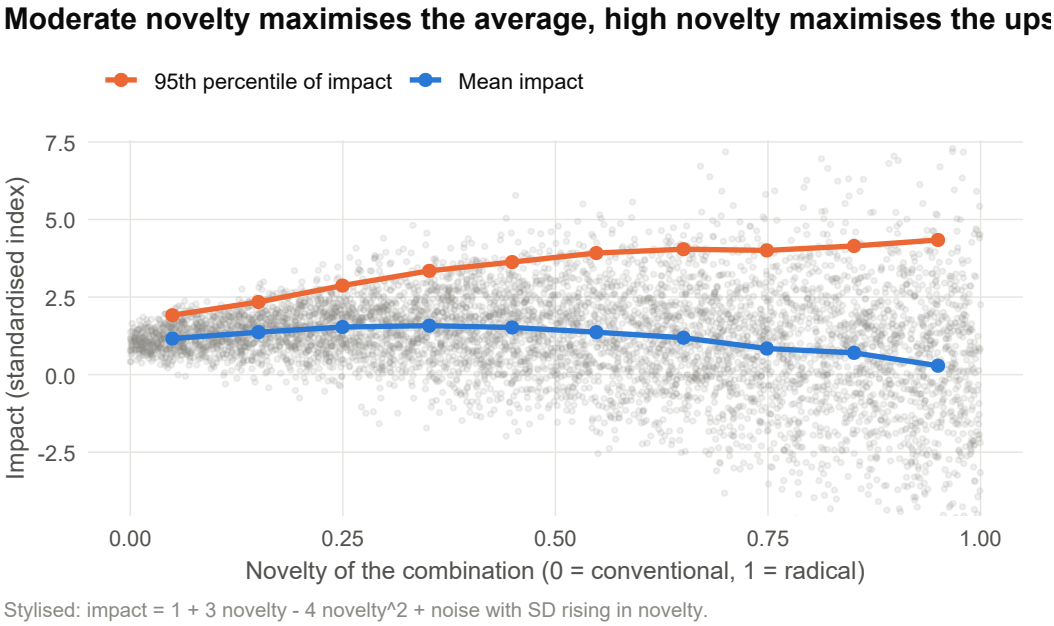


Figure 31.5. Stylised relationship between novelty and impact. Mean impact follows an inverted U (familiar combinations are reliable but unremarkable. Radical ones often fail), while the 95th percentile peaks at higher novelty because novel combinations have higher variance. Illustrative simulation of 6,000 hypothetical projects, not empirical data.

DEBATE AND EVIDENCE

How should novelty be measured? Combinatorial measures depend on how components are defined, patent subclasses, journal pairs, keywords, or film attributes, and on the baseline against which a combination counts as atypical. Different operationalisations can rank the same work differently. A second debate concerns causality: atypical combinations may be associated with impact because more talented teams both attempt and succeed at them. Researchers address these concerns with alternative baselines, author or team fixed effects,

and experiments that manipulate exposure to diverse components.

BUSINESS CASE

Knowledge brokering at IDEO and an innovation portfolio via recombination. Hargadon and Sutton (1997) studied the product design firm IDEO and showed how it innovated by acting as a **technology broker**: because it worked for clients in many industries, designers encountered solutions in one industry that could be recombined into products for another. The firm supported this with routines for capturing, storing, and retrieving design knowledge: brainstorming sessions, collections of physical objects, and a culture of asking colleagues for help. A portfolio manager can apply the same logic deliberately. *Illustrative scenario:* classify each funded project by how conventional its core components are and how many atypical pairings it contains. Allocate most of the budget to projects with a conventional core and a few novel pairings, the region where average returns peak in Figure 31.5, and a smaller share to radical combinations, judged not on average return but on their contribution to the portfolio's upside. Staff the radical projects with people who span domains, and keep ideation networks loosely connected to avoid redundant inspiration (Stephen, Zubcsek, and Goldenberg 2016).

RESEARCH TOOLKIT

Sample hypotheses. *H1.* Products that combine a conventional platform with a small number of atypical features will achieve higher average sales than either purely conventional or highly atypical products. *H2.* The variance of commercial outcomes will increase with the atypicality of feature combinations.
Designs. Archival studies of patents, papers, products, or media with combination-based novelty scores. Quantile regression to capture variance effects. Field experiments in idea contests that randomise exposure to others' ideas.
Measures. Pairwise atypicality z-scores relative to a randomised baseline (Uzzi et al. 2013). First appearance of a component pair (Fleming 2001). Similarity to prior products in an embedding space (Wei 2020). Citations, sales, or awards.

31.4 Porter's Five Forces

THEORY CARD

Core claim. An industry's long-run profitability is determined by five competitive forces: rivalry among existing competitors, the threat of new entrants, the bargaining power of buyers, the bargaining power of suppliers, and the threat of substitutes. The collectively strongest forces set the ceiling on returns.
Origins. Industrial organization economics, especially the structure-conduct-performance (Section @ref(structure-conduct-performance (Section 11.1))) tradition and research on entry barriers (Bain 1956). Translated into a managerial framework by Michael E. Porter (1979) and Michael E. Porter (1980b), with a companion article for financial analysts (Michael E. Porter 1980a). Restated in Michael E. Porter (2008).
Level of analysis. Industry (and the firm's position within it).
Key constructs. Entry barriers, supplier and buyer concentration, switching costs, substi-

tutes' price-performance, rivalry intensity, industry structure.
Boundary conditions. Assumes a definable industry with reasonably stable boundaries. It is less informative where boundaries shift rapidly, where complementors and ecosystems dominate value creation, or where firm-specific resources matter more than industry structure.
Related theories. Resource-based theory (Section 4.1), Dynamic capabilities (Section 5.1), Resource dependence theory (Section 7.1), Transaction cost economics (Section 6.3).

31.4.1 Origins and intellectual history

In the structure-conduct-performance paradigm of industrial organization, industry structure (concentration, entry barriers, product differentiation) shapes firm conduct, which determines industry performance. Bain (1956) showed that barriers to entry such as economies of scale, product differentiation, and absolute cost advantages allow incumbents to sustain prices above competitive levels. Economists traditionally viewed such barriers as a problem for public policy. Michael E. Porter (1979) inverted the perspective: what is a barrier to competition from society's point of view is, from the firm's point of view, a source of profitability to be understood and, where possible, shaped. Michael E. Porter (1980b) developed the framework into a full method for industry and competitor analysis and paired it with the generic strategies of cost leadership, differentiation, and focus.

Teece, Pisano, and Shuen (1997) later placed the five forces within a map of strategy paradigms. The **competitive forces** approach and the game-theoretic **strategic conflict** approach (Shapiro 1989) locate advantage in market position and the exploitation of market power. The **resource-based** and **dynamic capabilities** approaches locate it in firm-specific assets and the capacity to renew them. The five forces is thus one of two broad answers to the question of why firms differ in performance: because they compete in different industries, or because they bring different resources to the same industry.

31.4.2 Core logic

Figure 31.6 shows the framework. Each force captures a way in which the value created in an industry can be competed away or captured by someone other than the incumbents: by new entrants, by powerful buyers and suppliers, by substitutes that cap prices, or by rivals through price wars and costly competition.

Table 31.5 lists, for each force, the structural drivers that make it strong and typical strategic responses, following Michael E. Porter (2008).

Table 31.5. Drivers of each competitive force and typical strategic responses (organized following Porter, 2008).

Force	Strong when...	Typical strategic responses
Threat of entry	Low scale economies, weak network effects, low switching costs, low capital needs, open distribution, permissive regulation	Build scale and network effects, raise switching costs, secure channels, signal retaliation
Supplier power	Suppliers more concentrated than the industry, differentiated inputs, high switching costs, supplier can integrate forward	Multi-source, standardise inputs, develop alternatives, integrate backward

Table 31.5. Drivers of each competitive force and typical strategic responses (organized following Porter, 2008). *(continued)*

Force	Strong when...	Typical strategic responses
Buyer power	Few, large buyers, standardised products, low buyer switching costs, buyers can integrate backward, price-sensitive buyers	Differentiate, serve less powerful segments, create switching costs
Threat of substitutes	Attractive price-performance trade-off of substitutes, low cost of switching to them	Improve value relative to substitutes, reposition, bundle
Rivalry	Many or equally sized rivals, slow growth, high exit barriers, high fixed costs, perishable products, undifferentiated offerings	Differentiate, compete on non-price dimensions, avoid zero-sum price wars

Michael E. Porter (2008) also clarified what the framework is *not*. Industry growth, technology and innovation, government, and complementary products are better understood as **factors** that affect industry structure through the five forces than as additional forces in their own right. Rapid growth, for example, does not guarantee profitability if it attracts entry or empowers suppliers.

DEBATE AND EVIDENCE

How much does industry matter? If industry structure determines profits, industry should explain much of the variation in firm performance. Early variance-decomposition work found sizeable industry effects (Schmalensee 1985). Rumelt (1991), using multi-year data, found that stable business-unit effects dwarfed industry effects, a result widely read as support for the resource-based view. McGahan and Porter (1997) found that industry effects were meaningful and persistent, though business-specific effects remained larger. The emerging consensus is that both industry and firm matter, in proportions that vary across sectors, which is why the five forces is typically paired with resource-based analysis rather than used alone.

31.4.3 The value net: complements and co-opetition

A frequent criticism of the five forces is that it treats every other player as a threat to profits. Brandenburger and Nalebuff (1996) proposed the **value net**, which adds **complementors**, players whose products make customers value yours more (or make it more attractive for suppliers to supply you), and emphasises that the same player can be both competitor and complementor. Airlines compete for passengers yet cooperate in demanding aircraft that manufacturers can produce at scale. Hardware and software firms compete for margin while making each other's products more valuable. Figure 31.7 shows the structure.

Brandenburger and Nalebuff also offer a checklist for changing the game rather than merely playing it, summarised by the acronym **PARTS**: change the **players**, the **added values** each brings, the **rules**, the **tactics** that shape perceptions, and the **scope** of the game (Brandenburger and Nalebuff 1996). The value net and the five forces are complements: the five forces asks how value is captured and competed away. The value net asks how it is created and shared.

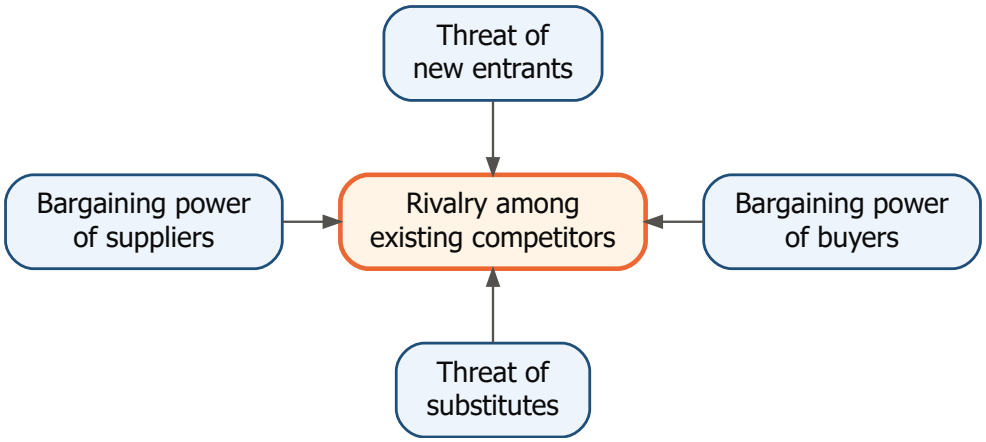


Figure 31.6. Porter's five competitive forces. Rivalry among existing competitors sits at the centre. The threat of entry, the threat of substitutes, and the bargaining power of suppliers and buyers jointly determine how much of the value an industry creates its incumbents can capture. Redrawn from Porter (1979, 2008).

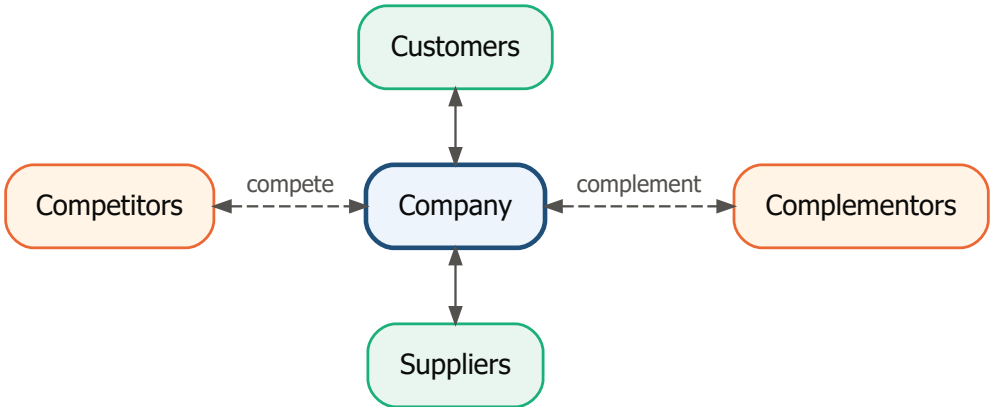


Figure 31.7. The value net. A firm interacts vertically with customers and suppliers and horizontally with competitors and complementors. The same player can occupy more than one role. Redrawn from Brandenburger and Nalebuff (1996).

31.4.4 Applications in business

BUSINESS CASE

Illustrative scenario: industry attractiveness assessment. A diversified firm is choosing between entering a regional meal-kit delivery market and an industrial testing-equipment niche. Demand for meal kits is growing faster, and the board initially favours that option. A five-forces assessment, summarised in Table 31.6, tells a different story. Meal kits face low entry barriers, low customer switching costs, close substitutes (grocery delivery, restaurants, cooking from scratch), and intense rivalry. Testing equipment benefits from certification requirements, high switching costs, and a fragmented customer base. Growth, a *factor* rather than a force, does not offset an unattractive structure (Michael E. Porter 2008). The analysis

does not end the conversation, a firm with unusual resources could still win in meal kits, but it shifts the burden of proof to identifying those resources.

Table 31.6. Illustrative five-forces attractiveness scoring for two hypothetical industries. Scores are stylised, not empirical.

Force	Meal-kit delivery (score)	Testing equipment (score)	Rationale (stylised)
Threat of entry	High (1)	Low (4)	Meal kits need little capital. Testing equipment requires certification and installed base
Supplier power	Moderate (3)	Moderate (3)	Both rely on commodity inputs plus a few specialised suppliers
Buyer power	High (2)	Low (4)	Consumers switch easily. Equipment buyers face retraining and recertification costs
Threat of substitutes	High (1)	Low (4)	Many alternatives to meal kits, few substitutes for certified testing
Rivalry	High (1)	Moderate (3)	Many funded rivals and heavy promotion versus a few specialised incumbents
Overall (mean, 1 = unattractive, 5 = attractive)	1.6	3.6	Structure favours the slower-growing niche

COMMON PITFALL

Mis-defining the industry. The most common error is drawing industry boundaries too broadly or too narrowly. Define the industry by product scope and geographic scope, asking where substitutes and rivals actually constrain pricing. A second error is treating the analysis as a static checklist. Forces change, and the value of the framework lies in anticipating how a technological, regulatory, or entry event will shift them (Michael E. Porter 2008).

31.5 Stimulus-Organism-Response Framework

THEORY CARD

Core claim. Environmental stimuli affect behaviour through the internal states they produce in the person, principally emotional states of pleasure and arousal, which in turn lead to approach or avoidance behaviours.

Origins. Environmental psychology, Mehrabian and Russell (1974). Applied to retailing by Donovan and Rossiter (1982) and to services by Bitner (1992). Rooted in a longer tradition that interposed the organism between stimulus and response (J. Jacoby 2002).

Level of analysis. Individual in a physical or digital environment.

Key constructs. Stimulus (environmental cues), organism (pleasure, arousal, dominance, cognitive and physiological states), response (approach versus avoidance).

Boundary conditions. Effects depend on the person's goals (task-oriented versus recreational shopping), personality, and culture, and on the congruence of cues with each other and with the setting.

> **Related theories.** Affect circumplex (Section 25.3), Feelings as information (Section 23.7), Scent and emotion (Section 30.12), Processing fluency (Section 23.8).

31.5.1 Core logic

Mehrabian and Russell (1974) proposed that people respond to environments along three emotional dimensions, **pleasure**, **arousal**, and **dominance** (PAD), and that these states mediate the effect of environmental features on **approach** behaviours (desire to stay, explore, affiliate, and perform) or **avoidance** behaviours (desire to leave). The pleasure and arousal dimensions later became the axes of the influential circumplex model of affect (Russell 1980). Figure 31.8 presents the framework.

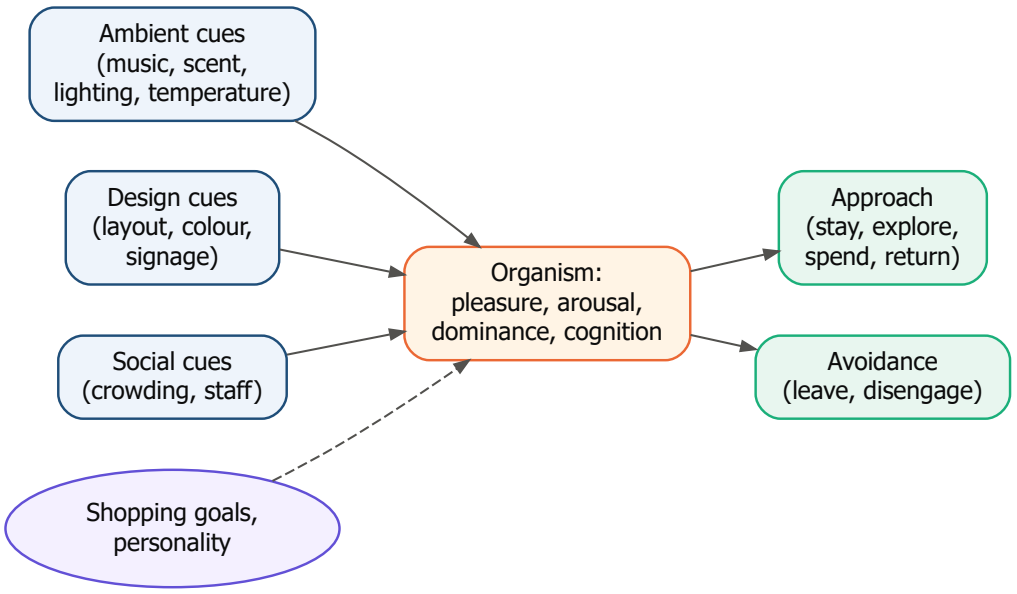


Figure 31.8. The stimulus-organism-response framework. Environmental cues produce emotional (and cognitive) states that mediate approach or avoidance responses. Shopping goals and individual differences moderate the links. Based on Mehrabian and Russell (1974) and Donovan and Rossiter (1982).

31.5.2 Evidence and extensions

Donovan and Rossiter (1982) brought the framework into retailing. Using the PAD measures in retail settings, they found that store-induced pleasure was a strong predictor of approach behaviours such as enjoyment of shopping, time spent browsing, and willingness to spend more than planned, and that arousal increased approach behaviours in environments that were already pleasant. Dominance added little. Earlier, Kotler (1973) had coined the term **atmospherics** for the deliberate design of buying environments to produce emotional effects that increase purchase probability. Bitner (1992) extended the logic to services in the **servicescape** model, which identifies ambient conditions, spatial layout and functionality, and signs, symbols, and artefacts as environmental dimensions that affect both customers *and* employees through cognitive, emotional, and physiological responses.

Field experiments have confirmed specific links. Milliman (1982) varied the tempo of background music in a supermarket and found that slower music was associated with a slower pace of in-store traffic and higher sales volume than faster music. Research on ambient scent, reviewed in the previous chapter (Section 30.12), shows similar effects of pleasant, congruent odours. Table 31.7 organizes the framework's elements.

Table 31.7. Elements of the stimulus-organism-response framework in retail and service settings.

Element	Examples	Typical measures
Stimulus: ambient	Music tempo and volume, scent, lighting, temperature	Experimental manipulation, sensor logs
Stimulus: design	Layout, colour, materials, signage, website design	Store or interface audits, A/B variants
Stimulus: social	Crowding, staff presence and behaviour, other customers	Footfall counts, observation
Organism: emotional	Pleasure, arousal, dominance	PAD semantic-differential scales, physiological measures
Organism: cognitive	Perceived quality, perceived waiting time, fluency	Survey ratings, response times
Response: approach	Dwell time, exploration, spending, return visits	Transaction and loyalty data, path tracking
Response: avoidance	Early exit, abandonment, complaints	Bounce and exit rates, complaint logs

DEBATE AND EVIDENCE

A framework or a theory? Critics note that SOR specifies *where* to look, stimuli, internal states, responses, more than *why* particular stimuli produce particular states, and that almost any mediator can be labelled an "organism" variable. J. Jacoby (2002) argues for a richer treatment of the organism, including stored knowledge and goals. In practice, SOR is most powerful when paired with a specific theory of the internal state, such as the circumplex model of affect (Russell 1980) or processing fluency (Section 23.8).

BUSINESS CASE

Retail atmospherics. A specialty retailer wants to increase dwell time and basket size without discounting. SOR suggests designing for pleasure first and calibrating arousal to the format: a calm, pleasant environment for considered purchases (furniture, wine) and a more energetic one for impulse-driven formats. Concrete levers follow from the evidence: slower music tempo where browsing is valuable (Milliman 1982). Pleasant, category-congruent scent. Lighting and layout that reduce perceived crowding, and staff behaviour that supports a sense of control. Because the servicescape also affects employees (Bitner 1992), changes should be evaluated on staff well-being and service quality as well as on sales. The firm should test levers one at a time across matched stores or time slots, as atmospheric cues interact and their effects depend on shopping goals.

31.6 A Mechanism Toolkit for Social Science Theory

The frameworks in this chapter organize problems. Theories explain them by specifying **mechanisms**. A mechanism is a set of entities and activities organized so that they produce a regular outcome: the cogs and wheels that connect cause and effect (Hedström and Swedberg 1998, Hedström and Ylikoski 2010). Mechanism-based explanation answers Whetten's "why" question (Whetten 1989): it is not enough to show that X and Y are correlated. A theory must say how X brings about Y.

Mechanism-based thinking is also the best way to see the unity behind the many theories in this book. Theories that look unrelated often share a mechanism: signaling theory (Section 8.2), screening theory (Section 8.3), and pack journalism (Section 30.7) all turn on *information*. Agency theory (Section 6.1), social loafing (Section 30.9), and the tragedy of the commons (Section 30.11) all turn on *incentives* when contributions and rewards are decoupled. Recognising the shared mechanism lets researchers transfer designs across literatures and lets managers transfer remedies across problems.

Hedström and Swedberg (1998) organize social mechanisms into three types, which correspond to the three links in a macro-micro-macro explanation (Figure 31.9). **Situational mechanisms** describe how social conditions shape individuals' beliefs, desires, and opportunities. **Action-formation mechanisms** describe how individuals combine these into action, and **transformational mechanisms** describe how individual actions aggregate into collective outcomes.

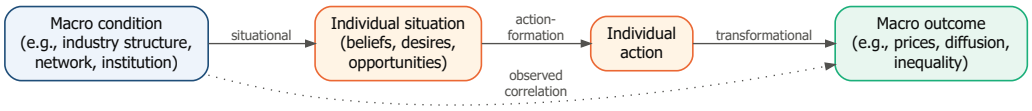


Figure 31.9. Three types of social mechanisms in a macro-micro-macro explanation. Macro conditions shape individuals' situations (situational mechanisms), individuals act (action-formation mechanisms), and actions aggregate into macro outcomes (transformational mechanisms). The dotted arrow is the macro-level correlation that mechanism-based theory seeks to explain. Based on Hedstrom and Swedberg (1998).

Table 31.8 is the capstone of the book's theoretical apparatus. It identifies ten recurring mechanisms, states the core logic of each, lists representative theories from earlier chapters, and names the managerial levers each mechanism implies.

Table 31.8. The mechanism toolkit: ten recurring mechanisms and the theories that use them.

Mechanism	Core logic	Representative theories in this book	Typical managerial levers
Information	Behaviour responds to what actors know and infer. Asymmetries create adverse selection, signals, and cascades	Information asymmetry, signaling, screening, countersignaling, wisdom of the crowd, pack journalism	Disclosure, certification, warranties, aggregation of independent judgements
Incentives	Actors respond to the pay-offs of their choices. Misaligned or diluted pay-offs produce shirking and overuse	Agency theory, transaction cost economics, expectancy theory, social loafing, tragedy of the commons	Contracts, monitoring, identifiable contributions, rules for shared resources

Table 31.8. The mechanism toolkit: ten recurring mechanisms and the theories that use them. *(continued)*

Mechanism	Core logic	Representative theories in this book	Typical managerial levers
Identity	People act to express and protect who they are and the groups they belong to	Social identity theory, optimal distinctiveness, identity signaling, snob effect, social magnet	Brand communities, positioning between belonging and distinctiveness
Social comparison and influence	People calibrate beliefs and behaviour by observing others	Social comparison, conformity, social norms, social facilitation, spiral of silence, bystander effect, groupthink	Descriptive norms, feedback relative to peers, structured dissent
Cognition and fluency	Bounded capacity and heuristics shape attention, memory, and judgement	Bounded rationality, heuristics and biases, prospect theory, processing fluency, similarity asymmetry, information processing model	Choice architecture, framing, reducing cognitive load
Emotion and affect	Feelings colour judgement and drive approach or avoidance	Broaden-and-build, affect circumplex, feelings as information, emotion and virality, SOR framework, scent and emotion	Atmospherics, emotional content, experience design
Power and dependence	Control over critical resources creates power and shapes bargaining	Resource dependence, social exchange, network exchange, power distance, five forces	Diversify dependencies, build switching costs, coalitions
Legitimacy	Organizations conform to institutional expectations to gain acceptance and resources	Institutional theory, legitimacy, stakeholder theory, organizational ecology	Certification, isomorphic practices, stakeholder engagement
Learning and capabilities	Accumulated knowledge and routines shape what firms and people can do next	Resource-based theory, knowledge-based view, dynamic capabilities, absorptive capacity, exploration-exploitation, Geneplore, combinatorial creativity	Capability building, recombination portfolios, ambidextrous structures
Networks and diffusion	Position in social structure governs access to information and the spread of behaviour	Weak ties, structural holes, social capital, diffusion of innovations, Bass model, homophily, Matthew effect	Seeding, brokerage, referral programmes, exposure design

The theories listed in the table can be found in the following sections:

- **Information:** Information asymmetry (Section 8.1), Signaling theory (Section 8.2), Screening theory (Section 8.3), Countersignaling (Section 8.6), Wisdom of the crowd (Section 30.5), Pack journalism (Section 30.7).
- **Incentives:** Agency theory (Section 6.1), Transaction cost economics (Section 6.3), Expectancy theory (Section 19.2), Social loafing (Section 30.9), Tragedy of the commons (Section 30.11).
- **Identity:** Social identity theory (Section 22.2), Optimal distinctiveness (Section 22.3), Identity signaling (Section 8.5), Snob effect (Section 29.5), Social magnet (Section 30.4).
- **Social comparison and influence:** Social comparison theory (Section 22.1), Conformity (Section 22.4), Social norms (Section 22.5), Social facilitation (Section 22.6), Spiral of silence (Section 22.8), Bystander effect (Section 30.10), Groupthink (Section 30.6).
- **Cognition and fluency:** Bounded rationality (Section 23.1), Heuristics and biases (Section

23.2), Prospect theory (Chapter 24), Processing fluency (Section 23.8), Similarity asymmetry (Section 30.1), Information processing model.
- **Emotion and affect:** Broaden-and-build theory (Section 25.1), Affect circumplex (Section 25.3), Feelings as information (Section 23.7), Emotion and virality (Section 25.4), SOR framework, Scent and emotion (Section 30.12).
- **Power and dependence:** Resource dependence theory (Section 7.1), Social exchange theory (Section 9.1), Network exchange theory (Section 9.2), Power distance (Section 18.3), Five forces.
- **Legitimacy (Section @ref(legitimacy (Section 7.3))):** Institutional theory (Section 7.2), Legitimacy (Section 7.3), Stakeholder theory (Section 7.4), Organizational ecology (Section 7.5).
- **Learning and capabilities:** Resource-based theory (Section 4.1), Knowledge-based view (Section 4.3), Dynamic capabilities (Section 5.1), Absorptive capacity (Section 5.2), Exploration and exploitation (Section 5.3), Geneplore model, Combinatorial creativity.
- **Networks and diffusion:** Weak ties (Section 9.3), Structural holes (Section 9.4), Social capital (Section 9.5), Diffusion of innovations (Section 10.1), Bass model (Section 10.2), Homophily (Section 30.2), Matthew effect (Section 30.8).

Many theories draw on more than one mechanism, and these bridging theories are often the most generative. Figure 31.10 plots a selection of theories as a network in which each theory is linked to its primary mechanism (solid) and, where relevant, to a secondary mechanism (dashed). Theories cluster around their primary mechanisms, while bridging theories, such as identity signaling (identity and information), the wisdom of the crowd (information and social influence), and the five forces (power and incentives), sit between clusters.

RESEARCH TOOLKIT

Using the toolkit in research. (1) *Name the mechanism.* For any hypothesised X-Y relationship, state which of the ten mechanisms links them and what intervening variable you could measure (beliefs for information, pay-offs for incentives, identification for identity). (2) *Look for rival mechanisms.* The same correlation often admits several mechanisms, peer adoption can reflect information, social influence, or homophily, and a design that separates them is a stronger contribution than one that merely documents the correlation. (3) *Borrow designs across literatures.* A remedy or identification strategy that works for one theory often transfers to others that share its mechanism. (4) *Specify the level.* Indicate whether the mechanism is situational, action- forming, or transformational (Hedström and Swedberg 1998), and measure it at that level.

TRY IT YOURSELF

Pick a business problem you know well: declining customer retention, slow adoption of a new internal tool, or rising supplier prices. Identify the two mechanisms from Table 31.8 most likely to be at work. For each, name one theory from this book that uses the mechanism, one observable indicator that the mechanism is operating, and one intervention it suggests. Where the two mechanisms imply different interventions, design a small test that would tell you which one dominates.

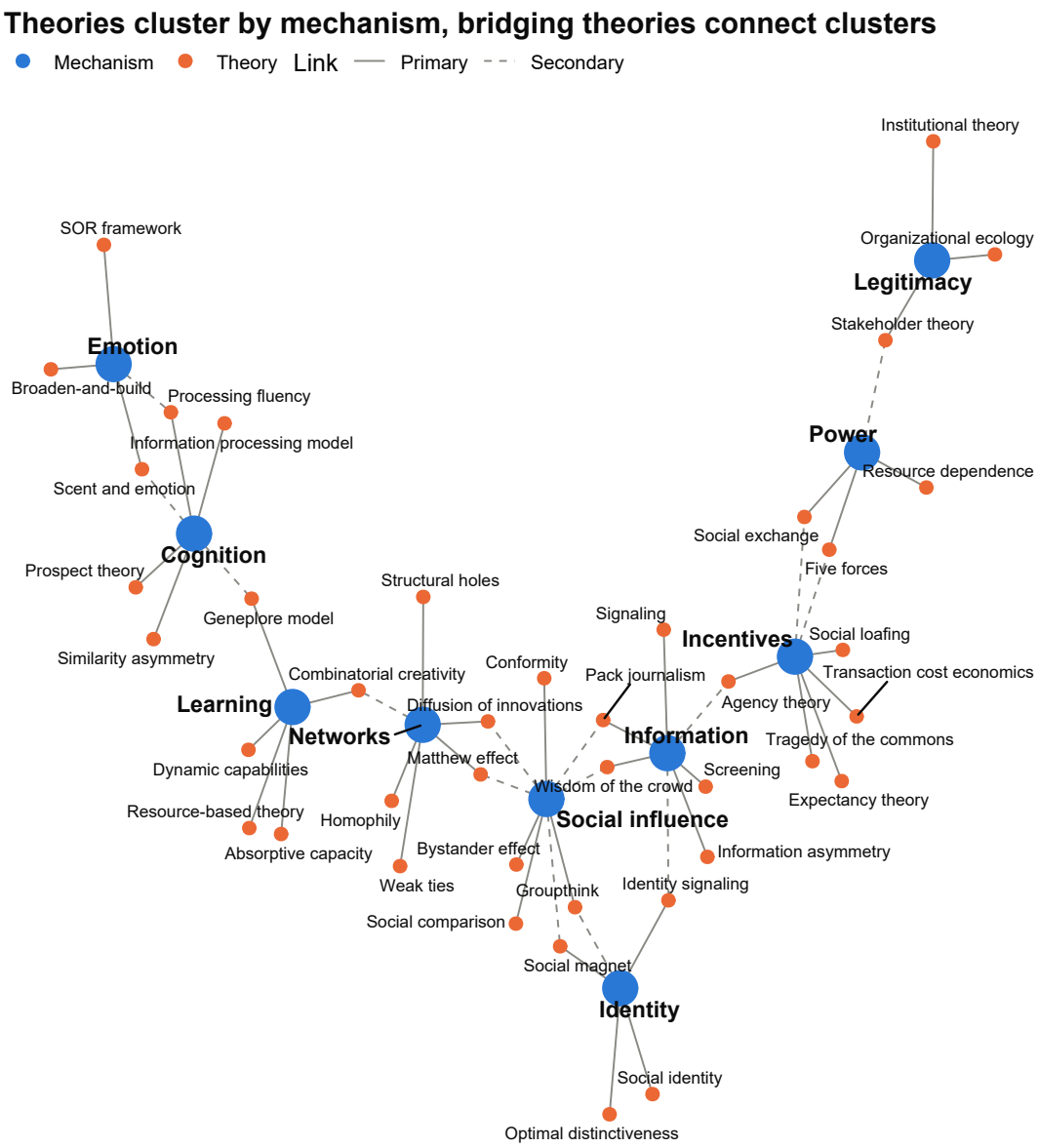


Figure 31.10. A map of selected theories by mechanism. Large nodes are mechanisms. Small nodes are theories from this book. Solid edges link a theory to its primary mechanism. Dashed edges to a secondary mechanism. The classification is the author's interpretive synthesis, not an empirical estimate.

31.7 Chapter Summary

KEY TAKEAWAYS

- **Frameworks** organize problems. **Theories** explain them by specifying mechanisms and boundary conditions (Whetten 1989).
- The **multi-store model** describes memory as a flow through sensory, short-term, and long-term stores governed by control processes. Capacity limits and the serial-position

curve are its signature evidence (Atkinson and Shiffrin 1968, G. A. Miller 1956, Glanzer and Cunitz 1966). Working memory and levels of processing refine it (A. D. Baddeley and Hitch 1974, Craik and Lockhart 1972).

- The **Geneplore model** treats creativity as a cycle of generating preinventive structures and exploring them, shaped by existing knowledge and vulnerable to fixation (Finke, Ward, and Smith 1992, Ward 1994).
- **Combinatorial creativity** holds that innovation is recombination. The most impactful work blends conventional combinations with a tail of novelty, and novel combinations raise variance as well as potential (Weitzman 1998, Fleming 2001, Uzzi et al. 2013).
- The **five forces** explain industry profitability through entry, supplier and buyer power, substitutes, and rivalry. The **value net** adds complementors and co-opetition (Michael E. Porter 2008, Brandenburger and Nalebuff 1996).
- The **SOR framework** links environmental cues to approach or avoidance through emotional states, with direct applications in retail atmospherics (Mehrabian and Russell 1974, Donovan and Rossiter 1982).
- A small set of **recurring mechanisms**, information, incentives, identity, social influence, cognition, emotion, power, legitimacy, learning, and networks, underlies the theories in this book (Hedström and Swedberg 1998).

Table 31.9 presents the comparison of the chapter's models and frameworks.

Table 31.9. Comparison of the chapter's models and frameworks.

Framework	Core elements	Central mechanism	Signature evidence or tool	Typical business use
Information processing model	Sensory register, short-term store, long-term store, control processes	Cognition (capacity limits)	Serial-position curve	Onboarding, training, advertising recall
Geneplore model	Generation, preinventive structures, exploration, constraints	Cognition and learning	Creative invention tasks, fixation studies	Ideation workshops, design processes
Combinatorial creativity	Components, combinations, novelty, conventionality	Learning and networks	Patent and citation atypicality	Innovation portfolios, R&D search
Five forces / value net	Entry, suppliers, buyers, substitutes, rivalry, complementors	Power and incentives	Industry analysis, variance decomposition	Entry, exit, and positioning decisions
SOR framework	Stimulus, organism (PAD), response	Emotion	PAD scales, atmospheric experiments	Store, service, and digital experience design
Mechanism toolkit	Situational, action-formation, transformational mechanisms	All ten mechanisms	Synthesis table and map	Diagnosing problems, transferring remedies

31.8 Discussion Questions

1. The levels-of-processing framework argues that depth of processing, not the store in which information resides, determines retention. Design an advertising experiment that would discriminate between the multi-store and levels-of-processing accounts of brand recall.
2. Using the Geneplore model, explain why showing a team competitors' products at the start of an ideation session might reduce the novelty of its ideas. How would you test this in a field

setting?

3. Uzzi et al. (2013) find that high-impact papers combine conventional and novel pairings. Propose an analogous measure for new consumer products and a hypothesis about its relationship to sales variance.
4. Apply the five forces and the value net to the same industry. Where do the two frameworks lead to different strategic recommendations, and why?
5. The SOR framework is sometimes criticised as a framework rather than a theory. Which specific theory of the "organism" would you pair it with to explain the effect of music tempo on spending, and what would that pairing predict?
6. Choose two theories from different chapters that share a mechanism in Table 31.8. Describe a research design or managerial remedy from one literature that could be transferred to the other.

31.9 Further Reading

- Atkinson and Shiffrin (1968): the original statement of the multi-store model, with a careful distinction between memory structures and control processes.
- A. D. Baddeley and Hitch (1974): the working-memory model that reshaped research on short-term memory. Read alongside Craik and Lockhart (1972) for the main alternative.
- Finke, Ward, and Smith (1992): the book that introduced the Geneplore model and the creative cognition approach.
- Uzzi et al. (2013): a landmark large-scale study of how conventional and atypical combinations shape scientific impact. Pair it with Fleming (2001).
- Michael E. Porter (2008): Porter's own restatement of the five forces, including common errors in applying them.
- Brandenburger and Nalebuff (1996): the value net and the logic of co-opetition, written for managers.
- Donovan and Rossiter (1982): the study that brought environmental psychology and the SOR framework into retailing.
- Hedström and Swedberg (1998): the foundational collection on mechanism-based explanation in the social sciences.

Part VII

Reference

Chapter 32

Theory Atlas

The Theory Atlas collects every entry in the book in one place. Use it to look up a theory's core claim, find theories that operate through the same mechanism, or locate candidates for a research question. In the online edition the table is searchable, sortable, and filterable. Each theory name links to its full entry.

32.1 Searchable atlas

Table 32.1 lists every entry in the book.

Table 32.1. The Theory Atlas: every entry in the book.

Theory	Ch.	Mechanism	Key source
Levels of analysis	2	Explanation	Klein, Dansereau & Hall (1994)
Theory building blocks	2	Explanation	Whetten (1989)
Mediation and moderation	3	Explanation	Baron & Kenny (1986)
Replication crisis	3	Explanation	Open Science Collaboration (2015)
Knowledge-based view	4	Resources & capabilities	Grant (1996)
Market-based assets	4	Resources & capabilities	Srivastava, Shervani & Fahey (1998)
Resource-advantage theory	4	Resources & capabilities	Hunt & Morgan (1995)
Resource-based view	4	Resources & capabilities	Barney (1991)
VRIO framework	4	Resources & capabilities	Barney (1991)
Absorptive capacity	5	Learning & adaptation	Cohen & Levinthal (1990)
Dynamic capabilities	5	Learning & adaptation	Teece, Pisano & Shuen (1997)
Exploration vs. exploitation	5	Learning & adaptation	March (1991)
Organizational ambidexterity	5	Learning & adaptation	Tushman & O'Reilly (1996)
Agency theory	6	Incentives	Jensen & Meckling (1976)
Stewardship theory	6	Motivation & goals	Davis, Schoorman & Donaldson (1997)
Transaction cost economics	6	Incentives	Williamson (1985)

Table 32.1. The Theory Atlas: every entry in the book. *(continued)*

Theory	Ch.	Mechanism	Key source
Upper echelons theory	6	Cognition & heuristics	Hambrick & Mason (1984)
Contingency theory	7	Learning & adaptation	Lawrence & Lorsch (1967)
Corporate social responsibility pyramid	7	Legitimacy & institutions	Carroll (1991)
Creating shared value	7	Incentives	Porter & Kramer (2011)
Institutional theory	7	Legitimacy & institutions	DiMaggio & Powell (1983)
Legitimacy	7	Legitimacy & institutions	Suchman (1995)
Organizational ecology	7	Learning & adaptation	Hannan & Freeman (1977)
Resource dependence theory	7	Power & dependence	Pfeffer & Salancik (1978)
Stakeholder theory	7	Power & dependence	Freeman (1984)
Costly signaling	8	Information	Zahavi (1975)
Countersignaling	8	Information	Feltovich, Harbaugh & To (2002)
Identity signaling	8	Identity & social comparison	Berger & Heath (2007)
Information asymmetry	8	Information	Akerlof (1970)
Screening theory	8	Information	Rothschild & Stiglitz (1976)
Signaling theory	8	Information	Spence (1973)
Commitment-trust theory	9	Networks & exchange	Morgan & Hunt (1994)
Network exchange theory	9	Power & dependence	Cook & Emerson (1978)
Social capital	9	Networks & exchange	Coleman (1988)
Social exchange theory	9	Networks & exchange	Blau (1964)
Strength of weak ties	9	Networks & exchange	Granovetter (1973)
Structural holes	9	Networks & exchange	Burt (1992)
Bass diffusion model	10	Diffusion & adoption	Bass (1969)
Diffusion of innovations	10	Diffusion & adoption	Rogers (2003)
Information systems success model	10	Diffusion & adoption	DeLone & McLean (1992, 2003)
Task-technology fit	10	Diffusion & adoption	Goodhue & Thompson (1995)
Technology acceptance model	10	Diffusion & adoption	Davis (1989)
Theory of planned behavior	10	Motivation & goals	Ajzen (1991)
UTAUT	10	Diffusion & adoption	Venkatesh et al. (2003)
Attention-based view	11	Cognition & heuristics	Ocasio (1997)

Table 32.1. The Theory Atlas: every entry in the book. *(continued)*

Theory	Ch.	Mechanism	Key source
Behavioral theory of the firm	11	Learning & adaptation	Cyert & March (1963)
Competitive dynamics	11	Incentives	Chen (1996)
First-mover advantage	11	Resources & capabilities	Lieberman & Montgomery (1988)
Game theory	11	Incentives	Nash (1950)
Real options theory	11	Learning & adaptation	Myers (1977), Bowman & Hurry (1993)
Strategic groups	11	Resources & capabilities	Caves & Porter (1977)
Structure-conduct-performance	11	Power & dependence	Bain (1956)
Architectural innovation	12	Information	Henderson & Clark (1990)
Bricolage	12	Resources & capabilities	Baker & Nelson (2005)
Creative destruction	12	Learning & adaptation	Schumpeter (1942), Aghion & Howitt (1992)
Disruptive innovation	12	Incentives	Bower & Christensen (1995), Christensen (1997)
Dominant design and technology cycles	12	Diffusion & adoption	Abernathy & Utterback (1978), Anderson & Tushman (1990)
Effectuation	12	Cognition & heuristics	Sarasvathy (2001)
Liability of newness	12	Legitimacy & institutions	Stinchcombe (1965), Freeman, Carroll & Hannan (1983)
Open innovation	12	Networks & exchange	Chesbrough (2003), Laursen & Salter (2006)
Opportunity discovery and creation	12	Information	Shane & Venkataraman (2000), Alvarez & Barney (2007)
Profiting from innovation	12	Resources & capabilities	Teece (1986)
Born globals	13	Networks & exchange	Oviatt & McDougall (1994)
Institutional voids	13	Legitimacy & institutions	Khanna & Palepu (1997)
Internalization theory	13	Incentives	Buckley & Casson (1976)
Liability of foreignness	13	Legitimacy & institutions	Zaheer (1995)
OLI paradigm	13	Resources & capabilities	Dunning (1980)
Uppsala model	13	Learning & adaptation	Johanson & Vahlne (1977)
Attraction-selection-attrition	14	Identity & social comparison	Schneider (1987)
Escalation of commitment	14	Cognition & heuristics	Staw (1976)
Garbage can model	14	Cognition & heuristics	Cohen, March & Olsen (1972)
Knowledge creation	14	Learning & adaptation	Nonaka (1994)

Table 32.1. The Theory Atlas: every entry in the book. *(continued)*

Theory	Ch.	Mechanism	Key source
Organizational culture	14	Culture & values	Schein (2010)
Organizational identity	14	Identity & social comparison	Albert & Whetten (1985)
Organizational learning	14	Learning & adaptation	Argyris & Schön (1978), Levitt & March (1988)
Paradox theory	14	Learning & adaptation	Smith & Lewis (2011)
Psychological safety	14	Learning & adaptation	Edmondson (1999)
Sensemaking	14	Cognition & heuristics	Weick (1995)
Social information processing	14	Information	Salancik & Pfeffer (1978)
Affective events theory	15	Emotion & affect	Weiss & Cropanzano (1996)
Conservation of resources theory	15	Resources & capabilities	Hobfoll (1989)
Human capital theory	15	Resources & capabilities	Becker (1964)
Job characteristics model	15	Motivation & goals	Hackman & Oldham (1976)
Job demands-resources theory	15	Resources & capabilities	Demerouti, Bakker, Nachreiner & Schaufeli (2001)
Leader-member exchange	15	Networks & exchange	Graen & Uhl-Bien (1995)
Organizational commitment (three-component model)	15	Motivation & goals	Meyer & Allen (1991)
Organizational justice	15	Identity & social comparison	Colquitt, Conlon, Wesson, Porter & Ng (2001)
Path-goal theory	15	Motivation & goals	House (1971)
Person-environment fit	15	Culture & values	Kristof (1996)
Psychological contract theory	15	Networks & exchange	Rousseau (1989)
Tournament theory	15	Incentives	Lazear & Rosen (1981)
Transformational leadership	15	Culture & values	Bass (1985)
Behavioral finance	16	Cognition & heuristics	Barberis & Thaler (2003)
Capital asset pricing model	16	Incentives	Sharpe (1964), Lintner (1965)
Efficient market hypothesis	16	Information	Fama (1970)
Free cash flow theory	16	Incentives	Jensen (1986)
Limits to arbitrage	16	Incentives	Shleifer & Vishny (1997)
Modigliani-Miller theorems	16	Incentives	Modigliani & Miller (1958, 1963)
Portfolio theory	16	Incentives	Markowitz (1952)
Trade-off and pecking order theories	16	Information	Kraus & Litzenberger (1973), Myers (1984)
Auction theory	17	Information	Vickrey (1961), Milgrom & Weber (1982)
Coase theorem	17	Incentives	Coase (1960)
Collective action	17	Incentives	Olson (1965)
Efficiency wages	17	Incentives	Akerlof (1982), Shapiro & Stiglitz (1984)
Network effects	17	Diffusion & adoption	Katz & Shapiro (1985), Arthur (1989)
New institutional economics	17	Legitimacy & institutions	North (1990), Williamson (2000)

Table 32.1. The Theory Atlas: every entry in the book. *(continued)*

Theory	Ch.	Mechanism	Key source
Property rights theory	17	Power & dependence	Grossman & Hart (1986), Hart & Moore (1990)
Search theory	17	Information	Stigler (1961), Mortensen & Pissarides (1994)
Social preferences	17	Identity & social comparison	Güth, Schmittberger & Schwarze (1982), Fehr & Schmidt (1999)
Two-sided markets	17	Networks & exchange	Rochet & Tirole (2003), Armstrong (2006)
GLOBE	18	Culture & values	House et al. (2004)
Hofstede's cultural dimensions	18	Culture & values	Hofstede (1980)
Power distance theory	18	Culture & values	Mulder (1977)
Schwartz value theory	18	Culture & values	Schwartz (1992)
Self-construal	18	Culture & values	Markus & Kitayama (1991)
Tight-loose cultures	18	Culture & values	Gelfand et al. (2011)
Equity theory	19	Identity & social comparison	Adams (1965)
Expectancy theory	19	Motivation & goals	Vroom (1964)
Goal-gradient hypothesis	19	Motivation & goals	Kivetz, Urminsky & Zheng (2006)
Goal-setting theory	19	Motivation & goals	Locke & Latham (2002)
Losing leads to winning	19	Motivation & goals	Berger & Pope (2011)
Self-determination theory	19	Motivation & goals	Ryan & Deci (2000)
Construal level theory	20	Cognition & heuristics	Trope & Liberman (2010)
Regulatory fit	20	Motivation & goals	Higgins (2000)
Regulatory focus theory	20	Motivation & goals	Higgins (1997)
Self-control and intertemporal choice	20	Motivation & goals	Thaler & Shefrin (1981)
Attribution theory	21	Cognition & heuristics	Heider (1958), Kelley (1973), Weiner (1985)
Balance theory	21	Cognition & heuristics	Heider (1946)
Cognitive dissonance theory	21	Motivation & goals	Festinger (1957)
Reactance theory	21	Motivation & goals	Brehm (1966)
Self-perception theory	21	Cognition & heuristics	Bem (1967, 1972)
Social cognitive theory	21	Learning & adaptation	Bandura (1977, 1986)
Conformity	22	Identity & social comparison	Deutsch & Gerard (1955)
Distraction-conflict theory	22	Cognition & heuristics	Baron (1986)
Optimal distinctiveness theory	22	Identity & social comparison	Brewer (1991)
Social comparison theory	22	Identity & social comparison	Festinger (1954)
Social facilitation	22	Emotion & affect	Zajonc (1965)
Social identity theory	22	Identity & social comparison	Tajfel & Turner (1979)
Social norms (focus theory)	22	Identity & social comparison	Cialdini, Reno & Kallgren (1990)

Table 32.1. The Theory Atlas: every entry in the book. *(continued)*

Theory	Ch.	Mechanism	Key source
Spiral of silence	22	Identity & social comparison	Noelle-Neumann (1974)
Bounded rationality	23	Cognition & heuristics	Simon (1955)
Choice architecture	23	Cognition & heuristics	Thaler & Sunstein (2008)
Discrepancy-attribution hypothesis	23	Cognition & heuristics	Whittlesea & Williams (2001)
Dual-process theory	23	Cognition & heuristics	Evans & Stanovich (2013)
Endowment effect	23	Cognition & heuristics	Kahneman, Knetsch & Thaler (1990)
Feelings-as-information theory	23	Emotion & affect	Schwarz & Clore (1983)
Heuristics and biases	23	Cognition & heuristics	Tversky & Kahneman (1974)
Mental accounting	23	Cognition & heuristics	Thaler (1985)
Processing fluency	23	Cognition & heuristics	Reber, Schwarz & Winkielman (2004)
Cumulative prospect theory	24	Cognition & heuristics	Tversky & Kahneman (1992)
Disposition effect	24	Cognition & heuristics	Shefrin & Statman (1985)
Expected utility theory	24	Cognition & heuristics	von Neumann & Morgenstern (1944)
Loss aversion	24	Emotion & affect	Tversky & Kahneman (1991)
Myopic loss aversion	24	Emotion & affect	Benartzi & Thaler (1995)
Prospect theory	24	Cognition & heuristics	Kahneman & Tversky (1979)
Reference-dependent preferences	24	Motivation & goals	Kőszegi & Rabin (2006)
Status quo bias	24	Cognition & heuristics	Samuelson & Zeckhauser (1988)
Affect circumplex	25	Emotion & affect	Russell (1980)
Appraisal theory	25	Emotion & affect	Smith & Ellsworth (1985)
Broaden-and-build theory	25	Emotion & affect	Fredrickson (2001)
Emotion and virality	25	Emotion & affect	Berger & Milkman (2012)
Hedonic adaptation	25	Emotion & affect	Brickman & Campbell (1971)
Peak-end rule	25	Emotion & affect	Kahneman et al. (1993)
Agenda setting	26	Information	McCombs & Shaw (1972)
Elaboration likelihood model	26	Cognition & heuristics	Petty & Cacioppo (1986)
Framing theory	26	Cognition & heuristics	Entman (1993)
Microblogging (undirected communication)	26	Networks & exchange	Buechel & Berger (2018)
Persuasion knowledge model	26	Cognition & heuristics	Friestad & Wright (1994)
Social transmission (word of mouth)	26	Networks & exchange	Berger (2014)
Uses and gratifications	26	Motivation & goals	Katz, Blumler & Gurevitch (1973)

Table 32.1. The Theory Atlas: every entry in the book. *(continued)*

Theory	Ch.	Mechanism	Key source
Brand personality	27	Identity & social comparison	Aaker (1997)
Consumer culture theory	27	Culture & values	Arnould & Thompson (2005)
Consumer-brand relationships	27	Networks & exchange	Fournier (1998)
Customer-based brand equity	27	Information	Keller (1993)
Expectancy-disconfirmation theory	27	Cognition & heuristics	Oliver (1980)
Hedonic and utilitarian consumption	27	Emotion & affect	Hirschman & Holbrook (1982)
Means-end chain theory	27	Motivation & goals	Gutman (1982)
Psychological ownership	27	Identity & social comparison	Pierce, Kostova & Dirks (2001)
Reference price theory	27	Cognition & heuristics	Winer (1986)
Schema congruity theory	27	Cognition & heuristics	Meyers-Levy & Tybout (1989)
Self-congruity theory	27	Identity & social comparison	Sirgy (1982)
Variety seeking	27	Motivation & goals	McAlister & Pessemier (1982)
Customer equity	28	Resources & capabilities	Gupta, Lehmann & Stuart (2004)
Gaps model of service quality	28	Information	Parasuraman, Zeithaml & Berry (1985)
Hierarchy of effects	28	Cognition & heuristics	Lavidge & Steiner (1961)
Market orientation	28	Information	Kohli & Jaworski (1990), Narver & Slater (1990)
Marketing mix	28	Resources & capabilities	Borden (1964), McCarthy (1960)
Product life cycle	28	Diffusion & adoption	Levitt (1965)
Service-dominant logic	28	Networks & exchange	Vargo & Lusch (2004)
Anchoring effect	29	Cognition & heuristics	Tversky & Kahneman (1974)
Bandwagon, snob, and Veblen effects	29	Identity & social comparison	Leibenstein (1950)
Beauty in averageness	29	Cognition & heuristics	Langlois & Roggman (1990)
Coolidge effect	29	Motivation & goals	Wilson, Kuehn & Beach (1963)
Decoy (attraction) effect	29	Cognition & heuristics	Huber, Payne & Puto (1982)
Glass cliff	29	Identity & social comparison	Ryan & Haslam (2005)
Goldilocks effect	29	Cognition & heuristics	Berlyne (1970)
IKEA effect	29	Motivation & goals	Norton, Mochon & Ariely (2012)
Mere exposure effect	29	Cognition & heuristics	Zajonc (1968)
Pratfall effect	29	Identity & social comparison	Aronson, Willerman & Floyd (1966)

Table 32.1. The Theory Atlas: every entry in the book. *(continued)*

Theory	Ch.	Mechanism	Key source
Warm-glow heuristic	29	Emotion & affect	Monin (2003)
Assortative mating	30	Identity & social comparison	Schwartz & Mare (2005)
Bullwhip effect	30	Information	Lee, Padmanabhan & Whang (1997)
Bystander effect	30	Identity & social comparison	Darley & Latané (1968)
Groupthink	30	Identity & social comparison	Janis (1972)
Homophily	30	Networks & exchange	McPherson, Smith-Lovin & Cook (2001)
Matthew effect	30	Networks & exchange	Merton (1968)
Pack journalism	30	Identity & social comparison	Crouse (1972)
Scent and emotion	30	Emotion & affect	Rolls (2004)
Similarity asymmetry	30	Cognition & heuristics	Tversky (1977)
Social loafing	30	Motivation & goals	Latané, Williams & Harkins (1979)
Social magnet	30	Identity & social comparison	Berger (2016)
Tragedy of the commons	30	Incentives	Hardin (1968), Ostrom (1990)
Wisdom of the crowd	30	Information	Galton (1907)
Combinatorial creativity	31	Learning & adaptation	Uzzi et al. (2013)
Five forces	31	Power & dependence	Porter (1980)
Geneplore model	31	Learning & adaptation	Finke, Ward & Smith (1992)
Information processing model	31	Cognition & heuristics	Atkinson & Shiffrin (1968)
Stimulus-organism-response	31	Emotion & affect	Mehrabian & Russell (1974)

The atlas is also available as a plain CSV file, `data/theory-atlas.csv`, for use in teaching materials, literature reviews, or your own analyses.

32.2 Theories by mechanism

Theories that share a mechanism are natural competitors or complements. Figure 32.1 counts the entries by dominant mechanism and part of the book.

Table 32.2 presents the mechanism families used in the atlas.

Table 32.2. Mechanism families used in the atlas.

Mechanism	What drives behaviour	Entries	Examples
Cognition & heuristics	Limited attention, mental shortcuts, and representation	39	Upper echelons theory, Attention-based view, Effectuation, Escalation of commitment

Table 32.2. Mechanism families used in the atlas. *(continued)*

Mechanism	What drives behaviour	Entries	Examples
Identity & social comparison	Who people think they are and how they compare with others	23	Identity signaling, Attraction-selection-attrition, Organizational identity, Organizational justice
Motivation & goals	Goals, expectations, rewards, and self-regulation	22	Stewardship theory, Theory of planned behavior, Job characteristics model, Organizational commitment (three-component model)
Information	What actors know, can observe, or can infer from others' actions	18	Costly signaling, Countersignaling, Information asymmetry, Screening theory
Incentives	Payoffs, contracts, and the costs of transacting	17	Agency theory, Transaction cost economics, Creating shared value, Competitive dynamics
Learning & adaptation	Search, experimentation, selection, and reconfiguration	17	Absorptive capacity, Dynamic capabilities, Exploration vs. exploitation, Organizational ambidexterity
Networks & exchange	Relationships, reciprocity, and positions in social structure	16	Commitment-trust theory, Social capital, Social exchange theory, Strength of weak ties
Emotion & affect	Feelings, arousal, and their influence on judgment	15	Affective events theory, Social facilitation, Feelings-as-information theory, Loss aversion
Resources & capabilities	Heterogeneous, hard-to-imitate assets and routines	15	Knowledge-based view, Market-based assets, Resource-advantage theory, Resource-based view
Culture & values	Shared values and norms that differ across societies	10	Organizational culture, Person-environment fit, Transformational leadership, GLOBE
Diffusion & adoption	Spread of practices and technologies through social systems	9	Bass diffusion model, Diffusion of innovations, Information systems success model, Task-technology fit
Legitimacy & institutions	Taken-for-granted rules, norms, and social approval	7	Corporate social responsibility pyramid, Institutional theory, Legitimacy, Liability of newness
Power & dependence	Control over resources that others need	6	Resource dependence theory, Stakeholder theory, Network exchange theory, Structure-conduct-performance
Explanation	Meta-theoretical tools for building and testing theory	4	Levels of analysis, Theory building blocks, Mediation and moderation, Replication crisis

32.3 Glossary

Table 32.3 presents the glossary of recurring terms.

Table 32.3. Glossary of recurring terms.

Term	Definition
Adverse selection	A hidden-information problem in which the parties most eager to transact are those the other side would least want to transact with.
Boundary condition	A limit on the who, where, and when to which a theory's predictions apply.

Table 32.3. Glossary of recurring terms. *(continued)*

Term	Definition
Construct	An abstract concept (e.g., trust, legitimacy) that a theory relates to other concepts. Measured imperfectly by variables.
Countersignaling	Deliberately not sending a signal because other information already separates the sender from low types.
Descriptive norm	Perception of what most people actually do.
Dynamic capability	A firm's ability to integrate, build, and reconfigure competences to address changing environments.
Equilibrium (pooling)	An outcome in which different types take the same action, so the action reveals nothing.
Equilibrium (separating)	An outcome in which different types take different actions, so the action reveals type.
HARKing	Hypothesizing after the results are known: presenting post hoc hypotheses as if they were a priori.
Injunctive norm	Perception of what most people approve or disapprove of.
Isomorphism	The tendency of organizations in a field to become similar in structure and practice.
Level of analysis	The unit (individual, dyad, group, firm, market) at which a construct or theory is defined.
Loss aversion	The tendency for losses to loom larger than gains of the same size.
Mechanism	The process through which a cause produces an effect. The 'why' of a theory.
Mediator	A variable that transmits the effect of an independent variable to a dependent variable.
Moderator	A variable that changes the strength or direction of a relationship.
Moral hazard	A hidden-action problem in which one party changes behaviour after a contract because the other cannot observe it.
Nomological network	The web of theoretical relationships in which a construct is embedded, used to establish construct validity.
Proposition	A statement of relationship between constructs. A hypothesis is its testable counterpart in terms of variables.
Reference point	The benchmark against which outcomes are coded as gains or losses.
Regulatory fit	The increased value people experience when the manner of goal pursuit matches their regulatory orientation.
Signal	An observable action chosen by an informed party that conveys information about an unobservable characteristic.
Social capital	Resources embedded in, and accessible through, social relationships.
Structural hole	A gap between two contacts who are not connected to each other, which a broker can span.
VRIO	Valuable, rare, inimitable, organized: the conditions for a resource to yield sustained competitive advantage.

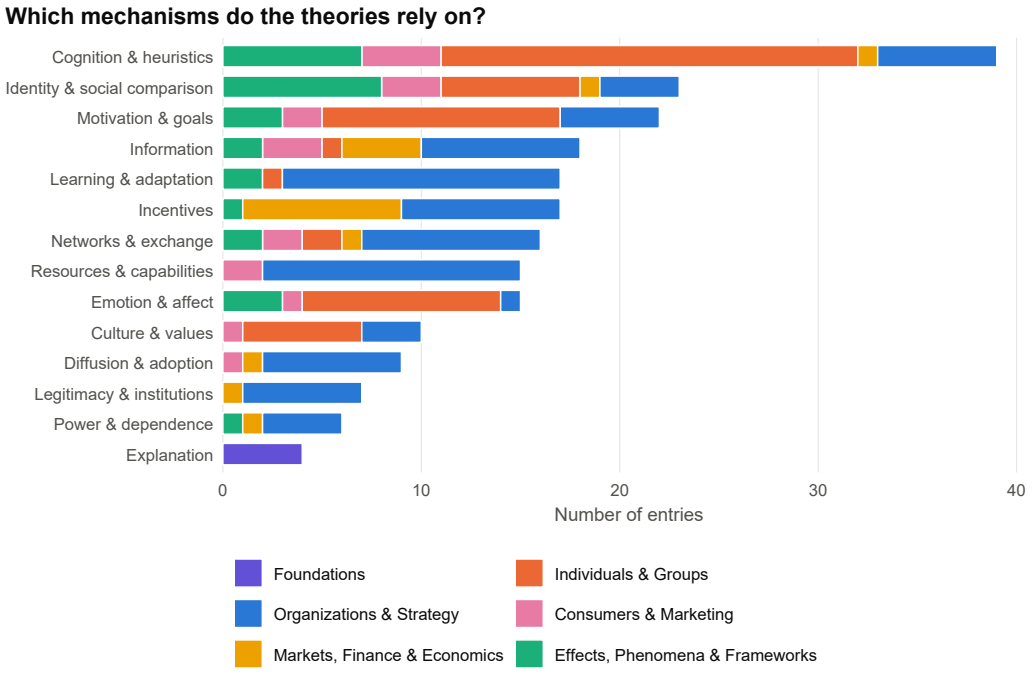


Figure 32.1. Entries in the book by dominant mechanism, stacked by part of the book.

References

Aaker, David A. 1991. *Managing Brand Equity: Capitalizing on the Value of a Brand Name*. New York: Free Press.

Aaker, David A., and Kevin Lane Keller. 1990. "Consumer Evaluations of Brand Extensions." *Journal of Marketing* 54 (1): 27-41.

Aaker, Jennifer L. 1997. "Dimensions of Brand Personality." *Journal of Marketing Research* 34 (3): 347-56. https://doi.org/10.2307/3151897.

Aaker, Jennifer L. 1999. "The Malleable Self: The Role of Self-Expression in Persuasion." *Journal of Marketing Research* 36 (1): 45-57. https://doi.org/10.2307/3151914.

Aaker, Jennifer L., and Angela Y. Lee. 2001. ""I" Seek Pleasures and "We" Avoid Pains: The Role of Self-Regulatory Goals in Information Processing and Persuasion." *Journal of Consumer Research* 28 (1): 33-49. https://doi.org/10.1086/321946.

Aaker, Jennifer, Susan Fournier, and S. Adam Brasel. 2004. "When Good Brands Do Bad." *Journal of Consumer Research* 31 (1): 1-16. https://doi.org/10.1086/383419.

Abdellaoui, Mohammed, Han Bleichrodt, and Corina Paraschiv. 2007. "Loss Aversion Under Prospect Theory: A Parameter-Free Measurement." *Management Science* 53 (10): 1659-74. https://doi.org/10.1287/mnsc.1070.0711.

Abeler, Johannes, Armin Falk, Lorenz Goette, and David Huffman. 2011. "Reference Points and Effort Provision." *American Economic Review* 101 (2): 470-92. https://doi.org/10.1257/aer.101.2.470.

Abernathy, William J., and James M. Utterback. 1978. "Patterns of Industrial Innovation." *Technology Review* 80 (7): 40-47.

Abrahamson, Eric. 1991. "Managerial Fads and Fashions: The Diffusion and Rejection of Innovations." *Academy of Management Review* 16 (3): 586-612. https://doi.org/10.5465/amr.1991.4279484.

Acemoglu, Daron, Simon Johnson, and James A. Robinson. 2001. "The Colonial Origins of Comparative Development: An Empirical Investigation." *American Economic Review* 91 (5): 1369-1401. https://doi.org/10.1257/aer.91.5.1369.

Acemoglu, Daron, and Jörn-Steffen Pischke. 1999. "Beyond Becker: Training in Imperfect Labour Markets." *The Economic Journal* 109 (453): F112-42. https://doi.org/10.1111/1468-0297.00405.

Adair, Wendi L., and Jeanne M. Brett. 2005. "The Negotiation Dance: Time, Culture, and Behavioral Sequences in Negotiation." *Organization Science* 16 (1): 33-51.

Adams, J. Stacy. 1963. "Toward an Understanding of Inequity." *Journal of Abnormal and Social Psychology* 67 (5): 422-36. https://doi.org/10.1037/h0040968.

Adams, J. Stacy. 1965. "Inequity in Social Exchange." *Advances in Experimental Social Psychology* 2: 267-99. https://doi.org/10.1016/S0065-2601(08)60108-2.

Adams, Susan M., Atul Gupta, and John D. Leeth. 2009. "Are Female Executives over-Represented in Precarious Leadership Positions?" *British Journal of Management* 20 (1): 1-12.

Adler, Paul S., and Seok-Woo Kwon. 2002. "Social Capital: Prospects for a New Concept." *Academy of Management Review* 27 (1): 17-40. https://doi.org/10.5465/amr.2002.5922314.

Adner, Ron, and Daniel A. Levinthal. 2004. "What Is Not a Real Option: Considering Boundaries for the Application of Real Options to Business Strategy." *The Academy of Management Review* 29 (1): 74-85. https://doi.org/10.2307/20159010.

Aggarwal, Pankaj. 2004. "The Effects of Brand Relationship Norms on Consumer Attitudes and Behavior." *Journal of Consumer Research* 31 (1): 87-101. https://doi.org/10.1086/383426.

Aghion, Philippe, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. 2005. "Competition and Innovation: An Inverted-U Relationship." *The Quarterly Journal of Economics* 120 (2): 701-28. https://doi.org/10.1093/qje/120.2.701.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica* 60 (2): 323-51. https://doi.org/10.2307/2951599.

Ahuja, Gautam. 2000. "Collaboration Networks, Structural Holes, and Innovation: A Longitudinal Study." *Administrative Science Quarterly* 45 (3): 425-55. https://doi.org/10.2307/2667105.

Aiello, John R., and Kathryn J. Kolb. 1995. "Electronic Performance Monitoring and Social Context: Impact on Productivity and Stress." *Journal of Applied Psychology* 80 (3): 339-53. https://doi.org/10.1037/0021-9010.80.3.339.

Aiken, Leona S., and Stephen G. West. 1991. *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park, CA: Sage.

Ailawadi, Kusum L., Donald R. Lehmann, and Scott A. Neslin. 2003. "Revenue Premium as an Outcome Measure of Brand Equity." *Journal of Marketing* 67 (4): 1-17. https://doi.org/10.1509/jmkg.67.4.1.18688.

Ainslie, George. 1975. "Specious Reward: A Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin* 82 (4): 463-96. https://doi.org/10.1037/h0076860.

Ajzen, Icek. 1991. "The Theory of Planned Behavior." *Organizational Behavior and Human Decision Processes* 50 (2): 179-211. https://doi.org/10.1016/0749-5978(91)90020-T.

Ajzen, Icek, and Martin Fishbein. 1980. *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice-Hall.

Akerlof, George A. 1970. "The Market for "Lemons": Quality Uncertainty and the Market Mechanism." *The Quarterly Journal of Economics* 84 (3): 488-500. https://doi.org/10.2307/1879431.

Akerlof, George A. 1982. "Labor Contracts as Partial Gift Exchange." *The Quarterly Journal of Economics* 97 (4): 543-69. https://doi.org/10.2307/1885099.

Aknin, Lara B., Christopher P. Barrington-Leigh, Elizabeth W. Dunn, John F. Helliwell, Justine Burns, Robert Biswas-Diener, Imelda Kemeza, Paul Nyende, Claire E. Ashton-James, and Michael I. Norton. 2013. "Prosocial Spending and Well-Being: Cross-Cultural Evidence for a Psychological Universal." *Journal of Personality and Social Psychology* 104 (4): 635-52.

Alba, Joseph W., and Elanor F. Williams. 2013. "Pleasure Principles: A Review of Research on Hedonic Consumption." *Journal of Consumer Psychology* 23 (1): 2-18. https://doi.org/10.1016/j.jcps.2012.07.003.

Albert, Stuart, and David A. Whetten. 1985. "Organizational Identity." In *Research in Organizational Behavior*, edited by L. L. Cummings and Barry M. Staw, 7:263-95. Greenwich, CT: JAI Press.

Albouy, David Y. 2012. "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment." *American Economic Review* 102 (6): 3059-76. https://doi.org/10.1257/aer.102.6.3059.

Aldrich, Howard E., and C. Marlene Fiol. 1994. "Fools Rush in? The Institutional Context of Industry Creation." *Academy of Management Review* 19 (4): 645-70.

Allaire, JJ, Yihui Xie, Christophe Dervieux, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, et al. 2025. *Rmarkdown: Dynamic Documents for r*. https://github.com/rstudio/rmarkdown.

Allais, Maurice. 1953. "Le Comportement de l'homme Rationnel Devant Le Risque: Critique Des Postulats Et Axiomes de l'École Américaine." *Econometrica* 21 (4): 503-46. https://doi.org/10.2307/1907921.

Allcott, Hunt. 2011. "Social Norms and Energy Conservation." *Journal of Public Economics* 95 (9-10): 1082-95. https://doi.org/10.1016/j.jpubeco.2011.03.003.

Allcott, Hunt. 2015. "Site Selection Bias in Program Evaluation." *The Quarterly Journal of Economics* 130 (3): 1117-65. https://doi.org/10.1093/qje/qjv015.

Allcott, Hunt, and Todd Rogers. 2014. "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." *American Economic Review* 104 (10): 3003-37. https://doi.org/10.1257/aer.104.10.3003.

Allen, Eric J., Patricia M. Dechow, Devin G. Pope, and George Wu. 2017. "Reference-Dependent Preferences: Evidence from Marathon Runners." *Management Science* 63 (6): 1657-72.

Allen, Natalie J., and John P. Meyer. 1990. "The Measurement and Antecedents of Affective, Continuance and Normative Commitment to the Organization." *Journal of Occupational Psychology* 63 (1): 1-18. https://doi.org/10.1111/j.2044-8325.1990.tb00506.x.

Allport, Floyd H. 1920. "The Influence of the Group Upon Association and Thought." *Journal of Experimental Psychology* 3 (3): 159-82. https://doi.org/10.1037/h0067891.

Almende B.V. and Contributors, and Benoit Thieurmel. 2025. *visNetwork: Network Visualization Using Vis.js Library*. https://datastorm-open.github.io/visNetwork/.

Alter, Adam L., and Daniel M. Oppenheimer. 2006. "Predicting Short-Term Stock Fluctuations by Using Processing Fluency." *Proceedings of the National Academy of Sciences* 103 (24): 9369-72. https://doi.org/10.1073/pnas.0601071103.

Alter, Adam L., and Daniel M. Oppenheimer. 2009. "Uniting the Tribes of Fluency to Form a Metacognitive Nation." *Personality and Social Psychology Review* 13 (3): 219-35. https://doi.org/10.1177/1088868309341564.

Alvarez, Sharon A., and Jay B. Barney. 2007. "Discovery and Creation: Alternative Theories of Entrepreneurial Action." *Strategic Entrepreneurship Journal* 1 (1-2): 11-26. https://doi.org/10.1002/sej.4.

Amabile, Teresa M. 1996. *Creativity in Context*. Boulder, CO: Westview Press.

Amit, Raphael, and Paul J. H. Schoemaker. 1993. "Strategic Assets and Organizational Rent." *Strategic Management Journal* 14 (1): 33-46. https://doi.org/10.1002/smj.4250140105.

Anderson, Erin, and Hubert Gatignon. 1986. "Modes of Foreign Entry: A Transaction Cost Analysis and Propositions." *Journal of International Business Studies* 17 (3): 1-26. https://doi.org/10.1057/palgrave.jibs.8490432.

Anderson, Eugene W., and Mary W. Sullivan. 1993. "The Antecedents and Consequences of Customer Satisfaction for Firms." *Marketing Science* 12 (2): 125-43. https://doi.org/10.1287/mksc.12.2.125.

Anderson, Philip, and Michael L. Tushman. 1990. "Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change." *Administrative Science Quarterly* 35 (4): 604-33. https://doi.org/10.2307/2393511.

Anderson, Rolph E. 1973. "Consumer Dissatisfaction: The Effect of Disconfirmed Expectancy on Perceived Product Performance." *Journal of Marketing Research* 10 (1): 38-44. https://doi.org/10.1177/002224377301000106.

Andreoni, James. 1990. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal* 100 (401): 464-77. https://doi.org/10.2307/2234133.

Antonakis, John, Samuel Bendahan, Philippe Jacquart, and Rafael Lalive. 2010. "On Making Causal Claims: A Review and Recommendations." *The Leadership Quarterly* 21 (6): 1086-1120. https://doi.org/10.1016/j.leaqua.2010.10.010.

Aral, Sinan, Lev Muchnik, and Arun Sundararajan. 2009. "Distinguishing Influence-Based Contagion from Homophily-Driven Diffusion in Dynamic Networks." *Proceedings of the National Academy of Sciences* 106 (51): 21544-49. https://doi.org/10.1073/pnas.0908800106.

Aral, Sinan, and Marshall Van Alstyne. 2011. "The Diversity-Bandwidth Trade-Off." *American Journal of Sociology* 117 (1): 90-171. https://doi.org/10.1086/661238.

Arend, Richard J., Hessamoddin Sarooghi, and Andrew Burkemper. 2015. "Effectuation as Ineffectual? Applying the 3E Theory-Assessment Framework to a Proposed New Theory of Entrepreneurship." *Academy of Management Review* 40 (4): 630-51. https://doi.org/10.5465/amr.2014.0455.

Argote, Linda, and Dennis Epple. 1990. "Learning Curves in Manufacturing." *Science* 247 (4945): 920-24. https://doi.org/10.1126/science.247.4945.920.

Argote, Linda, and Ella Miron-Spektor. 2011. "Organizational Learning: From Experience to Knowledge." *Organization Science* 22 (5): 1123-37. https://doi.org/10.1287/orsc.1100.0621.

Argyris, Chris. 1976. "Single-Loop and Double-Loop Models in Research on Decision Making." *Administrative Science Quarterly* 21 (3): 363-75. https://doi.org/10.2307/2391848.

Argyris, Chris, and Donald A. Schön. 1978. *Organizational Learning: A Theory of Action Perspective*. Reading, MA: Addison-Wesley.

Ariely, Dan. 2008. *Predictably Irrational: The Hidden Forces That Shape Our Decisions*. New York: HarperCollins.

Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "'Coherent Arbitrariness': Stable Demand Curves Without Stable Preferences." *The Quarterly Journal of Economics* 118 (1): 73-106. https://doi.org/10.1162/00335530360535153.

Ariely, Dan, and Klaus Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." *Psychological Science* 13 (3): 219-24. https://doi.org/10.1111/1467-9280.00441.

Arkes, Hal R., and Peter Ayton. 1999. "The Sunk Cost and Concorde Effects: Are Humans Less Rational Than Lower Animals?" *Psychological Bulletin* 125 (5): 591-600. https://doi.org/10.1037/0033-2909.125.5.591.

Arkes, Hal R., and Catherine Blumer. 1985. "The Psychology of Sunk Cost." *Organizational Behavior and Human Decision Processes* 35 (1): 124-40. https://doi.org/10.1016/0749-5978(85)90049-4.

Armitage, Christopher J., and Mark Conner. 2001. "Efficacy of the Theory of Planned Behaviour: A Meta-Analytic Review." *British Journal of Social Psychology* 40 (4): 471-99. https://doi.org/10.1348/014466601164939.

Armstrong, Mark. 2006. "Competition in Two-Sided Markets." *The RAND Journal of Economics* 37 (3): 668-91. https://doi.org/10.1111/j.1756-2171.2006.tb00037.x.

Arnould, Eric J., and Craig J. Thompson. 2005. "Consumer Culture Theory (CCT): Twenty Years of Research." *Journal of Consumer Research* 31 (4): 868-82. https://doi.org/10.1086/426626.

Aronson, Elliot, and Judson Mills. 1959. "The Effect of Severity of Initiation on Liking for a Group." *Journal of Abnormal and Social Psychology* 59 (2): 177-81. https://doi.org/10.1037/h0047195.

Aronson, Elliot, Ben Willerman, and Joanne Floyd. 1966. "The Effect of a Pratfall on Increasing Interpersonal Attractiveness." *Psychonomic Science* 4 (6): 227-28. https://doi.org/10.3758/BF03342263.

Arthur, W. Brian. 1989. "Competing Technologies, Increasing Returns, and Lock-in by Historical Events." *The Economic Journal* 99 (394): 116-31. https://doi.org/10.2307/2234208.

Asch, Solomon E. 1956. "Studies of Independence and Conformity: I. A Minority of One Against a Unanimous Majority." *Psychological Monographs: General and Applied* 70 (9): 1-70. https://doi.org/10.1037/h0093718.

Ashforth, Blake E., and Fred Mael. 1989. "Social Identity Theory and the Organization." *Academy of Management Review* 14 (1): 20-39. https://doi.org/10.2307/258189.

Ashraf, Nava, Dean Karlan, and Wesley Yin. 2006. "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines." *The Quarterly Journal of Economics* 121 (2): 635-72. https://doi.org/10.1162/qjec.2006.121.2.635.

Askegaard, Søren, and Jeppe Trolle Linnet. 2011. "Towards an Epistemology of Consumer Culture Theory: Phenomenology and the Context of Context." *Marketing Theory* 11 (4): 381-404. https://doi.org/10.1177/1470593111418796.

Atkinson, R. C., and R. M. Shiffrin. 1968. "Human Memory: A Proposed System and Its Control Processes." In *Psychology of Learning and Motivation*, 89-195. Elsevier. https://doi.org/10.1016/s0079-7421(08)60422-3.

Attewell, Paul. 1992. "Technology Diffusion and Organizational Learning: The Case of Business Computing." *Organization Science* 3 (1): 1-19. https://doi.org/10.1287/orsc.3.1.1.

Audia, Pino G., Edwin A. Locke, and Ken G. Smith. 2000. "The Paradox of Success: An Archival and a Laboratory Study of Strategic Persistence Following Radical Environmental Change." *Academy of Management Journal* 43 (5): 837-53. https://doi.org/10.2307/1556413.

Aupperle, Kenneth E., Archie B. Carroll, and John D. Hatfield. 1985. "An Empirical Examination of the Relationship Between Corporate Social Responsibility and Profitability." *Academy of Management Journal* 28 (2): 446-63. https://doi.org/10.2307/256210.

Austin, Jon R., Judy A. Siguaw, and Anna S. Mattila. 2003. "A Re-Examination of the Generalizability of the Aaker Brand Personality Measurement Framework." *Journal of Strategic Marketing* 11 (2): 77-92. https://doi.org/10.1080/0965254032000104469.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Azoulay, Audrey, and Jean-Noël Kapferer. 2003. "Do Brand Personality Scales Really Measure Brand Personality?" *Journal of Brand Management* 11 (2): 143-55. https://doi.org/10.1057/palgrave.bm.2540162.

Azoulay, Pierre, Toby Stuart, and Yanbo Wang. 2014. "Matthew: Effect or Fable?" *Management Science* 60 (1): 92-109.

Babin, Barry J., William R. Darden, and Mitch Griffin. 1994. "Work and/or Fun: Measuring Hedonic and Utilitarian Shopping Value." *Journal of Consumer Research* 20 (4): 644-56. https://doi.org/10.1086/209376.

Bacharach, Samuel B. 1989. "Organizational Theories: Some Criteria for Evaluation." *Academy of Management Review* 14 (4): 496-515. https://doi.org/10.5465/amr.1989.4308374.

Baddeley, Alan. 2000. "The Episodic Buffer: A New Component of Working Memory?" *Trends in Cognitive Sciences* 4 (11): 417-23.

Baddeley, Alan D., and Graham Hitch. 1974. "Working Memory." In *The Psychology of Learning and Motivation*, edited by Gordon H. Bower, 8:47-89. New York: Academic Press.

Bagozzi, Richard P. 2007. "The Legacy of the Technology Acceptance Model and a Proposal for a Paradigm Shift." *Journal of the Association for Information Systems* 8 (4): 244-54. https://doi.org/10.17705/1jais.00122.

Bagwell, Laurie Simon, and B. Douglas Bernheim. 1996. "Veblen Effects in a Theory of Conspicuous Consumption." *The American Economic Review* 86 (3): 349-73.

Bain, Joe S. 1951. "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-1940." *The Quarterly Journal of Economics* 65 (3): 293-324. https://doi.org/10.2307/1882217.

Bain, Joe S. 1956. *Barriers to New Competition: Their Character and Consequences in Manufacturing Industries*. Cambridge, MA: Harvard University Press.

Baines, T. S., H. W. Lightfoot, O. Benedettini, and J. M. Kay. 2009. "The Servitization of Manufacturing: A Review of Literature and Reflection on Future Challenges." *Journal of Manufacturing Technology Management* 20 (5): 547-67. https://doi.org/10.1108/17410380910960984.

Baker, George P., and Thomas N. Hubbard. 2003. "Make Versus Buy in Trucking: Asset Ownership, Job Design, and Information." *American Economic Review* 93 (3): 551-72.

Baker, Malcolm, and Jeffrey Wurgler. 2002. "Market Timing and Capital Structure." *The Journal of Finance* 57 (1): 1-32. https://doi.org/10.1111/1540-6261.00414.

Baker, Malcolm, and Jeffrey Wurgler. 2006. "Investor Sentiment and the Cross-Section of Stock Returns." *The Journal of Finance* 61 (4): 1645-80. https://doi.org/10.1111/j.1540-6261.2006.00885.x.

Baker, Ted, and Reed E. Nelson. 2005. "Creating Something from Nothing: Resource Construction Through Entrepreneurial Bricolage." *Administrative Science Quarterly* 50 (3): 329-66. https://doi.org/10.2189/asqu.2005.50.3.329.

Bakker, Arnold B., and Evangelia Demerouti. 2007. "The Job Demands-Resources Model: State of the Art." *Journal of Managerial Psychology* 22 (3): 309-28. https://doi.org/10.1108/02683940710733115.

Bakker, Arnold B., and Evangelia Demerouti. 2017. "Job Demands-Resources Theory: Taking Stock and Looking Forward." *Journal of Occupational Health Psychology* 22 (3): 273-85. https://doi.org/10.1037/ocp0000056.

Bandura, Albert. 1977. "Self-Efficacy: Toward a Unifying Theory of Behavioral Change." *Psychological Review* 84 (2): 191-215.

Bandura, Albert. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, Albert. 1997. *Self-Efficacy: The Exercise of Control*. New York: W. H. Freeman.

Bandura, Albert, and Edwin A. Locke. 2003. "Negative Self-Efficacy and Goal Effects Revisited." *Journal of Applied Psychology* 88 (1): 87-99. https://doi.org/10.1037/0021-9010.88.1.87.

Bandura, Albert, Dorothea Ross, and Sheila A. Ross. 1961. "Transmission of Aggression Through Imitation of Aggressive Models." *Journal of Abnormal and Social Psychology* 63 (3): 575-82. https://doi.org/10.1037/h0045925.

Banerjee, Abhijit V. 1992. "A Simple Model of Herd Behavior." *The Quarterly Journal of Economics* 107 (3): 797-817.

Barabási, Albert-László, and Réka Albert. 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439): 509-12. https://doi.org/10.1126/science.286.5439.509.

Barasch, Alixandra, and Jonah Berger. 2014. "Broadcasting and Narrowcasting: How Audience Size Affects What People Share." *Journal of Marketing Research* 51 (3): 286-99. https://doi.org/10.1509/jmr.13.0238.

Barber, Brad M., and Terrance Odean. 2000. "Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors." *The Journal of Finance* 55 (2): 773-806. https://doi.org/10.1111/0022-1082.00226.

Barber, Brad M., and Terrance Odean. 2001. "Boys Will Be Boys: Gender, Overconfidence, and Common Stock Investment." *The Quarterly Journal of Economics* 116 (1): 261-92. https://doi.org/10.1162/003355301556400.

Barberis, Nicholas C. 2013. "Thirty Years of Prospect Theory in Economics: A Review and Assessment." *Journal of Economic Perspectives* 27 (1): 173-96. https://doi.org/10.1257/jep.27.1.173.

Barberis, Nicholas, and Ming Huang. 2008. "Stocks as Lotteries: The Implications of Probability Weighting for Security Prices." *American Economic Review* 98 (5): 2066-2100. https://doi.org/

10.1257/aer.98.5.2066.
Barberis, Nicholas, Ming Huang, and Tano Santos. 2001. "Prospect Theory and Asset Prices." *The Quarterly Journal of Economics* 116 (1): 1-53. https://doi.org/10.1162/003355301556310.
Barberis, Nicholas, Ming Huang, and Richard H. Thaler. 2006. "Individual Preferences, Monetary Gambles, and Stock Market Participation: A Case for Narrow Framing." *American Economic Review* 96 (4): 1069-90. https://doi.org/10.1257/aer.96.4.1069.
Barberis, Nicholas, Abhiroop Mukherjee, and Baolian Wang. 2016. "Prospect Theory and Stock Returns: An Empirical Test." *Review of Financial Studies* 29 (11): 3068-3107. https://doi.org/10.1093/rfs/hhw049.
Barberis, Nicholas, and Richard Thaler. 2003. "A Survey of Behavioral Finance." In *Handbook of the Economics of Finance*, edited by George M. Constantinides, Milton Harris, and René M. Stulz, 1B:1053-1128. Amsterdam: Elsevier. https://doi.org/10.1016/S1574-0102(03)01027-6.
Barberis, Nicholas, and Wei Xiong. 2009. "What Drives the Disposition Effect? An Analysis of a Long-Standing Preference-Based Explanation." *The Journal of Finance* 64 (2): 751-84. https://doi.org/10.1111/j.1540-6261.2009.01448.x.
Barling, Julian, Tom Weber, and E. Kevin Kelloway. 1996. "Effects of Transformational Leadership Training on Attitudinal and Financial Outcomes: A Field Experiment." *Journal of Applied Psychology* 81 (6): 827-32. https://doi.org/10.1037/0021-9010.81.6.827.
Barnett, Michael L., and Robert M. Salomon. 2006. "Beyond Dichotomy: The Curvilinear Relationship Between Social Responsibility and Financial Performance." *Strategic Management Journal* 27 (11): 1101-22. https://doi.org/10.1002/smj.557.
Barney, J., and & A. Arikan. 2001. "The Resource-Based View: Origins and Implications." In *Handbook of Strategic Management*.
Barney, Jay. 1991. "Firm Resources and Sustained Competitive Advantage." *Journal of Management* 17 (1): 99-120. https://doi.org/10.1177/014920639101700108.
Barney, Jay B. 1986. "Strategic Factor Markets: Expectations, Luck, and Business Strategy." *Management Science* 32 (10): 1231-41. https://doi.org/10.1287/mnsc.32.10.1231.
Barney, Jay B. 2001. "Is the Resource-Based 'View' a Useful Perspective for Strategic Management Research? Yes." *Academy of Management Review* 26 (1): 41-56.
Barney, Jay B., and Asli M. Arikan. 2017. "The Resource-Based View." In *The Blackwell Handbook of Strategic Management*, 123-82. Blackwell Publishing Ltd. https://doi.org/10.1111/b.9780631218616.2006.00006.x.
Barney, Jay B., and William S. Hesterly. 2012. *Strategic Management and Competitive Advantage Concepts*.
Barney, Jay B., and Robert E. Hoskisson. 1990. "Strategic Groups: Untested Assertions and Research Proposals." *Managerial and Decision Economics* 11 (3): 187-98. https://doi.org/10.1002/mde.4090110306.
Barney, Jay B., David J. Ketchen, and Mike Wright. 2011. "The Future of Resource-Based Theory." Edited by Jay B. Barney, David J. Ketchen, and Mike Wright. *Journal of Management* 37 (5): 1299-1315. https://doi.org/10.1177/0149206310391805.
Baron, Reuben M., and David A. Kenny. 1986. "The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51 (6): 1173-82. https://doi.org/10.1037/0022-3514.51.6.1173.
Baron, Robert S. 1986. "Distraction-Conflict Theory: Progress and Problems." *Advances in Experimental Social Psychology* 19: 1-40. https://doi.org/10.1016/S0065-2601(08)60211-7.
Barrett, Lisa Feldman, and James A. Russell. 1999. "The Structure of Current Affect: Controversies and Emerging Consensus." *Current Directions in Psychological Science* 8 (1): 10-14.

Barry, Thomas E., and Daniel J. Howard. 1990. "A Review and Critique of the Hierarchy of Effects in Advertising." *International Journal of Advertising* 9 (2): 121-35. https://doi.org/10.1080/02650487.1990.11107138.

Barsade, Sigal G. 2002. "The Ripple Effect: Emotional Contagion and Its Influence on Group Behavior." *Administrative Science Quarterly* 47 (4): 644-75.

Bass, Bernard M. 1985. *Leadership and Performance Beyond Expectations*. New York: Free Press.

Bass, Frank M. 1969. "A New Product Growth for Model Consumer Durables." *Management Science* 15 (5): 215-27. https://doi.org/10.1287/mnsc.15.5.215.

Bass, Frank M., Trichy V. Krishnan, and Dipak C. Jain. 1994. "Why the Bass Model Fits Without Decision Variables." *Marketing Science* 13 (3): 203-23. https://doi.org/10.1287/mksc.13.3.203.

Batra, Rajeev, Aaron Ahuvia, and Richard P. Bagozzi. 2012. "Brand Love." *Journal of Marketing* 76 (2): 1-16. https://doi.org/10.1509/jm.09.0339.

Baum, Joel A. C., and Helaine J. Korn. 1996. "Competitive Dynamics of Interfirm Rivalry." *Academy of Management Journal* 39 (2): 255-91. https://doi.org/10.2307/256781.

Baumeister, Roy F., Ellen Bratslavsky, Mark Muraven, and Dianne M. Tice. 1998. "Ego Depletion: Is the Active Self a Limited Resource?" *Journal of Personality and Social Psychology* 74 (5): 1252-65. https://doi.org/10.1037/0022-3514.74.5.1252.

Beaman, Arthur L., C. Maureen Cole, Marilyn Preston, Bonnel Klentz, and Nancy Mehrkens Steblay. 1983. "Fifteen Years of Foot-in-the-Door Research: A Meta-Analysis." *Personality and Social Psychology Bulletin* 9 (2): 181-96. https://doi.org/10.1177/0146167283092002.

Bebchuk, Lucian A., and Roberto Tallarita. 2020. "The Illusory Promise of Stakeholder Governance." *Cornell Law Review* 106 (1): 91-178.

Bebchuk, Lucian, and Jesse Fried. 2004. *Pay Without Performance: The Unfulfilled Promise of Executive Compensation*. Cambridge, MA: Harvard University Press.

Becker, Brian E., and Mark A. Huselid. 1992. "The Incentive Effects of Tournament Compensation Systems." *Administrative Science Quarterly* 37 (2): 336-50. https://doi.org/10.2307/2393228.

Becker, Gary S. 1962. "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy* 70 (5, Part 2): 9-49. https://doi.org/10.1086/258724.

Becker, Gary S. 1964. *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: National Bureau of Economic Research.

Belk, Russell W. 1988. "Possessions and the Extended Self." *Journal of Consumer Research* 15 (2): 139-68. https://doi.org/10.1086/209154.

Belk, Russell W. 2013. "Extended Self in a Digital World." *Journal of Consumer Research* 40 (3): 477-500. https://doi.org/10.1086/671052.

Belk, Russell W., Melanie Wallendorf, and Jr. Sherry John F. 1989. "The Sacred and the Profane in Consumer Behavior: Theodicy on the Odyssey." *Journal of Consumer Research* 16 (1): 1-38. https://doi.org/10.1086/209191.

Bell, David R., and James M. Lattin. 2000. "Looking for Loss Aversion in Scanner Panel Data: The Confounding Effect of Price Response Heterogeneity." *Marketing Science* 19 (2): 185-200. https://doi.org/10.1287/mksc.19.2.185.11802.

Bem, Daryl J. 1967. "Self-Perception: An Alternative Interpretation of Cognitive Dissonance Phenomena." *Psychological Review* 74 (3): 183-200. https://doi.org/10.1037/h0024835.

Bem, Daryl J. 1972. "Self-Perception Theory." In *Advances in Experimental Social Psychology*, edited by Leonard Berkowitz, 6:1-62. New York: Academic Press. https://doi.org/10.1016/S0065-2601(08)60024-6.

Benartzi, Shlomo, and Richard H. Thaler. 1995. "Myopic Loss Aversion and the Equity Premium Puzzle." *The Quarterly Journal of Economics* 110 (1): 73-92. https://doi.org/10.2307/2118511.

Benartzi, Shlomo, and Richard H. Thaler. 1999. "Risk Aversion or Myopia? Choices in Repeated Gambles and Retirement Investments." *Management Science* 45 (3): 364-81. https://doi.org/10.1287/mnsc.45.3.364.

Benbasat, Izak, and Henri Barki. 2007. "Quo Vadis, TAM?" *Journal of the Association for Information Systems* 8 (4): 211-18. https://doi.org/10.17705/1jais.00126.

Ben-David, Itzhak, and David Hirshleifer. 2012. "Are Investors Really Reluctant to Realize Their Losses? Trading Responses to Past Returns and the Disposition Effect." *Review of Financial Studies* 25 (8): 2485-2532. https://doi.org/10.1093/rfs/hhs077.

Bendor, Jonathan, Terry M. Moe, and Kenneth W. Shotts. 2001. "Recycling the Garbage Can: An Assessment of the Research Program." *American Political Science Review* 95 (1): 169-90. https://doi.org/10.1017/S0003055401000041.

Benner, Mary J., and Michael L. Tushman. 2003. "Exploitation, Exploration, and Process Management: The Productivity Dilemma Revisited." *Academy of Management Review* 28 (2): 238-56.

Berger, Jonah. 2011. "Arousal Increases Social Transmission of Information." *Psychological Science* 22 (7): 891-93. https://doi.org/10.1177/0956797611413294.

Berger, Jonah. 2013. *Contagious: Why Things Catch on*.

Berger, Jonah. 2014. "Word of Mouth and Interpersonal Communication: A Review and Directions for Future Research." *Journal of Consumer Psychology* 24 (4): 586-607. https://doi.org/10.1016/j.jcps.2014.05.002.

Berger, Jonah. 2016. *Invisible Influence: The Hidden Forces That Shape Behavior*.

Berger, Jonah, and Chip Heath. 2007. "Where Consumers Diverge from Others: Identity Signaling and Product Domains." *Journal of Consumer Research* 34 (2): 121-34. https://doi.org/10.1086/519142.

Berger, Jonah, and Chip Heath. 2008. "Who Drives Divergence? Identity Signaling, Outgroup Dissimilarity, and the Abandonment of Cultural Tastes." *Journal of Personality and Social Psychology* 95 (3): 593-607. https://doi.org/10.1037/0022-3514.95.3.593.

Berger, Jonah, and Katherine L. Milkman. 2012. "What Makes Online Content Viral?" *Journal of Marketing Research* 49 (2): 192-205. https://doi.org/10.1509/jmr.10.0353.

Berger, Jonah, and Devin Pope. 2011. "Can Losing Lead to Winning?" *Management Science* 57 (5): 817-27. https://doi.org/10.1287/mnsc.1110.1328.

Berger, Jonah, and Eric M. Schwartz. 2011. "What Drives Immediate and Ongoing Word of Mouth?" *Journal of Marketing Research* 48 (5): 869-80. https://doi.org/10.1509/jmkr.48.5.869.

Bergh, Donald D., Brian L. Connelly, David J. Ketchen, and Lu M. Shannon. 2014. "Signalling Theory and Equilibrium in Strategic Management Research: An Assessment and a Research Agenda." *Journal of Management Studies* 51 (8): 1334-60. https://doi.org/10.1111/joms.12097.

Bergstresser, Daniel, and Thomas Philippon. 2006. "CEO Incentives and Earnings Management." *Journal of Financial Economics* 80 (3): 511-29. https://doi.org/10.1016/j.jfineco.2004.10.011.

Berle, Adolf A., and Gardiner C. Means. 1932. *The Modern Corporation and Private Property*. New York: Macmillan.

Berlyne, D. E. 1970. "Novelty, Complexity, and Hedonic Value." *Perception & Psychophysics* 8 (5): 279-86.

Berlyne, D. E. 1971. *Aesthetics and Psychobiology*. New York: Appleton-Century-Crofts.

Bernard, Victor L., and Jacob K. Thomas. 1989. "Post-Earnings-Announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research* 27: 1-36. https://doi.org/10.2307/2491062.

Bernheim, B. Douglas, and Michael D. Whinston. 1990. "Multimarket Contact and Collusive Behavior." *The RAND Journal of Economics* 21 (1): 1-26. https://doi.org/10.2307/2555490.

Bernoulli, Daniel. 1954. "Exposition of a New Theory on the Measurement of Risk." *Econometrica*

22 (1): 23-36. https://doi.org/10.2307/1909829.
Bernstein, Ethan S., and Stephen Turban. 2018. "The Impact of the 'Open' Workspace on Human Collaboration." *Philosophical Transactions of the Royal Society B: Biological Sciences* 373 (1753): 20170239. https://doi.org/10.1098/rstb.2017.0239.
Bertrand, Marianne, and Antoinette Schoar. 2003. "Managing with Style: The Effect of Managers on Firm Policies." *The Quarterly Journal of Economics* 118 (4): 1169-1208. https://doi.org/10.1162/003355303322552775.
Bettman, James R., and Barton A. Weitz. 1983. "Attributions in the Board Room: Causal Reasoning in Corporate Annual Reports." *Administrative Science Quarterly* 28 (2): 165-83. https://doi.org/10.2307/2392616.
Bewley, Truman F. 1999. *Why Wages Don't Fall During a Recession*. Cambridge, MA: Harvard University Press.
Bhattacharya, C. B., and Sankar Sen. 2003. "Consumer-Company Identification: A Framework for Understanding Consumers' Relationships with Companies." *Journal of Marketing* 67 (2): 76-88. https://doi.org/10.1509/jmkg.67.2.76.18609.
Bhattacharya, Sudipto. 1979. "Imperfect Information, Dividend Policy, and "the Bird in the Hand" Fallacy." *The Bell Journal of Economics* 10 (1): 259-70. https://doi.org/10.2307/3003330.
Bhattacherjee, Anol. 2001. "Understanding Information Systems Continuance: An Expectation-Confirmation Model." *MIS Quarterly* 25 (3): 351-70. https://doi.org/10.2307/3250921.
Bian, Yanjie. 1997. "Bringing Strong Ties Back in: Indirect Ties, Network Bridges, and Job Searches in China." *American Sociological Review* 62 (3): 366-85. https://doi.org/10.2307/2657311.
Bies, Robert J., and Joseph S. Moag. 1986. "Interactional Justice: Communication Criteria of Fairness." In *Research on Negotiation in Organizations*, edited by Roy J. Lewicki, Blair H. Sheppard, and Max H. Bazerman, 1:43-55. Greenwich, CT: JAI Press.
Bikhchandani, Sushil, David Hirshleifer, and Ivo Welch. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades." *Journal of Political Economy* 100 (5): 992-1026.
Bitner, Mary Jo. 1992. "Servicescapes: The Impact of Physical Surroundings on Customers and Employees." *Journal of Marketing* 56 (2): 57-71.
Blank, Steve. 2013. "Why the Lean Start-up Changes Everything." *Harvard Business Review* 91 (5): 63-72.
Blattberg, Robert C., and John Deighton. 1996. "Manage Marketing by the Customer Equity Test." *Harvard Business Review* 74 (4): 136-44.
Blau, Peter M. 1964. *Exchange and Power in Social Life*. New York: John Wiley & Sons.
Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *Quarterly Journal of Economics* 130 (1): 165-218. https://doi.org/10.1093/qje/qju032.
Blumler, Jay G., and Elihu Katz. 1974. *The Uses of Mass Communications: Current Perspectives on Gratifications Research*. Beverly Hills, CA: Sage.
Boczkowski, Pablo J. 2010. *News at Work: Imitation in an Age of Information Abundance*. Chicago: University of Chicago Press.
Bogen, James, and James Woodward. 1988. "Saving the Phenomena." *The Philosophical Review* 97 (3): 303-52.
Bolton, Gary E., and Axel Ockenfels. 2000. "ERC: A Theory of Equity, Reciprocity, and Competition." *American Economic Review* 90 (1): 166-93. https://doi.org/10.1257/aer.90.1.166.
Bonacich, Phillip. 1987. "Power and Centrality: A Family of Measures." *American Journal of Sociology* 92 (5): 1170-82. https://doi.org/10.1086/228631.
Bond, Charles F., and Linda J. Titus. 1983. "Social Facilitation: A Meta-Analysis of 241 Studies."

Psychological Bulletin 94 (2): 265-92. https://doi.org/10.1037/0033-2909.94.2.265.

Bond, Rod, and Peter B. Smith. 1996. "Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task." *Psychological Bulletin* 119 (1): 111-37. https://doi.org/10.1037/0033-2909.119.1.111.

Bonezzi, Andrea, C. Miguel Brendl, and Matteo De Angelis. 2011. "Stuck in the Middle: The Psychophysics of Goal Pursuit." *Psychological Science* 22 (5): 607-12.

Booms, Bernard H., and Mary Jo Bitner. 1981. "Marketing Strategies and Organization Structures for Service Firms." In *Marketing of Services*, edited by James H. Donnelly and William R. George, 47-51. Chicago: American Marketing Association.

Bordalo, Pedro, Nicola Gennaioli, and Andrei Shleifer. 2012. "Salience Theory of Choice Under Risk." *The Quarterly Journal of Economics* 127 (3): 1243-85. https://doi.org/10.1093/qje/qjs018.

Borden, Neil H. 1964. "The Concept of the Marketing Mix." *Journal of Advertising Research* 4 (2): 2-7. https://doi.org/10.1080/00218499.1964.12519724.

Borgatti, Stephen P., Ajay Mehra, Daniel J. Brass, and Giuseppe Labianca. 2009. "Network Analysis in the Social Sciences." *Science* 323 (5916): 892-95. https://doi.org/10.1126/science.1165821.

Bornstein, Robert F. 1989. "Exposure and Affect: Overview and Meta-Analysis of Research, 1968-1987." *Psychological Bulletin* 106 (2): 265-89. https://doi.org/10.1037/0033-2909.106.2.265.

Bottomley, Paul A., and Stephen J. S. Holden. 2001. "Do We Really Know How Consumers Evaluate Brand Extensions? Empirical Generalizations Based on Secondary Analysis of Eight Studies." *Journal of Marketing Research* 38 (4): 494-500. https://doi.org/10.1509/jmkr.38.4.494.18901.

Boulding, William, and Amna Kirmani. 1993. "A Consumer-Side Experimental Examination of Signaling Theory: Do Consumers Perceive Warranties as Signals of Quality?" *Journal of Consumer Research* 20 (1): 111-23. https://doi.org/10.1086/209337.

Bourgeois, L. J. 1981. "On the Measurement of Organizational Slack." *The Academy of Management Review* 6 (1): 29-39. https://doi.org/10.2307/257138.

Bower, Joseph L., and Clayton M. Christensen. 1995. "Disruptive Technologies: Catching the Wave." *Harvard Business Review* 73 (1): 43-53.

Bowman, Edward H., and Dileep Hurry. 1993. "Strategy Through the Option Lens: An Integrated View of Resource Investments and the Incremental-Choice Process." *The Academy of Management Review* 18 (4): 760-82. https://doi.org/10.2307/258597.

Brandenburger, Adam M., and Barry J. Nalebuff. 1996. *Co-Opetition*. New York: Currency Doubleday.

Brav, Alon, John R. Graham, Campbell R. Harvey, and Roni Michaely. 2005. "Payout Policy in the 21st Century." *Journal of Financial Economics* 77 (3): 483-527. https://doi.org/10.1016/j.jfineco.2004.07.004.

Brehm, Jack W. 1956. "Postdecision Changes in the Desirability of Alternatives." *Journal of Abnormal and Social Psychology* 52 (3): 384-89. https://doi.org/10.1037/h0041006.

Brehm, Jack W. 1966. *A Theory of Psychological Reactance*. New York: Academic Press.

Brehm, Sharon S., and Jack W. Brehm. 1981. *Psychological Reactance: A Theory of Freedom and Control*. New York: Academic Press.

Brett, Jeanne M. 2007. *Negotiating Globally: How to Negotiate Deals, Resolve Disputes, and Make Decisions Across Cultural Boundaries*. 2nd ed. San Francisco: Jossey-Bass.

Brewer, Marilynn B. 1991. "The Social Self: On Being the Same and Different at the Same Time." *Personality and Social Psychology Bulletin* 17 (5): 475-82. https://doi.org/10.1177/0146167291175001.

Brickley, James A., and Frederick H. Dark. 1987. "The Choice of Organizational Form: The Case of Franchising." *Journal of Financial Economics* 18 (2): 401-20. https://doi.org/10.1016/0304-405X(87)90046-8.

Brickman, Philip, and Donald T. Campbell. 1971. "Hedonic Relativism and Planning the Good Society." In *Adaptation-Level Theory: A Symposium*, edited by M. H. Appley, 287-302. New York: Academic Press.

Brickman, Philip, Dan Coates, and Ronnie Janoff-Bulman. 1978. "Lottery Winners and Accident Victims: Is Happiness Relative?" *Journal of Personality and Social Psychology* 36 (8): 917-27. https://doi.org/10.1037/0022-3514.36.8.917.

Briesch, Richard A., Lakshman Krishnamurthi, Tridib Mazumdar, and S. P. Raj. 1997. "A Comparative Analysis of Reference Price Models." *Journal of Consumer Research* 24 (2): 202-14. https://doi.org/10.1086/209505.

Brockner, Joel. 1992. "The Escalation of Commitment to a Failing Course of Action: Toward Theoretical Progress." *Academy of Management Review* 17 (1): 39-61. https://doi.org/10.2307/258647.

Brockner, Joel, Grant Ackerman, Jerald Greenberg, Michele J. Gelfand, et al. 2001. "Culture and Procedural Justice: The Influence of Power Distance on Reactions to Voice." *Journal of Experimental Social Psychology* 37 (4): 300-315.

Brockner, Joel, and E. Tory Higgins. 2001. "Regulatory Focus Theory: Implications for the Study of Emotions at Work." *Organizational Behavior and Human Decision Processes* 86 (1): 35-66. https://doi.org/10.1006/obhd.2001.2972.

Brockner, Joel, and Batia M. Wiesenfeld. 1996. "An Integrative Framework for Explaining Reactions to Decisions: Interactive Effects of Outcomes and Procedures." *Psychological Bulletin* 120 (2): 189-208. https://doi.org/10.1037/0033-2909.120.2.189.

Bromiley, Philip. 1991. "Testing a Causal Model of Corporate Risk Taking and Performance." *Academy of Management Journal* 34 (1): 37-59. https://doi.org/10.2307/256301.

Bromley, Patricia, and Walter W. Powell. 2012. "From Smoke and Mirrors to Walking the Talk: Decoupling in the Contemporary World." *Academy of Management Annals* 6 (1): 483-530. https://doi.org/10.5465/19416520.2012.684462.

Brown, Alexander L., Taisuke Imai, Ferdinand M. Vieider, and Colin F. Camerer. 2024. "Meta-Analysis of Empirical Estimates of Loss Aversion." *Journal of Economic Literature* 62 (2): 485-516. https://doi.org/10.1257/jel.20221698.

Brown, Jeffrey R., and Austan Goolsbee. 2002. "Does the Internet Make Markets More Competitive? Evidence from the Life Insurance Industry." *Journal of Political Economy* 110 (3): 481-507. https://doi.org/10.1086/339714.

Brown, Judson S. 1948. "Gradients of Approach and Avoidance Responses and Their Relation to Level of Motivation." *Journal of Comparative and Physiological Psychology* 41 (6): 450-65. https://doi.org/10.1037/h0055463.

Brown, Nicholas J. L., and Julia M. Rohrer. 2020. "Easy as (Happiness) Pie? A Critical Evaluation of a Popular Model of the Determinants of Well-Being." *Journal of Happiness Studies* 21 (4): 1285-1301.

Brown, Nicholas J. L., Alan D. Sokal, and Harris L. Friedman. 2013. "The Complex Dynamics of Wishful Thinking: The Critical Positivity Ratio." *American Psychologist* 68 (9): 801-13. https://doi.org/10.1037/a0032850.

Bruckmüller, Susanne, and Nyla R. Branscombe. 2010. "The Glass Cliff: When and Why Women Are Selected as Leaders in Crisis Contexts." *British Journal of Social Psychology* 49 (3): 433-51. https://doi.org/10.1348/014466609X466594.

Brüderl, Josef, and Rudolf Schüssler. 1990. "Organizational Mortality: The Liabilities of Newness and Adolescence." *Administrative Science Quarterly* 35 (3): 530-47. https://doi.org/10.2307/2393316.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond. 2025. "Generative AI at Work." *The Quarterly*

Journal of Economics 140 (2): 889-942. https://doi.org/10.1093/qje/qjae044.
Brynjolfsson, Erik, and Michael D. Smith. 2000. "Frictionless Commerce? A Comparison of Internet and Conventional Retailers." *Management Science* 46 (4): 563-85. https://doi.org/10.1287/mnsc.46.4.563.12061.
Buckley, Peter J., and Mark Casson. 1976. *The Future of the Multinational Enterprise*. London: Macmillan.
Buckley, Peter J., and Mark C. Casson. 2009. "The Internalisation Theory of the Multinational Enterprise: A Review of the Progress of a Research Agenda After 30 Years." *Journal of International Business Studies* 40 (9): 1563-80. https://doi.org/10.1057/jibs.2009.49.
Buechel, Eva C., and Jonah Berger. 2018. "Microblogging and the Value of Undirected Communication." *Journal of Consumer Psychology* 28 (1): 40-55. https://doi.org/10.1002/jcpy.1013.
Bullock, John G., Donald P. Green, and Shang E. Ha. 2010. "Yes, but What's the Mechanism? (Don't Expect an Easy Answer)." *Journal of Personality and Social Psychology* 98 (4): 550-58. https://doi.org/10.1037/a0018933.
Burdett, Kenneth, and Kenneth L. Judd. 1983. "Equilibrium Price Dispersion." *Econometrica* 51 (4): 955-69. https://doi.org/10.2307/1912045.
Burgelman, Robert A. 1994. "Fading Memories: A Process Theory of Strategic Business Exit in Dynamic Environments." *Administrative Science Quarterly* 39 (1): 24-56. https://doi.org/10.2307/2393493.
Burger, Jerry M. 1999. "The Foot-in-the-Door Compliance Procedure: A Multiple-Process Analysis and Review." *Personality and Social Psychology Review* 3 (4): 303-25. https://doi.org/10.1207/s15327957pspr0304_2.
Burks, Stephen V., Bo Cowgill, Mitchell Hoffman, and Michael Housman. 2015. "The Value of Hiring Through Employee Referrals." *The Quarterly Journal of Economics* 130 (2): 805-39. https://doi.org/10.1093/qje/qjv010.
Burns, James MacGregor. 1978. *Leadership*. New York: Harper & Row.
Burns, Tom, and G. M. Stalker. 1961. *The Management of Innovation*. London: Tavistock.
Burt, Ronald S. 1992. *Structural Holes: The Social Structure of Competition*. Cambridge, MA: Harvard University Press.
Burt, Ronald S. 2004. "Structural Holes and Good Ideas." *American Journal of Sociology* 110 (2): 349-99. https://doi.org/10.1086/421787.
Burt, Ronald S. 2005. *Brokerage and Closure: An Introduction to Social Capital*. Oxford: Oxford University Press.
Business Roundtable. 2019. "Statement on the Purpose of a Corporation."
Byrne, Donn. 1971. *The Attraction Paradigm*. New York: Academic Press.
Cable, Daniel M., and D. Scott DeRue. 2002. "The Convergent and Discriminant Validity of Subjective Fit Perceptions." *Journal of Applied Psychology* 87 (5): 875-84. https://doi.org/10.1037/0021-9010.87.5.875.
Cachon, Gérard P., Taylor Randall, and Glen M. Schmidt. 2007. "In Search of the Bullwhip Effect." *Manufacturing & Service Operations Management* 9 (4): 457-79. https://doi.org/10.1287/msom.1060.0149.
Cacioppo, John T., and Richard E. Petty. 1979. "Effects of Message Repetition and Position on Cognitive Response, Recall, and Persuasion." *Journal of Personality and Social Psychology* 37 (1): 97-109. https://doi.org/10.1037/0022-3514.37.1.97.
Cacioppo, John T., and Richard E. Petty. 1982. "The Need for Cognition." *Journal of Personality and Social Psychology* 42 (1): 116-31. https://doi.org/10.1037/0022-3514.42.1.116.
Cai, Hongbin, Yuyu Chen, and Hanming Fang. 2009. "Observational Learning: Evidence from a Randomized Natural Field Experiment." *American Economic Review* 99 (3): 864-82. https://do

i.org/10.1257/aer.99.3.864.
Caillaud, Bernard, and Bruno Jullien. 2003. "Chicken & Egg: Competition Among Intermediation Service Providers." *The RAND Journal of Economics* 34 (2): 309-28. https://doi.org/10.2307/1593720.
Camerer, Colin F. 1991. "Does Strategy Research Need Game Theory?" *Strategic Management Journal* 12 (S2): 137-52. https://doi.org/10.1002/smj.4250121010.
Camerer, Colin F., Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, et al. 2018. "Evaluating the Replicability of Social Science Experiments in Nature and Science Between 2010 and 2015." *Nature Human Behaviour* 2 (9): 637-44. https://doi.org/10.1038/s41562-018-0399-z.
Camerer, Colin, Linda Babcock, George Loewenstein, and Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics* 112 (2): 407-41. https://doi.org/10.1162/003355397555244.
Camerer, Colin, and Richard H. Thaler. 1995. "Anomalies: Ultimatums, Dictators and Manners." *Journal of Economic Perspectives* 9 (2): 209-19. https://doi.org/10.1257/jep.9.2.209.
Cameron, Judy, and W. David Pierce. 1994. "Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis." *Review of Educational Research* 64 (3): 363-423.
Campbell, Donald T. 1960. "Blind Variation and Selective Retention in Creative Thought as in Other Knowledge Processes." *Psychological Review* 67 (6): 380-400.
Campbell, Donald T., and Donald W. Fiske. 1959. "Convergent and Discriminant Validation by the Multitrait-Multimethod Matrix." *Psychological Bulletin* 56 (2): 81-105. https://doi.org/10.1037/h0046016.
Campbell, Margaret C., and Ronald C. Goodstein. 2001. "The Moderating Effect of Perceived Risk on Consumers' Evaluations of Product Incongruity: Preference for the Norm." *Journal of Consumer Research* 28 (3): 439-49. https://doi.org/10.1086/323731.
Campbell, Margaret C., and Amna Kirmani. 2000. "Consumers' Use of Persuasion Knowledge: The Effects of Accessibility and Cognitive Capacity on Perceptions of an Influence Agent." *Journal of Consumer Research* 27 (1): 69-83. https://doi.org/10.1086/314309.
Camuffo, Arnaldo, Alessandro Cordova, Alfonso Gambardella, and Chiara Spina. 2020. "A Scientific Approach to Entrepreneurial Decision Making: Evidence from a Randomized Control Trial." *Management Science* 66 (2): 564-86. https://doi.org/10.1287/mnsc.2018.3249.
Cannon, Mark D., and Amy C. Edmondson. 2005. "Failing to Learn and Learning to Fail (Intelligently): How Great Organizations Put Failure to Work to Innovate and Improve." *Long Range Planning* 38 (3): 299-319. https://doi.org/10.1016/j.lrp.2005.04.005.
Capen, E. C., R. V. Clapp, and W. M. Campbell. 1971. "Competitive Bidding in High-Risk Situations." *Journal of Petroleum Technology* 23 (6): 641-53. https://doi.org/10.2118/2993-PA.
Cappelli, Peter, and Keith Chauvin. 1991. "An Interplant Test of the Efficiency Wage Hypothesis." *The Quarterly Journal of Economics* 106 (3): 769-87. https://doi.org/10.2307/2937926.
Card, David, and Gordon B. Dahl. 2011. "Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior." *The Quarterly Journal of Economics* 126 (1): 103-43. https://doi.org/10.1093/qje/qjr001.
Card, David, Alexandre Mas, Enrico Moretti, and Emmanuel Saez. 2012. "Inequality at Work: The Effect of Peer Salaries on Job Satisfaction." *American Economic Review* 102 (6): 2981-3003.
Carhart, Mark M. 1997. "On Persistence in Mutual Fund Performance." *The Journal of Finance* 52 (1): 57-82. https://doi.org/10.1111/j.1540-6261.1997.tb03808.x.
Carmon, Ziv, and Dan Ariely. 2000. "Focusing on the Forgone: How Value Can Appear so Different to Buyers and Sellers." *Journal of Consumer Research* 27 (3): 360-70.
Carpenter, Christopher J. 2013. "A Meta-Analysis of the Effectiveness of the 'but You Are Free'

Compliance-Gaining Technique." *Communication Studies* 64 (1): 6-17. https://doi.org/10.1080/10510974.2012.727941.

Carpenter, Gregory S., and Kent Nakamoto. 1989. "Consumer Preference Formation and Pioneering Advantage." *Journal of Marketing Research* 26 (3): 285-98. https://doi.org/10.2307/3172901.

Carroll, Archie B. 1979. "A Three-Dimensional Conceptual Model of Corporate Performance." *Academy of Management Review* 4 (4): 497-505. https://doi.org/10.2307/257850.

Carroll, Archie B. 1991. "The Pyramid of Corporate Social Responsibility: Toward the Moral Management of Organizational Stakeholders." *Business Horizons* 34 (4): 39-48. https://doi.org/10.1016/0007-6813(91)90005-G.

Carroll, Glenn R. 1985. "Concentration and Specialization: Dynamics of Niche Width in Populations of Organizations." *American Journal of Sociology* 90 (6): 1262-83. https://doi.org/10.1086/228210.

Carroll, Glenn R., and Anand Swaminathan. 2000. "Why the Microbrewery Movement? Organizational Dynamics of Resource Partitioning in the U.S. Brewing Industry." *American Journal of Sociology* 106 (3): 715-62. https://doi.org/10.1086/318962.

Carter, Travis J., and Thomas Gilovich. 2010. "The Relative Relativity of Material and Experiential Purchases." *Journal of Personality and Social Psychology* 98 (1): 146-59.

Cartwright, Dorwin, and Frank Harary. 1956. "Structural Balance: A Generalization of Heider's Theory." *Psychological Review* 63 (5): 277-93. https://doi.org/10.1037/h0046049.

Casadesus-Masanell, Ramon, and Daniel F. Spulber. 2000. "The Fable of Fisher Body." *The Journal of Law and Economics* 43 (1): 67-104. https://doi.org/10.1086/467448.

Casciaro, Tiziana, and Mikołaj Jan Piskorski. 2005. "Power Imbalance, Mutual Dependence, and Constraint Absorption: A Closer Look at Resource Dependence Theory." *Administrative Science Quarterly* 50 (2): 167-99. https://doi.org/10.2189/asqu.2005.50.2.167.

Cascio, Wayne F. 2006. "Decency Means More Than 'Always Low Prices': A Comparison of Costco to Wal-Mart's Sam's Club." *Academy of Management Perspectives* 20 (3): 26-37. https://doi.org/10.5465/amp.2006.21903478.

Caves, Richard E., and Michael E. Porter. 1977. "From Entry Barriers to Mobility Barriers: Conjectural Decisions and Contrived Deterrence to New Competition." *The Quarterly Journal of Economics* 91 (2): 241-61. https://doi.org/10.2307/1885416.

Centola, Damon. 2010. "The Spread of Behavior in an Online Social Network Experiment." *Science* 329 (5996): 1194-97. https://doi.org/10.1126/science.1185231.

Cepeda, Nicholas J., Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. 2006. "Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis." *Psychological Bulletin* 132 (3): 354-80.

Cerasoli, Christopher P., Jessica M. Nicklin, and Michael T. Ford. 2014. "Intrinsic Motivation and Extrinsic Incentives Jointly Predict Performance: A 40-Year Meta-Analysis." *Psychological Bulletin* 140 (4): 980-1008.

Certo, S. Trevis. 2003. "Influencing Initial Public Offering Investors with Prestige: Signaling with Board Structures." *Academy of Management Review* 28 (3): 432-46. https://doi.org/10.5465/amr.2003.10196754.

Cesario, Joseph, Heidi Grant, and E. Tory Higgins. 2004. "Regulatory Fit and Persuasion: Transfer from "Feeling Right"." *Journal of Personality and Social Psychology* 86 (3): 388-404. https://doi.org/10.1037/0022-3514.86.3.388.

Chaiken, Shelly. 1980. "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion." *Journal of Personality and Social Psychology* 39 (5): 752-66. https://doi.org/10.1037/0022-3514.39.5.752.

Chan, Cindy, Jonah Berger, and Leaf Van Boven. 2012. "Identifiable but Not Identical: Combining

Social Identity and Uniqueness Motives in Choice." *Journal of Consumer Research* 39 (3): 561-73. https://doi.org/10.1086/664804.

Chartrand, Tanya L., and John A. Bargh. 1999. "The Chameleon Effect: The Perception-Behavior Link and Social Interaction." *Journal of Personality and Social Psychology* 76 (6): 893-910. https://doi.org/10.1037/0022-3514.76.6.893.

Chase, Richard B., and Sriram Dasu. 2001. "Want to Perfect Your Company's Service? Use Behavioral Science." *Harvard Business Review* 79 (6): 78-84.

Chatman, Jennifer A. 1989. "Improving Interactional Organizational Research: A Model of Person-Organization Fit." *Academy of Management Review* 14 (3): 333-49. https://doi.org/10.2307/258171.

Chatman, Jennifer A., and Charles A. O'Reilly. 2016. "Paradigm Lost: Reinvigorating the Study of Organizational Culture." *Research in Organizational Behavior* 36: 199-224. https://doi.org/10.1016/j.riob.2016.11.004.

Chatterjee, Arijit, and Donald C. Hambrick. 2007. "It's All about Me: Narcissistic Chief Executive Officers and Their Effects on Company Strategy and Performance." *Administrative Science Quarterly* 52 (3): 351-86. https://doi.org/10.2189/asqu.52.3.351.

Chen, Frank, Zvi Drezner, Jennifer K. Ryan, and David Simchi-Levi. 2000. "Quantifying the Bullwhip Effect in a Simple Supply Chain: The Impact of Forecasting, Lead Times, and Information." *Management Science* 46 (3): 436-43. https://doi.org/10.1287/mnsc.46.3.436.12069.

Chen, M. Keith, and Jane L. Risen. 2010. "How Choice Affects and Reflects Preferences: Revisiting the Free-Choice Paradigm." *Journal of Personality and Social Psychology* 99 (4): 573-94. https://doi.org/10.1037/a0020217.

Chen, Ming-Jer. 1996. "Competitor Analysis and Interfirm Rivalry: Toward a Theoretical Integration." *The Academy of Management Review* 21 (1): 100-134. https://doi.org/10.2307/258631.

Chen, Ming-Jer, and Danny Miller. 2012. "Competitive Dynamics: Themes, Trends, and a Prospective Research Platform." *Academy of Management Annals* 6 (1): 135-210. https://doi.org/10.5465/19416520.2012.660762.

Chen, Ming-Jer, Ken G. Smith, and Curtis M. Grimm. 1992. "Action Characteristics as Predictors of Competitive Responses." *Management Science* 38 (3): 439-55. https://doi.org/10.1287/mnsc.38.3.439.

Chen, Shisan C. 1937. "Social Modification of the Activity of Ants in Nest-Building." *Physiological Zoology* 10 (4): 420-36. https://doi.org/10.1086/physzool.10.4.30151428.

Chernev, Alexander, Ulf Böckenholt, and Joseph Goodman. 2015. "Choice Overload: A Conceptual Review and Meta-Analysis." *Journal of Consumer Psychology* 25 (2): 333-58.

Chesbrough, Henry W. 2003. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston, MA: Harvard Business School Press.

Cheung, Steven N. S. 1973. "The Fable of the Bees: An Economic Investigation." *The Journal of Law and Economics* 16 (1): 11-33. https://doi.org/10.1086/466753.

Chevalier, Judith A., and Dina Mayzlin. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews." *Journal of Marketing Research* 43 (3): 345-54. https://doi.org/10.1509/jmkr.43.3.345.

Child, John. 1972. "Organizational Structure, Environment and Performance: The Role of Strategic Choice." *Sociology* 6 (1): 1-22. https://doi.org/10.1177/003803857200600101.

Cho, Theresa S., and Donald C. Hambrick. 2006. "Attention as the Mediator Between Top Management Team Characteristics and Strategic Change: The Case of Airline Deregulation." *Organization Science* 17 (4): 453-69. https://doi.org/10.1287/orsc.1060.0192.

Christensen, Clayton M. 1997. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston, MA: Harvard Business School Press.

Christensen, Clayton M., and Joseph L. Bower. 1996. "Customer Power, Strategic Investment, and the Failure of Leading Firms." *Strategic Management Journal* 17 (3): 197-218.

Christensen, Clayton M., and Michael E. Raynor. 2003. "Why Hard-Nosed Executives Should Care about Management Theory." *Harvard Business Review* 81 (9): 66-74.

Christensen, Clayton M., Michael E. Raynor, and Rory McDonald. 2015. "What Is Disruptive Innovation?" *Harvard Business Review* 93 (12): 44-53.

Christensen, Clayton M., and Richard S. Rosenbloom. 1995. "Explaining the Attacker's Advantage: Technological Paradigms, Organizational Dynamics, and the Value Network." *Research Policy* 24 (2): 233-57. https://doi.org/10.1016/0048-7333(93)00764-K.

Christopherson, Susan. 2007. "Barriers to 'US Style' Lean Retailing: The Case of Wal-Mart's Failure in Germany." *Journal of Economic Geography* 7 (4): 451-69. https://doi.org/10.1093/jeg/lbm010.

Chua, Roy Y. J., Yannig Roth, and Jean-François Lemoine. 2015. "The Impact of Culture on Creativity: How Cultural Tightness and Cultural Distance Affect Global Innovation Crowdsourcing Work." *Administrative Science Quarterly* 60 (2): 189-227.

Chung, Doug J., Thomas Steenburgh, and K. Sudhir. 2014. "Do Bonuses Enhance Sales Productivity? A Dynamic Structural Analysis of Bonus-Based Compensation Plans." *Marketing Science* 33 (2): 165-87.

Churchill, Jr., Gilbert A., and Carol Surprenant. 1982. "An Investigation into the Determinants of Customer Satisfaction." *Journal of Marketing Research* 19 (4): 491-504. https://doi.org/10.1177/002224378201900410.

Cialdini, Robert B., and Noah J. Goldstein. 2004. "Social Influence: Compliance and Conformity." *Annual Review of Psychology* 55: 591-621. https://doi.org/10.1146/annurev.psych.55.090902.142015.

Cialdini, Robert B., Raymond R. Reno, and Carl A. Kallgren. 1990. "A Focus Theory of Normative Conduct: Recycling the Concept of Norms to Reduce Littering in Public Places." *Journal of Personality and Social Psychology* 58 (6): 1015-26. https://doi.org/10.1037/0022-3514.58.6.1015.

Clee, Mona A., and Robert A. Wicklund. 1980. "Consumer Behavior and Psychological Reactance." *Journal of Consumer Research* 6 (4): 389-405. https://doi.org/10.1086/208782.

Clemen, Robert T. 1989. "Combining Forecasts: A Review and Annotated Bibliography." *International Journal of Forecasting* 5 (4): 559-83.

Coase, R. H. 1937. "The Nature of the Firm." *Economica* 4 (16): 386-405. https://doi.org/10.1111/j.1468-0335.1937.tb00002.x.

Coase, R. H. 1959. "The Federal Communications Commission." *The Journal of Law and Economics* 2: 1-40. https://doi.org/10.1086/466549.

Coase, R. H. 1960. "The Problem of Social Cost." *The Journal of Law and Economics* 3: 1-44. https://doi.org/10.1086/466560.

Coase, R. H. 2000. "The Acquisition of Fisher Body by General Motors." *The Journal of Law and Economics* 43 (1): 15-32. https://doi.org/10.1086/467446.

Coff, Russell W. 1997. "Human Assets and Management Dilemmas: Coping with Hazards on the Road to Resource-Based Theory." *Academy of Management Review* 22 (2): 374-402. https://doi.org/10.2307/259327.

Cohen, Bernard C. 1963. *The Press and Foreign Policy*. Princeton, NJ: Princeton University Press.

Cohen, Michael D., James G. March, and Johan P. Olsen. 1972. "A Garbage Can Model of Organizational Choice." *Administrative Science Quarterly* 17 (1): 1-25. https://doi.org/10.2307/2392088.

Cohen, Wesley M., and Daniel A. Levinthal. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation." *Administrative Science Quarterly* 35 (1): 128-52. https://doi.org/10

.2307/2393553.
Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94: S95-120. https://doi.org/10.1086/228943.
Coleman, James S. 1990. *Foundations of Social Theory*. Cambridge, MA: Belknap Press of Harvard University Press.
Coleman, James S., Elihu Katz, and Herbert Menzel. 1966. *Medical Innovation: A Diffusion Study*. Indianapolis: Bobbs-Merrill.
Collins, Rebecca L. 1996. "For Better or Worse: The Impact of Upward Social Comparison on Self-Evaluations." *Psychological Bulletin* 119 (1): 51-69. https://doi.org/10.1037/0033-2909.119.1.51.
Collis, David J. 1994. "Research Note: How Valuable Are Organizational Capabilities?" *Strategic Management Journal* 15 (S1): 143-52. https://doi.org/10.1002/smj.4250150910.
Colquitt, Jason A. 2001. "On the Dimensionality of Organizational Justice: A Construct Validation of a Measure." *Journal of Applied Psychology* 86 (3): 386-400. https://doi.org/10.1037/0021-9010.86.3.386.
Colquitt, Jason A., Donald E. Conlon, Michael J. Wesson, Christopher O. L. H. Porter, and K. Yee Ng. 2001. "Justice at the Millennium: A Meta-Analytic Review of 25 Years of Organizational Justice Research." *Journal of Applied Psychology* 86 (3): 425-45. https://doi.org/10.1037/0021-9010.86.3.425.
Colquitt, Jason A., and Cindy P. Zapata-Phelan. 2007. "Trends in Theory Building and Theory Testing: A Five-Decade Study of the Academy of Management Journal." *Academy of Management Journal* 50 (6): 1281-1303. https://doi.org/10.5465/amj.2007.28165855.
Columbia Accident Investigation Board. 2003. *Report of the Columbia Accident Investigation Board, Volume I*. Washington, DC: National Aeronautics and Space Administration and U.S. Government Printing Office.
Compeau, Deborah R., and Christopher A. Higgins. 1995. "Computer Self-Efficacy: Development of a Measure and Initial Test." *MIS Quarterly* 19 (2): 189-211. https://doi.org/10.2307/249688.
Connelly, Brian L., S. Trevis Certo, R. Duane Ireland, and Christopher R. Reutzel. 2011. "Signaling Theory: A Review and Assessment." *Journal of Management* 37 (1): 39-67. https://doi.org/10.1177/0149206310388419.
Connelly, Brian L., Laszlo Tihanyi, T. Russell Crook, and K. Ashley Gangloff. 2014. "Tournament Theory: Thirty Years of Contests and Competitions." *Journal of Management* 40 (1): 16-47. https://doi.org/10.1177/0149206313498902.
Constantinides, Efthymios. 2006. "The Marketing Mix Revisited: Towards the 21st Century Marketing." *Journal of Marketing Management* 22 (3-4): 407-38. https://doi.org/10.1362/026725706776861190.
Cook, Alison, and Christy Glass. 2014. "Above the Glass Ceiling: When Are Women and Racial/Ethnic Minorities Promoted to CEO?" *Strategic Management Journal* 35 (7): 1080-89. https://doi.org/10.1002/smj.2161.
Cook, Karen S., and Richard M. Emerson. 1978. "Power, Equity and Commitment in Exchange Networks." *American Sociological Review* 43 (5): 721-39. https://doi.org/10.2307/2094546.
Cook, Karen S., Richard M. Emerson, Mary R. Gillmore, and Toshio Yamagishi. 1983. "The Distribution of Power in Exchange Networks: Theory and Experimental Results." *American Journal of Sociology* 89 (2): 275-305. https://doi.org/10.1086/227866.
Cook, Karen S., and J. M. Whitmeyer. 1992. "Two Approaches to Social Structure: Exchange Theory and Network Analysis." *Annual Review of Sociology* 18: 109-27. https://doi.org/10.1146/annurev.so.18.080192.000545.
Cool, Karel O., and Dan Schendel. 1987. "Strategic Group Formation and Performance: The Case

of the U.S. Pharmaceutical Industry, 1963-1982." *Management Science* 33 (9): 1102-24. https://doi.org/10.1287/mnsc.33.9.1102.

Cool, Karel, and Dan Schendel. 1988. "Performance Differences Among Strategic Group Members." *Strategic Management Journal* 9 (3): 207-23. https://doi.org/10.1002/smj.4250090302.

Coombs, W. Timothy. 2007. "Protecting Organization Reputations During a Crisis: The Development and Application of Situational Crisis Communication Theory." *Corporate Reputation Review* 10 (3): 163-76. https://doi.org/10.1057/palgrave.crr.1550049.

Cooper, Joel, and Russell H. Fazio. 1984. "A New Look at Dissonance Theory." In *Advances in Experimental Social Psychology*, edited by Leonard Berkowitz, 17:229-66. Orlando, FL: Academic Press. https://doi.org/10.1016/S0065-2601(08)60121-5.

Corley, Kevin G., and Dennis A. Gioia. 2011. "Building Theory about Theory Building: What Constitutes a Theoretical Contribution?" *Academy of Management Review* 36 (1): 12-32. https://doi.org/10.5465/amr.2009.0486.

Cottrell, Nickolas B., Dennis L. Wack, Gary J. Sekerak, and Robert H. Rittle. 1968. "Social Facilitation of Dominant Responses by the Presence of an Audience and the Mere Presence of Others." *Journal of Personality and Social Psychology* 9 (3): 245-50. https://doi.org/10.1037/h0025902.

Coughlan, Anne T., and David A. Soberman. 2005. "Strategic Segmentation Using Outlet Malls." *International Journal of Research in Marketing* 22 (1): 61-86. https://doi.org/10.1016/j.ijresmar.2004.03.003.

Cowan, Nelson. 2001. "The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity." *Behavioral and Brain Sciences* 24 (1): 87-114.

Cowgill, Bo, and Eric Zitzewitz. 2015. "Corporate Prediction Markets: Evidence from Google, Ford, and Firm x." *The Review of Economic Studies* 82 (4): 1309-41.

Craik, Fergus I. M., and Robert S. Lockhart. 1972. "Levels of Processing: A Framework for Memory Research." *Journal of Verbal Learning and Verbal Behavior* 11 (6): 671-84.

Crane, Andrew, Guido Palazzo, Laura J. Spence, and Dirk Matten. 2014. "Contesting the Value of 'Creating Shared Value'." *California Management Review* 56 (2): 130-53. https://doi.org/10.1525/cmr.2014.56.2.130.

Crawford, Eean R., Jeffery A. LePine, and Bruce Louis Rich. 2010. "Linking Job Demands and Resources to Employee Engagement and Burnout: A Theoretical Extension and Meta-Analytic Test." *Journal of Applied Psychology* 95 (5): 834-48. https://doi.org/10.1037/a0019364.

Crawford, Vincent P., and Juanjuan Meng. 2011. "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income." *American Economic Review* 101 (5): 1912-32. https://doi.org/10.1257/aer.101.5.1912.

Cronbach, Lee J., and Paul E. Meehl. 1955. "Construct Validity in Psychological Tests." *Psychological Bulletin* 52 (4): 281-302. https://doi.org/10.1037/h0040957.

Cronin, J. Joseph, and Steven A. Taylor. 1992. "Measuring Service Quality: A Reexamination and Extension." *Journal of Marketing* 56 (3): 55-68. https://doi.org/10.1177/002224299205600304.

Crook, T. Russell, Jr. Ketchen David J., James G. Combs, and Samuel Y. Todd. 2008. "Strategic Resources and Performance: A Meta-Analysis." *Strategic Management Journal* 29 (11): 1141-54. https://doi.org/10.1002/smj.703.

Crook, T. Russell, Samuel Y. Todd, James G. Combs, David J. Woehr, and Jr. Ketchen David J. 2011. "Does Human Capital Matter? A Meta-Analysis of the Relationship Between Human Capital and Firm Performance." *Journal of Applied Psychology* 96 (3): 443-56. https://doi.org/10.1037/a0022147.

Cropanzano, Russell, and Marie S. Mitchell. 2005. "Social Exchange Theory: An Interdisciplinary Review." *Journal of Management* 31 (6): 874-900. https://doi.org/10.1177/0149206305279602

02.
Crouse, Timothy. 1973. *The Boys on the Bus*. New York: Random House.
Crowe, Ellen, and E. Tory Higgins. 1997. "Regulatory Focus and Strategic Inclinations: Promotion and Prevention in Decision-Making." *Organizational Behavior and Human Decision Processes* 69 (2): 117-32. https://doi.org/10.1006/obhd.1996.2675.
Cuddy, Amy J. C., Kyle Todd Doherty, and Maarten W. Bos. 2010. "OPOWER: Increasing Energy Efficiency Through Normative Influence (a)." *Harvard Business School Case*, September. https://www.hbs.edu/faculty/Pages/item.aspx?num=39383.
Cummings, William H., and M. Venkatesan. 1976. "Cognitive Dissonance and Consumer Behavior: A Review of the Evidence." *Journal of Marketing Research* 13 (3): 303-8. https://doi.org/10.1177/002224377601300313.
Cutolo, Donato, and Martin Kenney. 2021. "Platform-Dependent Entrepreneurs: Power Asymmetries, Risks, and Strategies in the Platform Economy." *Academy of Management Perspectives* 35 (4): 584-605. https://doi.org/10.5465/amp.2019.0103.
Cyert, Richard M., and James G. March. 1963. *A Behavioral Theory of the Firm*. Englewood Cliffs, NJ: Prentice-Hall.
Dahlander, Linus, and David M. Gann. 2010. "How Open Is Innovation?" *Research Policy* 39 (6): 699-709. https://doi.org/10.1016/j.respol.2010.01.013.
Dai, Hengchen, Berkeley J. Dietvorst, Bradford Tuckfield, Katherine L. Milkman, and Maurice E. Schweitzer. 2018. "Quitting When the Going Gets Tough: A Downside of High Performance Expectations." *Academy of Management Journal* 61 (5): 1667-91. https://doi.org/10.5465/amj.2014.1045.
Daily, Catherine M., Dan R. Dalton, and Jr. Cannella Albert A. 2003. "Corporate Governance: Decades of Dialogue and Data." *Academy of Management Review* 28 (3): 371-82. https://doi.org/10.2307/30040727.
Dalton, Dan R., Catherine M. Daily, Alan E. Ellstrand, and Jonathan L. Johnson. 1998. "Meta-Analytic Reviews of Board Composition, Leadership Structure, and Financial Performance." *Strategic Management Journal* 19 (3): 269-90.
Dansereau, Jr., Fred, George Graen, and William J. Haga. 1975. "A Vertical Dyad Linkage Approach to Leadership Within Formal Organizations: A Longitudinal Investigation of the Role Making Process." *Organizational Behavior and Human Performance* 13 (1): 46-78. https://doi.org/10.1016/0030-5073(75)90005-7.
Darley, John M., and Bibb Latané. 1968. "Bystander Intervention in Emergencies: Diffusion of Responsibility." *Journal of Personality and Social Psychology* 8 (4): 377-83. https://doi.org/10.1037/h0025589.
Darr, Eric D., Linda Argote, and Dennis Epple. 1995. "The Acquisition, Transfer, and Depreciation of Knowledge in Service Organizations: Productivity in Franchises." *Management Science* 41 (11): 1750-62. https://doi.org/10.1287/mnsc.41.11.1750.
David, Paul A. 1985. "Clio and the Economics of QWERTY." *American Economic Review* 75 (2): 332-37.
David, Robert J., and Shin-Kap Han. 2004. "A Systematic Assessment of the Empirical Support for Transaction Cost Economics." *Strategic Management Journal* 25 (1): 39-58. https://doi.org/10.1002/smj.359.
Davis, Fred D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology." *MIS Quarterly* 13 (3): 319-40. https://doi.org/10.2307/249008.
Davis, Fred D., Richard P. Bagozzi, and Paul R. Warshaw. 1989. "User Acceptance of Computer Technology: A Comparison of Two Theoretical Models." *Management Science* 35 (8): 982-1003. https://doi.org/10.1287/mnsc.35.8.982.

Davis, James H., F. David Schoorman, and Lex Donaldson. 1997. "Toward a Stewardship Theory of Management." *Academy of Management Review* 22 (1): 20-47. https://doi.org/10.5465/amr.1997.9707180258.

Davis, Murray S. 1971. "That's Interesting! Towards a Phenomenology of Sociology and a Sociology of Phenomenology." *Philosophy of the Social Sciences* 1 (2): 309-44. https://doi.org/10.1177/004839317100100211.

Dawkins, Richard, and Tamsin R. Carlisle. 1976. "Parental Investment, Mate Desertion and a Fallacy." *Nature* 262 (5564): 131-33. https://doi.org/10.1038/262131a0.

Day, George S. 1981. "The Product Life Cycle: Analysis and Applications Issues." *Journal of Marketing* 45 (4): 60-67. https://doi.org/10.1177/002224298104500408.

Day, George S. 1994. "The Capabilities of Market-Driven Organizations." *Journal of Marketing* 58 (4): 37-52. https://doi.org/10.1177/002224299405800404.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann. 1990. "Noise Trader Risk in Financial Markets." *Journal of Political Economy* 98 (4): 703-38. https://doi.org/10.1086/261703.

Deci, Edward L., Richard Koestner, and Richard M. Ryan. 1999. "A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation." *Psychological Bulletin* 125 (6): 627-68. https://doi.org/10.1037/0033-2909.125.6.627.

Deci, Edward L., and Richard M. Ryan. 1985. *Intrinsic Motivation and Self-Determination in Human Behavior*. New York: Plenum.

Deephouse, David L. 1999. "To Be Different, or to Be the Same? It's a Question (and Theory) of Strategic Balance." *Strategic Management Journal* 20 (2): 147-66.

DellaVigna, Stefano. 2009. "Psychology and Economics: Evidence from the Field." *Journal of Economic Literature* 47 (2): 315-72. https://doi.org/10.1257/jel.47.2.315.

DellaVigna, Stefano, Attila Lindner, Balázs Reizer, and Johannes F. Schmieder. 2017. "Reference-Dependent Job Search: Evidence from Hungary." *The Quarterly Journal of Economics* 132 (4): 1969-2018. https://doi.org/10.1093/qje/qjx015.

DellaVigna, Stefano, and Ulrike Malmendier. 2006. "Paying Not to Go to the Gym." *American Economic Review* 96 (3): 694-719. https://doi.org/10.1257/aer.96.3.694.

DeLone, William H., and Ephraim R. McLean. 1992. "Information Systems Success: The Quest for the Dependent Variable." *Information Systems Research* 3 (1): 60-95. https://doi.org/10.1287/isre.3.1.60.

DeLone, William H., and Ephraim R. McLean. 2003. "The DeLone and McLean Model of Information Systems Success: A Ten-Year Update." *Journal of Management Information Systems* 19 (4): 9-30. https://doi.org/10.1080/07421222.2003.11045748.

Demerouti, Evangelia, Arnold B. Bakker, Friedhelm Nachreiner, and Wilmar B. Schaufeli. 2001. "The Job Demands-Resources Model of Burnout." *Journal of Applied Psychology* 86 (3): 499-512. https://doi.org/10.1037/0021-9010.86.3.499.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2009. "Optimal Versus Naive Diversification: How Inefficient Is the 1/n Portfolio Strategy?" *The Review of Financial Studies* 22 (5): 1915-53. https://doi.org/10.1093/rfs/hhm075.

Demsetz, Harold. 1973. "Industry Structure, Market Rivalry, and Public Policy." *The Journal of Law and Economics* 16 (1): 1-9. https://doi.org/10.1086/466752.

Detert, James R., and Ethan R. Burris. 2007. "Leadership Behavior and Employee Voice: Is the Door Really Open?" *Academy of Management Journal* 50 (4): 869-84.

Deutsch, Morton, and Harold B. Gerard. 1955. "A Study of Normative and Informational Social Influences Upon Individual Judgment." *The Journal of Abnormal and Social Psychology* 51 (3): 629-36. https://doi.org/10.1037/h0046408.

Dewsbury, Donald A. 1981. "Effects of Novelty on Copulatory Behavior: The Coolidge Effect and Related Phenomena." *Psychological Bulletin* 89 (3): 464-82. https://doi.org/10.1037/0033-2909.89.3.464.

Dhalla, Nariman K., and Sonia Yuspeh. 1976. "Forget the Product Life Cycle Concept!" *Harvard Business Review* 54 (1): 102-12.

Dhar, Ravi, and Klaus Wertenbroch. 2000. "Consumer Choice Between Hedonic and Utilitarian Goods." *Journal of Marketing Research* 37 (1): 60-71. https://doi.org/10.1509/jmkr.37.1.60.18718.

Diamond, Peter A. 1971. "A Model of Price Adjustment." *Journal of Economic Theory* 3 (2): 156-68. https://doi.org/10.1016/0022-0531(71)90013-5.

Diener, Ed, Richard E. Lucas, and Christie Napa Scollon. 2006. "Beyond the Hedonic Treadmill: Revising the Adaptation Theory of Well-Being." *American Psychologist* 61 (4): 305-14. https://doi.org/10.1037/0003-066x.61.4.305.

Diener, Ed, Eunkook M. Suh, Richard E. Lucas, and Heidi L. Smith. 1999. "Subjective Well-Being: Three Decades of Progress." *Psychological Bulletin* 125 (2): 276-302. https://doi.org/10.1037/0033-2909.125.2.276.

Dierickx, Ingemar, and Karel Cool. 1989. "Asset Stock Accumulation and Sustainability of Competitive Advantage." *Management Science* 35 (12): 1504-11. https://doi.org/10.1287/mnsc.35.12.1504.

Dillard, James Price, and Lijiang Shen. 2005. "On the Nature of Reactance and Its Role in Persuasive Health Communication." *Communication Monographs* 72 (2): 144-68. https://doi.org/10.1080/03637750500111815.

DiMaggio, Paul J. 1995. "Comments on "What Theory Is Not"." *Administrative Science Quarterly* 40 (3): 391-97. https://doi.org/10.2307/2393790.

DiMaggio, Paul J., and Walter W. Powell. 1983. "The Iron Cage Revisited: Institutional Isomorphism and Collective Rationality in Organizational Fields." *American Sociological Review* 48 (2): 147-60. https://doi.org/10.2307/2095101.

Dinner, Isaac, Eric J. Johnson, Daniel G. Goldstein, and Kaiya Liu. 2011. "Partitioning Default Effects: Why People Choose Not to Choose." *Journal of Experimental Psychology: Applied* 17 (4): 332-41.

DiPrete, Thomas A., and Gregory M. Eirich. 2006. "Cumulative Advantage as a Mechanism for Inequality: A Review of Theoretical and Empirical Developments." *Annual Review of Sociology* 32: 271-97.

Dishaw, Mark T., and Diane M. Strong. 1999. "Extending the Technology Acceptance Model with Task-Technology Fit Constructs." *Information and Management* 36 (1): 9-21. https://doi.org/10.1016/S0378-7206(98)00101-3.

Dixit, Avinash. 1980. "The Role of Investment in Entry-Deterrence." *The Economic Journal* 90 (357): 95-106. https://doi.org/10.2307/2231658.

Dixit, Avinash K., and Robert S. Pindyck. 1994. *Investment Under Uncertainty*. Princeton, NJ: Princeton University Press.

Donaldson, Lex. 2001. *The Contingency Theory of Organizations*. Thousand Oaks, CA: Sage.

Donaldson, Lex, and James H. Davis. 1991. "Stewardship Theory or Agency Theory: CEO Governance and Shareholder Returns." *Australian Journal of Management* 16 (1): 49-64. https://doi.org/10.1177/031289629101600103.

Donaldson, Thomas, and Lee E. Preston. 1995. "The Stakeholder Theory of the Corporation: Concepts, Evidence, and Implications." *Academy of Management Review* 20 (1): 65-91. https://doi.org/10.2307/258887.

Doney, Patricia M., and Joseph P. Cannon. 1997. "An Examination of the Nature of Trust in Buyer-Seller Relationships." *Journal of Marketing* 61 (2): 35-51. https://doi.org/10.1177/00222429

9706100203.
Donovan, Robert J., and John R. Rossiter. 1982. "Store Atmosphere: An Environmental Psychology Approach." *Journal of Retailing* 58 (1): 34-57.
Dranove, David, and Ginger Zhe Jin. 2010. "Quality Disclosure and Certification: Theory and Practice." *Journal of Economic Literature* 48 (4): 935-63. https://doi.org/10.1257/jel.48.4.935.
Drees, Johannes M., and Pursey P. M. A. R. Heugens. 2013. "Synthesizing and Extending Resource Dependence Theory: A Meta-Analysis." *Journal of Management* 39 (6): 1666-98. https://doi.org/10.1177/0149206312471391.
Dubin, Robert. 1978. *Theory Building*. Revised. New York: Free Press.
Duhigg, Charles. 2016. "What Google Learned from Its Quest to Build the Perfect Team." *The New York Times Magazine*, February.
Dulebohn, James H., William H. Bommer, Robert C. Liden, Robyn L. Brouer, and Gerald R. Ferris. 2012. "A Meta-Analysis of Antecedents and Consequences of Leader-Member Exchange: Integrating the Past with an Eye Toward the Future." *Journal of Management* 38 (6): 1715-59. https://doi.org/10.1177/0149206311415280.
Dunn, Elizabeth W., Lara B. Aknin, and Michael I. Norton. 2008. "Spending Money on Others Promotes Happiness." *Science* 319 (5870): 1687-88. https://doi.org/10.1126/science.1150952.
Dunning, John H. 1980. "Toward an Eclectic Theory of International Production: Some Empirical Tests." *Journal of International Business Studies* 11 (1): 9-31. https://doi.org/10.1057/palgrave.jibs.8490593.
Dunning, John H. 1988. "The Eclectic Paradigm of International Production: A Restatement and Some Possible Extensions." *Journal of International Business Studies* 19 (1): 1-31. https://doi.org/10.1057/palgrave.jibs.8490372.
Dutton, Jane E., and Janet M. Dukerich. 1991. "Keeping an Eye on the Mirror: Image and Identity in Organizational Adaptation." *Academy of Management Journal* 34 (3): 517-54. https://doi.org/10.2307/256405.
Dwyer, F. Robert, Paul H. Schurr, and Sejo Oh. 1987. "Developing Buyer-Seller Relationships." *Journal of Marketing* 51 (2): 11-27. https://doi.org/10.1177/002224298705100202.
Dyer, Jeffrey H., and Kentaro Nobeoka. 2000. "Creating and Managing a High-Performance Knowledge-Sharing Network: The Toyota Case." *Strategic Management Journal* 21 (3): 345-67.
Dyer, Jeffrey H., and Harbir Singh. 1998. "The Relational View: Cooperative Strategy and Sources of Interorganizational Competitive Advantage." *Academy of Management Review* 23 (4): 660-79. https://doi.org/10.5465/amr.1998.1255632.
Eagly, Alice H., and Shelly Chaiken. 1993. *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.
Edelman, Benjamin, Michael Ostrovsky, and Michael Schwarz. 2007. "Internet Advertising and the Generalized Second-Price Auction: Selling Billions of Dollars Worth of Keywords." *American Economic Review* 97 (1): 242-59. https://doi.org/10.1257/aer.97.1.242.
Edmans, Alex. 2011. "Does the Stock Market Fully Value Intangibles? Employee Satisfaction and Equity Prices." *Journal of Financial Economics* 101 (3): 621-40.
Edmans, Alex, Diego García, and Øyvind Norli. 2007. "Sports Sentiment and Stock Returns." *The Journal of Finance* 62 (4): 1967-98. https://doi.org/10.1111/j.1540-6261.2007.01262.x.
Edmondson, Amy. 1999. "Psychological Safety and Learning Behavior in Work Teams." *Administrative Science Quarterly* 44 (2): 350-83.
Edmondson, Amy C. 1996. "Learning from Mistakes Is Easier Said Than Done: Group and Organizational Influences on the Detection and Correction of Human Error." *Journal of Applied Behavioral Science* 32 (1): 5-28. https://doi.org/10.1177/0021886396321001.
Edmondson, Amy C. 2018. *The Fearless Organization: Creating Psychological Safety in the Workplace*

for Learning, Innovation, and Growth. Hoboken, NJ: Wiley.

Edmondson, Amy C., Richard M. Bohmer, and Gary P. Pisano. 2001. "Disrupted Routines: Team Learning and New Technology Implementation in Hospitals." *Administrative Science Quarterly* 46 (4): 685-716. https://doi.org/10.2307/3094828.

Edmondson, Amy C., and Zhike Lei. 2014. "Psychological Safety: The History, Renaissance, and Future of an Interpersonal Construct." *Annual Review of Organizational Psychology and Organizational Behavior* 1: 23-43. https://doi.org/10.1146/annurev-orgpsych-031413-091305.

Edmondson, Amy C., and Stacy E. McManus. 2007. "Methodological Fit in Management Field Research." *Academy of Management Review* 32 (4): 1246-64. https://doi.org/10.5465/amr.2007.26586086.

Edwards, Jeffrey R., and Mark E. Parry. 1993. "On the Use of Polynomial Regression Equations as an Alternative to Difference Scores in Organizational Research." *Academy of Management Journal* 36 (6): 1577-1613. https://doi.org/10.2307/256822.

Egan, Louisa C., Laurie R. Santos, and Paul Bloom. 2007. "The Origins of Cognitive Dissonance: Evidence from Children and Monkeys." *Psychological Science* 18 (11): 978-83. https://doi.org/10.1111/j.1467-9280.2007.02012.x.

Eggers, J. P., and Sarah Kaplan. 2009. "Cognition and Renewal: Comparing CEO and Organizational Effects on Incumbent Adaptation to Technical Change." *Organization Science* 20 (2): 461-77. https://doi.org/10.1287/orsc.1080.0401.

Ehrenberg, Andrew S. C., Gerald J. Goodhardt, and T. Patrick Barwise. 1990. "Double Jeopardy Revisited." *Journal of Marketing* 54 (3): 82-91. https://doi.org/10.1177/002224299005400307.

Ehrenberg, Ronald G., and Michael L. Bognanno. 1990. "Do Tournaments Have Incentive Effects?" *Journal of Political Economy* 98 (6): 1307-24. https://doi.org/10.1086/261736.

Ein-Gar, Danit, Baba Shiv, and Zakary L. Tormala. 2012. "When Blemishing Leads to Blossoming: The Positive Effect of Negative Information." *Journal of Consumer Research* 38 (5): 846-59. https://doi.org/10.1086/660807.

Eisenberger, Robert, Robin Huntington, Steven Hutchison, and Debora Sowa. 1986. "Perceived Organizational Support." *Journal of Applied Psychology* 71 (3): 500-507. https://doi.org/10.1037/0021-9010.71.3.500.

Eisenhardt, Kathleen M. 1989a. "Agency Theory: An Assessment and Review." *Academy of Management Review* 14 (1): 57-74. https://doi.org/10.5465/amr.1989.4279003.

Eisenhardt, Kathleen M. 1989b. "Building Theories from Case Study Research." *Academy of Management Review* 14 (4): 532-50. https://doi.org/10.5465/amr.1989.4308385.

Eisenhardt, Kathleen M., and Jeffrey A. Martin. 2000. "Dynamic Capabilities: What Are They?" *Strategic Management Journal* 21 (10-11): 1105-21.

Ekman, Paul. 1992. "An Argument for Basic Emotions." *Cognition and Emotion* 6 (3-4): 169-200.

Elliot, Andrew J., and Patricia G. Devine. 1994. "On the Motivational Nature of Cognitive Dissonance: Dissonance as Psychological Discomfort." *Journal of Personality and Social Psychology* 67 (3): 382-94. https://doi.org/10.1037/0022-3514.67.3.382.

Ellison, Glenn, and Sara Fisher Ellison. 2009. "Search, Obfuscation, and Price Elasticities on the Internet." *Econometrica* 77 (2): 427-52. https://doi.org/10.3982/ECTA5708.

Ellsberg, Daniel. 1961. "Risk, Ambiguity, and the Savage Axioms." *The Quarterly Journal of Economics* 75 (4): 643-69. https://doi.org/10.2307/1884324.

Elsbach, Kimberly D., and Roderick M. Kramer. 1996. "Members' Responses to Organizational Identity Threats: Encountering and Countering the Business Week Rankings." *Administrative Science Quarterly* 41 (3): 442-76. https://doi.org/10.2307/2393938.

Elwert, Felix, and Christopher Winship. 2014. "Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable." *Annual Review of Sociology* 40: 31-53. https://doi.org/10.114

6/annurev-soc-071913-043455.
Emerson, Richard M. 1962. "Power-Dependence Relations." *American Sociological Review* 27 (1): 31-41. https://doi.org/10.2307/2089716.
Emerson, Richard M. 1976. "Social Exchange Theory." *Annual Review of Sociology* 2: 335-62. https://doi.org/10.1146/annurev.so.02.080176.002003.
Englis, Basil G., and Michael R. Solomon. 1995. "To Be and Not to Be: Lifestyle Imagery, Reference Groups, and the Clustering of America." *Journal of Advertising* 24 (1): 13-28. https://doi.org/10.1080/00913367.1995.10673465.
Engström, Per, Katarina Nordblom, Henry Ohlsson, and Annika Persson. 2015. "Tax Compliance and Loss Aversion." *American Economic Journal: Economic Policy* 7 (4): 132-64. https://doi.org/10.1257/pol.20130134.
Entman, Robert M. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43 (4): 51-58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x.
Epley, Nicholas, and Thomas Gilovich. 2006. "The Anchoring-and-Adjustment Heuristic: Why the Adjustments Are Insufficient." *Psychological Science* 17 (4): 311-18. https://doi.org/10.1111/j.1467-9280.2006.01704.x.
Erdem, Tülin, and Joffre Swait. 1998. "Brand Equity as a Signaling Phenomenon." *Journal of Consumer Psychology* 7 (2): 131-57. https://doi.org/10.1207/s15327663jcp0702_02.
Ericson, Keith M. Marzilli, and Andreas Fuster. 2011. "Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments." *The Quarterly Journal of Economics* 126 (4): 1879-1907. https://doi.org/10.1093/qje/qjr034.
Ert, Eyal, and Ido Erev. 2013. "On the Descriptive Value of Loss Aversion in Decisions Under Risk: Six Clarifications." *Judgment and Decision Making* 8 (3): 214-35. https://doi.org/10.1017/S1930297500005945.
Escalas, Jennifer Edson, and James R. Bettman. 2003. "You Are What They Eat: The Influence of Reference Groups on Consumers' Connections to Brands." *Journal of Consumer Psychology* 13 (3): 339-48. https://doi.org/10.1207/S15327663JCP1303_14.
Esser, James K. 1998. "Alive and Well After 25 Years: A Review of Groupthink Research." *Organizational Behavior and Human Decision Processes* 73 (2-3): 116-41.
Evans, Jonathan St. B. T., and Keith E. Stanovich. 2013. "Dual-Process Theories of Higher Cognition: Advancing the Debate." *Perspectives on Psychological Science* 8 (3): 223-41. https://doi.org/10.1177/1745691612460685.
Evans, William N., and Ioannis N. Kessides. 1994. "Living by the 'Golden Rule': Multimarket Contact in the U.S. Airline Industry." *The Quarterly Journal of Economics* 109 (2): 341-66. https://doi.org/10.2307/2118466.
Fader, Peter S., Bruce G. S. Hardie, and Ka Lok Lee. 2005. "'Counting Your Customers' the Easy Way: An Alternative to the Pareto/NBD Model." *Marketing Science* 24 (2): 275-84. https://doi.org/10.1287/mksc.1040.0098.
Fama, Eugene F. 1970. "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance* 25 (2): 383-417. https://doi.org/10.2307/2325486.
Fama, Eugene F. 1980. "Agency Problems and the Theory of the Firm." *Journal of Political Economy* 88 (2): 288-307. https://doi.org/10.1086/260866.
Fama, Eugene F. 1998. "Market Efficiency, Long-Term Returns, and Behavioral Finance." *Journal of Financial Economics* 49 (3): 283-306. https://doi.org/10.1016/S0304-405X(98)00026-9.
Fama, Eugene F., and Kenneth R. French. 1992. "The Cross-Section of Expected Stock Returns." *The Journal of Finance* 47 (2): 427-65. https://doi.org/10.1111/j.1540-6261.1992.tb04398.x.
Fama, Eugene F., and Kenneth R. French. 1993. "Common Risk Factors in the Returns on Stocks and Bonds." *Journal of Financial Economics* 33 (1): 3-56. https://doi.org/10.1016/0304-

405X(93)90023-5.
Fama, Eugene F., and Kenneth R. French. 2004. "The Capital Asset Pricing Model: Theory and Evidence." *Journal of Economic Perspectives* 18 (3): 25-46. https://doi.org/10.1257/0895330042162430.
Fama, Eugene F., and Kenneth R. French. 2015. "A Five-Factor Asset Pricing Model." *Journal of Financial Economics* 116 (1): 1-22. https://doi.org/10.1016/j.jfineco.2014.10.010.
Fama, Eugene F., and Michael C. Jensen. 1983. "Separation of Ownership and Control." *The Journal of Law and Economics* 26 (2): 301-25. https://doi.org/10.1086/467037.
Farber, Henry S. 2008. "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers." *American Economic Review* 98 (3): 1069-82. https://doi.org/10.1257/aer.98.3.1069.
Farber, Henry S. 2015. "Why You Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers." *The Quarterly Journal of Economics* 130 (4): 1975-2026. https://doi.org/10.1093/qje/qjv026.
Farrell, Joseph, and Garth Saloner. 1985. "Standardization, Compatibility, and Innovation." *The RAND Journal of Economics* 16 (1): 70-83. https://doi.org/10.2307/2555589.
Fazio, Russell H., Mark P. Zanna, and Joel Cooper. 1977. "Dissonance and Self-Perception: An Integrative View of Each Theory's Proper Domain of Application." *Journal of Experimental Social Psychology* 13 (5): 464-79. https://doi.org/10.1016/0022-1031(77)90031-2.
Fehr, Ernst, and Simon Gächter. 2000a. "Cooperation and Punishment in Public Goods Experiments." *American Economic Review* 90 (4): 980-94. https://doi.org/10.1257/aer.90.4.980.
Fehr, Ernst, and Simon Gächter. 2000b. "Fairness and Retaliation: The Economics of Reciprocity." *Journal of Economic Perspectives* 14 (3): 159-82. https://doi.org/10.1257/jep.14.3.159.
Fehr, Ernst, and Lorenz Goette. 2007. "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment." *American Economic Review* 97 (1): 298-317. https://doi.org/10.1257/aer.97.1.298.
Fehr, Ernst, Georg Kirchsteiger, and Arno Riedl. 1993. "Does Fairness Prevent Market Clearing? An Experimental Investigation." *The Quarterly Journal of Economics* 108 (2): 437-59. https://doi.org/10.2307/2118338.
Fehr, Ernst, and Klaus M. Schmidt. 1999. "A Theory of Fairness, Competition, and Cooperation." *The Quarterly Journal of Economics* 114 (3): 817-68. https://doi.org/10.1162/003355399556151.
Feltovich, Nick, Richmond Harbaugh, and Ted To. 2002. "Too Cool for School? Signalling and Countersignalling." *The RAND Journal of Economics* 33 (4): 630-49. https://doi.org/10.2307/3087478.
Fernandez, Roberto M., Emilio J. Castilla, and Paul Moore. 2000. "Social Capital at Work: Networks and Employment at a Phone Center." *American Journal of Sociology* 105 (5): 1288-1356. https://doi.org/10.1086/210432.
Fershtman, Chaim, and Uri Gneezy. 2011. "THE TRADEOFF BETWEEN PERFORMANCE AND QUITTING IN HIGH POWER TOURNAMENTS." *Journal of the European Economic Association* 9 (2): 318-36. https://doi.org/10.1111/j.1542-4774.2010.01012.x.
Festinger, Leon. 1954. "A Theory of Social Comparison Processes." *Human Relations* 7 (2): 117-40. https://doi.org/10.1177/001872675400700202.
Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford, CA: Stanford University Press.
Festinger, Leon, and James M. Carlsmith. 1959. "Cognitive Consequences of Forced Compliance." *Journal of Abnormal and Social Psychology* 58 (2): 203-10. https://doi.org/10.1037/h0041593.
Fichman, Robert G., and Chris F. Kemerer. 1999. "The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps." *Information Systems Research* 10 (3): 255-75. https://doi.org/10.1287/isre.10.3.255.

Finke, Ronald A., Thomas B. Ward, and Steven M. Smith. 1992. *Creative Cognition: Theory, Research, and Applications*. Cambridge, MA: MIT Press.

Finke, Ronald A., Thomas B. Ward, and Steven M. Smith. 1996. *Creative Cognition: Theory, Research, and Applications*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/7722.001.0001.

Finkelstein, Sydney, and Donald C. Hambrick. 1990. "Top-Management-Team Tenure and Organizational Outcomes: The Moderating Role of Managerial Discretion." *Administrative Science Quarterly* 35 (3): 484-503. https://doi.org/10.2307/2393314.

Finkelstein, Sydney, Donald C. Hambrick, and Jr. Cannella Albert A. 2009. *Strategic Leadership: Theory and Research on Executives, Top Management Teams, and Boards*. New York: Oxford University Press.

Fiol, C. Marlene, and Marjorie A. Lyles. 1985. "Organizational Learning." *Academy of Management Review* 10 (4): 803-13. https://doi.org/10.2307/258048.

Fischbacher, Urs, Simon Gächter, and Ernst Fehr. 2001. "Are People Conditionally Cooperative? Evidence from a Public Goods Experiment." *Economics Letters* 71 (3): 397-404. https://doi.org/10.1016/S0165-1765(01)00394-9.

Fischer, Peter, Joachim I. Krueger, Tobias Greitemeyer, Claudia Vogrincic, Andreas Kastenmüller, Dieter Frey, Moritz Heene, Magdalena Wicher, and Martina Kainbacher. 2011. "The Bystander-Effect: A Meta-Analytic Review on Bystander Intervention in Dangerous and Non-Dangerous Emergencies." *Psychological Bulletin* 137 (4): 517-37. https://doi.org/10.1037/a0023304.

Fishbein, Martin, and Icek Ajzen. 1975. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley.

Fitzsimons, Gavan J., and Donald R. Lehmann. 2004. "Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses." *Marketing Science* 23 (1): 82-94. https://doi.org/10.1287/mksc.1030.0033.

Fleming, Lee. 2001. "Recombinant Uncertainty in Technological Search." *Management Science* 47 (1): 117-32.

Folkes, Valerie S. 1984. "Consumer Reactions to Product Failure: An Attributional Approach." *Journal of Consumer Research* 10 (4): 398-409. https://doi.org/10.1086/208978.

Folkes, Valerie S. 1988. "Recent Attribution Research in Consumer Behavior: A Review and New Directions." *Journal of Consumer Research* 14 (4): 548-65. https://doi.org/10.1086/209135.

Folkes, Valerie S., Susan Koletsky, and John L. Graham. 1987. "A Field Study of Causal Inferences and Consumer Reaction: The View from the Airport." *Journal of Consumer Research* 13 (4): 534-39. https://doi.org/10.1086/209086.

Fornell, Claes, Michael D. Johnson, Eugene W. Anderson, Jaesung Cha, and Barbara Everitt Bryant. 1996. "The American Customer Satisfaction Index: Nature, Purpose, and Findings." *Journal of Marketing* 60 (4): 7-18. https://doi.org/10.1177/002224299606000403.

Fornell, Claes, Sunil Mithas, Forrest V. Morgeson, and M. S. Krishnan. 2006. "Customer Satisfaction and Stock Prices: High Returns, Low Risk." *Journal of Marketing* 70 (1): 3-14. https://doi.org/10.1509/jmkg.70.1.003.qxd.

Forrester, Jay W. 1961. *Industrial Dynamics*. Cambridge, MA: MIT Press.

Forsgren, Mats. 2002. "The Concept of Learning in the Uppsala Internationalization Process Model: A Critical Review." *International Business Review* 11 (3): 257-77. https://doi.org/10.1016/S0969-5931(01)00060-9.

Fournier, Susan. 1998. "Consumers and Their Brands: Developing Relationship Theory in Consumer Research." *Journal of Consumer Research* 24 (4): 343-73. https://doi.org/10.1086/209515.

Frank, Murray Z., and Vidhan K. Goyal. 2003. "Testing the Pecking Order Theory of Capital Struc-

ture." *Journal of Financial Economics* 67 (2): 217-48. https://doi.org/10.1016/S0304-405X(02)00252-0.

Franke, Nikolaus, Martin Schreier, and Ulrike Kaiser. 2010. "The 'i Designed It Myself' Effect in Mass Customization." *Management Science* 56 (1): 125-40. https://doi.org/10.1287/mnsc.1090.1077.

Frazier, M. Lance, Stav Fainshmidt, Ryan L. Klinger, Amir Pezeshkan, and Veselina Vracheva. 2017. "Psychological Safety: A Meta-Analytic Review and Extension." *Personnel Psychology* 70 (1): 113-65. https://doi.org/10.1111/peps.12183.

Frederick, Shane. 2005. "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives* 19 (4): 25-42. https://doi.org/10.1257/089533005775196732.

Frederick, Shane, Leonard Lee, and Ernest Baskin. 2014. "The Limits of Attraction." *Journal of Marketing Research* 51 (4): 487-507. https://doi.org/10.1509/jmr.12.0061.

Frederick, Shane, and George Loewenstein. 1999. "Hedonic Adaptation." In *Well-Being: The Foundations of Hedonic Psychology*, edited by Daniel Kahneman, Ed Diener, and Norbert Schwarz, 302-29. New York: Russell Sage Foundation.

Frederick, Shane, George Loewenstein, and Ted O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature* 40 (2): 351-401. https://doi.org/10.1257/jel.40.2.351.

Fredrickson, Barbara L. 1998. "What Good Are Positive Emotions?" *Review of General Psychology* 2 (3): 300-319. https://doi.org/10.1037/1089-2680.2.3.300.

Fredrickson, Barbara L. 2001. "The Role of Positive Emotions in Positive Psychology: The Broaden-and-Build Theory of Positive Emotions." *American Psychologist* 56 (3): 218-26. https://doi.org/10.1037/0003-066X.56.3.218.

Fredrickson, Barbara L. 2013. "Updated Thinking on Positivity Ratios." *American Psychologist* 68 (9): 814-22.

Fredrickson, Barbara L., and Christine Branigan. 2005. "Positive Emotions Broaden the Scope of Attention and Thought-Action Repertoires." *Cognition and Emotion* 19 (3): 313-32. https://doi.org/10.1080/02699930441000238.

Fredrickson, Barbara L., Michael A. Cohn, Kimberly A. Coffey, Jolynn Pek, and Sandra M. Finkel. 2008. "Open Hearts Build Lives: Positive Emotions, Induced Through Loving-Kindness Meditation, Build Consequential Personal Resources." *Journal of Personality and Social Psychology* 95 (5): 1045-62.

Fredrickson, Barbara L., and Daniel Kahneman. 1993. "Duration Neglect in Retrospective Evaluations of Affective Episodes." *Journal of Personality and Social Psychology* 65 (1): 45-55. https://doi.org/10.1037/0022-3514.65.1.45.

Fredrickson, Barbara L., and Marcial F. Losada. 2005. "Positive Affect and the Complex Dynamics of Human Flourishing." *American Psychologist* 60 (7): 678-86. https://doi.org/10.1037/0003-066X.60.7.678.

Fredrickson, Barbara L., Roberta A. Mancuso, Christine Branigan, and Michele M. Tugade. 2000. "The Undoing Effect of Positive Emotions." *Motivation and Emotion* 24 (4): 237-58.

Freedman, Jonathan L., and Scott C. Fraser. 1966. "Compliance Without Pressure: The Foot-in-the-Door Technique." *Journal of Personality and Social Psychology* 4 (2): 195-202. https://doi.org/10.1037/h0023552.

Freeland, Robert F. 2000. "Creating Holdup Through Vertical Integration: Fisher Body Revisited." *The Journal of Law and Economics* 43 (1): 33-66. https://doi.org/10.1086/467447.

Freeman, John, Glenn R. Carroll, and Michael T. Hannan. 1983. "The Liability of Newness: Age Dependence in Organizational Death Rates." *American Sociological Review* 48 (5): 692-710. https://doi.org/10.2307/2094928.

Freeman, Linton C. 1977. "A Set of Measures of Centrality Based on Betweenness." *Sociometry* 40 (1): 35-41. https://doi.org/10.2307/3033543.

Freeman, R. Edward. 1984. *Strategic Management: A Stakeholder Approach*. Boston: Pitman.

French, Kenneth R. 2008. "Presidential Address: The Cost of Active Investing." *The Journal of Finance* 63 (4): 1537-73. https://doi.org/10.1111/j.1540-6261.2008.01368.x.

Frey, Bruno S., and Reto Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys* 15 (5): 589-611.

Fried, Yitzhak, and Gerald R. Ferris. 1987. "The Validity of the Job Characteristics Model: A Review and Meta-Analysis." *Personnel Psychology* 40 (2): 287-322. https://doi.org/10.1111/j.1744-6570.1987.tb00605.x.

Friedman, Milton. 1970. "The Social Responsibility of Business Is to Increase Its Profits." *The New York Times Magazine*, September.

Friestad, Marian, and Peter Wright. 1994. "The Persuasion Knowledge Model: How People Cope with Persuasion Attempts." *Journal of Consumer Research* 21 (1): 1-31. https://doi.org/10.1086/209380.

Froot, Kenneth A., and Emil M. Dabora. 1999. "How Are Stock Prices Affected by the Location of Trade?" *Journal of Financial Economics* 53 (2): 189-216. https://doi.org/10.1016/S0304-405X(99)00020-3.

Fryer, Roland G., Steven D. Levitt, John List, and Sally Sadoff. 2022. "Enhancing the Efficacy of Teacher Incentives Through Framing: A Field Experiment." *American Economic Journal: Economic Policy* 14 (4): 269-99. https://doi.org/10.1257/pol.20190287.

Fuchs, Christoph, Emanuela Prandelli, and Martin Schreier. 2010. "The Psychological Effects of Empowerment Strategies on Consumers' Product Demand." *Journal of Marketing* 74 (1): 65-79. https://doi.org/10.1509/jmkg.74.1.65.

Fujita, Kentaro, Yaacov Trope, Nira Liberman, and Maya Levin-Sagi. 2006. "Construal Levels and Self-Control." *Journal of Personality and Social Psychology* 90 (3): 351-67. https://doi.org/10.1037/0022-3514.90.3.351.

Furnham, Adrian, and Hua Chu Boo. 2011. "A Literature Review of the Anchoring Effect." *The Journal of Socio-Economics* 40 (1): 35-42. https://doi.org/10.1016/j.socec.2010.10.008.

Gable, Philip A., and Eddie Harmon-Jones. 2008. "Approach-Motivated Positive Affect Reduces Breadth of Attention." *Psychological Science* 19 (5): 476-82.

Gächter, Simon, Eric J. Johnson, and Andreas Herrmann. 2022. "Individual-Level Loss Aversion in Riskless and Risky Choices." *Theory and Decision* 92 (3-4): 599-624. https://doi.org/10.1007/s11238-021-09839-8.

Gagné, Marylène, and Edward L. Deci. 2005. "Self-Determination Theory and Work Motivation." *Journal of Organizational Behavior* 26 (4): 331-62.

Gajendran, Ravi S., and David A. Harrison. 2007. "The Good, the Bad, and the Unknown about Telecommuting: Meta-Analysis of Psychological Mediators and Individual Consequences." *Journal of Applied Psychology* 92 (6): 1524-41. https://doi.org/10.1037/0021-9010.92.6.1524.

Gal, David. 2006. "A Psychological Law of Inertia and the Illusion of Loss Aversion." *Judgment and Decision Making* 1 (1): 23-32. https://doi.org/10.1017/S1930297500000322.

Gal, David, and Derek D. Rucker. 2018a. "Loss Aversion, Intellectual Inertia, and a Call for a More Contrarian Science: A Reply to Simonson and Kivetz and Higgins and Liberman." *Journal of Consumer Psychology* 28 (3): 533-39. https://doi.org/10.1002/jcpy.1044.

Gal, David, and Derek D. Rucker. 2018b. "The Loss of Loss Aversion: Will It Loom Larger Than Its Gain?" *Journal of Consumer Psychology* 28 (3): 497-516.

Galbraith, Jay R. 1973. *Designing Complex Organizations*. Reading, MA: Addison-Wesley.

Galinsky, Adam D., and Thomas Mussweiler. 2001. "First Offers as Anchors: The Role of Perspective-

Taking and Negotiator Focus." *Journal of Personality and Social Psychology* 81 (4): 657-69. https://doi.org/10.1037/0022-3514.81.4.657.

Galton, Francis. 1907. "Vox Populi." *Nature* 75 (1949): 450-51. https://doi.org/10.1038/075450a0.

Gans, Joshua S., and Scott Stern. 2003. "The Product Market and the Market for 'Ideas': Commercialization Strategies for Technology Entrepreneurs." *Research Policy* 32 (2): 333-50. https://doi.org/10.1016/S0048-7333(02)00103-8.

Ganzach, Yoav, and Nili Karsahi. 1995. "Message Framing and Buying Behavior: A Field Experiment." *Journal of Business Research* 32 (1): 11-17. https://doi.org/10.1016/0148-2963(93)00038-3.

Garbarino, Ellen, and Mark S. Johnson. 1999. "The Different Roles of Satisfaction, Trust, and Commitment in Customer Relationships." *Journal of Marketing* 63 (2): 70-87. https://doi.org/10.1177/002224299906300205.

Garud, Raghu, and Peter Karnøe. 2003. "Bricolage Versus Breakthrough: Distributed and Embedded Agency in Technology Entrepreneurship." *Research Policy* 32 (2): 277-300. https://doi.org/10.1016/S0048-7333(02)00100-2.

Gavetti, Giovanni, Henrich R. Greve, Daniel A. Levinthal, and William Ocasio. 2012. "The Behavioral Theory of the Firm: Assessment and Prospects." *Academy of Management Annals* 6 (1): 1-40. https://doi.org/10.5465/19416520.2012.656841.

Gelfand, Michele J. 2018. *Rule Makers, Rule Breakers: How Tight and Loose Cultures Wire Our World*. New York: Scribner.

Gelfand, Michele J., Jana L. Raver, Lisa Nishii, Lisa M. Leslie, Janetta Lun, Beng Chong Lim, et al. 2011. "Differences Between Tight and Loose Cultures: A 33-Nation Study." *Science* 332 (6033): 1100-1104. https://doi.org/10.1126/science.1197754.

Gelman, Andrew, and Eric Loken. 2014. "The Statistical Crisis in Science." *American Scientist* 102 (6): 460-65. https://doi.org/10.1511/2014.111.460.

Genesove, David, and Christopher Mayer. 2001. "Loss Aversion and Seller Behavior: Evidence from the Housing Market." *The Quarterly Journal of Economics* 116 (4): 1233-60. https://doi.org/10.1162/003355301753265561.

Gerber, Jonathan P., Ladd Wheeler, and Jerry Suls. 2018. "A Social Comparison Theory Meta-Analysis 60+ Years On." *Psychological Bulletin* 144 (2): 177-97. https://doi.org/10.1037/bul0000127.

Gerstner, Charlotte R., and David V. Day. 1997. "Meta-Analytic Review of Leader-Member Exchange Theory: Correlates and Construct Issues." *Journal of Applied Psychology* 82 (6): 827-44. https://doi.org/10.1037/0021-9010.82.6.827.

Gerstner, Jr., Louis V. 2002. *Who Says Elephants Can't Dance? Inside IBM's Historic Turnaround*. New York: HarperBusiness.

Geuens, Maggie, Bert Weijters, and Kristof De Wulf. 2009. "A New Measure of Brand Personality." *International Journal of Research in Marketing* 26 (2): 97-107. https://doi.org/10.1016/j.ijresmar.2008.12.002.

Ghemawat, Pankaj. 1991. *Commitment: The Dynamic of Strategy*. New York: Free Press.

Ghemawat, Pankaj. 2001. "Distance Still Matters: The Hard Reality of Global Expansion." *Harvard Business Review* 79 (8): 137-47.

Ghoshal, Sumantra. 2005. "Bad Management Theories Are Destroying Good Management Practices." *Academy of Management Learning & Education* 4 (1): 75-91. https://doi.org/10.5465/amle.2005.16132558.

Ghoshal, Sumantra, and Peter Moran. 1996. "Bad for Practice: A Critique of the Transaction Cost Theory." *Academy of Management Review* 21 (1): 13-47. https://doi.org/10.5465/amr.1996.9602161563.

Gibson, Cristina B., and Julian Birkinshaw. 2004. "The Antecedents, Consequences, and Mediating Role of Organizational Ambidexterity." *Academy of Management Journal* 47 (2): 209-26.

Gigerenzer, Gerd, and Wolfgang Gaissmaier. 2011. "Heuristic Decision Making." *Annual Review of Psychology* 62: 451-82. https://doi.org/10.1146/annurev-psych-120709-145346.

Gigerenzer, Gerd, and Daniel G. Goldstein. 1996. "Reasoning the Fast and Frugal Way: Models of Bounded Rationality." *Psychological Review* 103 (4): 650-69. https://doi.org/10.1037/0033-295X.103.4.650.

Gilbert, Daniel T., Gary King, Stephen Pettigrew, and Timothy D. Wilson. 2016. "Comment on "Estimating the Reproducibility of Psychological Science"." *Science* 351 (6277): 1037. https://doi.org/10.1126/science.aad7243.

Gilbert, Daniel T., and Patrick S. Malone. 1995. "The Correspondence Bias." *Psychological Bulletin* 117 (1): 21-38. https://doi.org/10.1037/0033-2909.117.1.21.

Gilbert, Daniel T., Elizabeth C. Pinel, Timothy D. Wilson, Stephen J. Blumberg, and Thalia P. Wheatley. 1998. "Immune Neglect: A Source of Durability Bias in Affective Forecasting." *Journal of Personality and Social Psychology* 75 (3): 617-38. https://doi.org/10.1037/0022-3514.75.3.617.

Gill, David, and Victoria Prowse. 2012. "A Structural Analysis of Disappointment Aversion in a Real Effort Competition." *American Economic Review* 102 (1): 469-503. https://doi.org/10.1257/aer.102.1.469.

Gilovich, Thomas, Amit Kumar, and Lily Jampol. 2015. "A Wonderful Life: Experiential Consumption and the Pursuit of Happiness." *Journal of Consumer Psychology* 25 (1): 152-65.

Gimeno, Javier. 1999. "Reciprocal Threats in Multimarket Rivalry: Staking Out 'Spheres of Influence' in the U.S. Airline Industry." *Strategic Management Journal* 20 (2): 101-28.

Gimeno, Javier, and Carolyn Y. Woo. 1996. "Hypercompetition in a Multimarket Environment: The Role of Strategic Similarity and Multimarket Contact in Competitive de-Escalation." *Organization Science* 7 (3): 322-41. https://doi.org/10.1287/orsc.7.3.322.

Giné, Xavier, Dean Karlan, and Jonathan Zinman. 2010. "Put Your Money Where Your Butt Is: A Commitment Contract for Smoking Cessation." *American Economic Journal: Applied Economics* 2 (4): 213-35. https://doi.org/10.1257/app.2.4.213.

Gioia, Dennis A., and Kumar Chittipeddi. 1991. "Sensemaking and Sensegiving in Strategic Change Initiation." *Strategic Management Journal* 12 (6): 433-48. https://doi.org/10.1002/smj.4250120604.

Gioia, Dennis A., Majken Schultz, and Kevin G. Corley. 2000. "Organizational Identity, Image, and Adaptive Instability." *Academy of Management Review* 25 (1): 63-81. https://doi.org/10.2307/259263.

Gladwell, Malcolm. 2008. *Outliers: The Story of Success*. New York: Little, Brown and Company.

Glaeser, Edward L., Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2004. "Do Institutions Cause Growth?" *Journal of Economic Growth* 9 (3): 271-303.

Glanzer, Murray, and Anita R. Cunitz. 1966. "Two Storage Mechanisms in Free Recall." *Journal of Verbal Learning and Verbal Behavior* 5 (4): 351-60.

Glaser, Barney G., and Anselm L. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.

Glynn, Carroll J., Andrew F. Hayes, and James Shanahan. 1997. "Perceived Support for One's Opinions and Willingness to Speak Out: A Meta-Analysis of Survey Studies on the "Spiral of Silence"." *Public Opinion Quarterly* 61 (3): 452-63. https://doi.org/10.1086/297808.

Gneezy, Uri, Lorenz Goette, Charles Sprenger, and Florian Zimmermann. 2017. "The Limits of Expectations-Based Reference Dependence." *Journal of the European Economic Association* 15 (4): 861-76. https://doi.org/10.1093/jeea/jvw020.

Gneezy, Uri, and John A. List. 2006. "Putting Behavioral Economics to Work: Testing for Gift Ex-

change in Labor Markets Using Field Experiments." *Econometrica* 74 (5): 1365-84. https://doi.org/10.1111/j.1468-0262.2006.00707.x.

Gneezy, Uri, and Jan Potters. 1997. "An Experiment on Risk Taking and Evaluation Periods." *The Quarterly Journal of Economics* 112 (2): 631-45. https://doi.org/10.1162/003355397555217.

Gneezy, Uri, and Aldo Rustichini. 2000a. "A Fine Is a Price." *The Journal of Legal Studies* 29 (1): 1-17.

Gneezy, Uri, and Aldo Rustichini. 2000b. "Pay Enough or Don't Pay at All." *The Quarterly Journal of Economics* 115 (3): 791-810.

Godes, David, and Dina Mayzlin. 2004. "Using Online Conversations to Study Word-of-Mouth Communication." *Marketing Science* 23 (4): 545-60. https://doi.org/10.1287/mksc.1040.0071.

Goldenberg, Jacob, Barak Libai, and Eitan Muller. 2002. "Riding the Saddle: How Cross-Market Communications Can Create a Major Slump in Sales." *Journal of Marketing* 66 (2): 1-16. https://doi.org/10.1509/jmkg.66.2.1.18472.

Golder, Peter N., and Gerard J. Tellis. 1993. "Pioneer Advantage: Marketing Logic or Marketing Legend?" *Journal of Marketing Research* 30 (2): 158-70. https://doi.org/10.1177/002224379303000203.

Golder, Peter N., and Gerard J. Tellis. 1997. "Will It Ever Fly? Modeling the Takeoff of Really New Consumer Durables." *Marketing Science* 16 (3): 256-70. https://doi.org/10.1287/mksc.16.3.256.

Golder, Peter N., and Gerard J. Tellis. 2004. "Growing, Growing, Gone: Cascades, Diffusion, and Turning Points in the Product Life Cycle." *Marketing Science* 23 (2): 207-18. https://doi.org/10.1287/mksc.1040.0057.

Goldstein, Daniel G., and Gerd Gigerenzer. 2002. "Models of Ecological Rationality: The Recognition Heuristic." *Psychological Review* 109 (1): 75-90. https://doi.org/10.1037/0033-295X.109.1.75.

Goldstein, Noah J., Robert B. Cialdini, and Vladas Griskevicius. 2008. "A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels." *Journal of Consumer Research* 35 (3): 472-82. https://doi.org/10.1086/586910.

Gonzales, Marti Hope, Elliot Aronson, and Mark A. Costanzo. 1988. "Using Social Cognition and Persuasion to Promote Energy Conservation: A Quasi-Experiment." *Journal of Applied Social Psychology* 18 (12): 1049-66. https://doi.org/10.1111/j.1559-1816.1988.tb01192.x.

Gonzalez, Richard, and George Wu. 1999. "On the Shape of the Probability Weighting Function." *Cognitive Psychology* 38 (1): 129-66. https://doi.org/10.1006/cogp.1998.0710.

Goodhue, Dale L. 1995. "Understanding User Evaluations of Information Systems." *Management Science* 41 (12): 1827-44. https://doi.org/10.1287/mnsc.41.12.1827.

Goodhue, Dale L., and Ronald L. Thompson. 1995. "Task-Technology Fit and Individual Performance." *MIS Quarterly* 19 (2): 213-36. https://doi.org/10.2307/249689.

Gouldner, Alvin W. 1960. "The Norm of Reciprocity: A Preliminary Statement." *American Sociological Review* 25 (2): 161-78. https://doi.org/10.2307/2092623.

Gourlay, Stephen. 2006. "Conceptualizing Knowledge Creation: A Critique of Nonaka's Theory." *Journal of Management Studies* 43 (7): 1415-36. https://doi.org/10.1111/j.1467-6486.2006.00637.x.

Gourville, John T., and Dilip Soman. 1998. "Payment Depreciation: The Behavioral Effects of Temporally Separating Payments from Consumption." *Journal of Consumer Research* 25 (2): 160-74.

Graen, George B., and Mary Uhl-Bien. 1995. "Relationship-Based Approach to Leadership: Development of Leader-Member Exchange (LMX) Theory of Leadership over 25 Years: Applying a Multi-Level Multi-Domain Perspective." *The Leadership Quarterly* 6 (2): 219-47. https://doi.org/10.1016/1048-9843(95)90036-5.

Graham, John R., and Campbell R. Harvey. 2001. "The Theory and Practice of Corporate Finance:

Evidence from the Field." *Journal of Financial Economics* 60 (2-3): 187-243. https://doi.org/10.1016/S0304-405X(01)00044-7.

Graham, John R., Campbell R. Harvey, and Shiva Rajgopal. 2005. "The Economic Implications of Corporate Financial Reporting." *Journal of Accounting and Economics* 40 (1-3): 3-73. https://doi.org/10.1016/j.jacceco.2005.01.002.

Grammer, Karl, and Randy Thornhill. 1994. "Human (Homo Sapiens) Facial Attractiveness and Sexual Selection: The Role of Symmetry and Averageness." *Journal of Comparative Psychology* 108 (3): 233-42. https://doi.org/10.1037/0735-7036.108.3.233.

Granovetter, Mark. 1978. "Threshold Models of Collective Behavior." *American Journal of Sociology* 83 (6): 1420-43. https://doi.org/10.1086/226707.

Granovetter, Mark. 1985. "Economic Action and Social Structure: The Problem of Embeddedness." *American Journal of Sociology* 91 (3): 481-510. https://doi.org/10.1086/228311.

Granovetter, Mark S. 1973. "The Strength of Weak Ties." *American Journal of Sociology* 78 (6): 1360-80. https://doi.org/10.1086/225469.

Granovetter, Mark S. 1974. *Getting a Job: A Study of Contacts and Careers*. Cambridge, MA: Harvard University Press.

Grant, Adam M. 2008. "The Significance of Task Significance: Job Performance Effects, Relational Mechanisms, and Boundary Conditions." *Journal of Applied Psychology* 93 (1): 108-24. https://doi.org/10.1037/0021-9010.93.1.108.

Grant, Robert M. 1996. "Toward a Knowledge-Based Theory of the Firm." *Strategic Management Journal* 17 (S2): 109-22. https://doi.org/10.1002/smj.4250171110.

Greenberg, Jerald. 1990. "Employee Theft as a Reaction to Underpayment Inequity: The Hidden Cost of Pay Cuts." *Journal of Applied Psychology* 75 (5): 561-68. https://doi.org/10.1037/0021-9010.75.5.561.

Gregor, Shirley. 2006. "The Nature of Theory in Information Systems." *MIS Quarterly* 30 (3): 611-42. https://doi.org/10.2307/25148742.

Greif, Avner. 1993. "Contract Enforceability and Economic Institutions in Early Trade: The Maghribi Traders' Coalition." *American Economic Review* 83 (3): 525-48.

Greifeneder, Rainer, Herbert Bless, and Michel Tuan Pham. 2011. "When Do People Rely on Affective and Cognitive Feelings in Judgment? A Review." *Personality and Social Psychology Review* 15 (2): 107-41.

Greve, Henrich R. 1998. "Performance, Aspirations, and Risky Organizational Change." *Administrative Science Quarterly* 43 (1): 58-86. https://doi.org/10.2307/2393591.

Greve, Henrich R. 2003. *Organizational Learning from Performance Feedback: A Behavioral Perspective on Innovation and Change*. Cambridge: Cambridge University Press.

Grewal, Dhruv, Kent B. Monroe, and R. Krishnan. 1998. "The Effects of Price-Comparison Advertising on Buyers' Perceptions of Acquisition Value, Transaction Value, and Behavioral Intentions." *Journal of Marketing* 62 (2): 46-59. https://doi.org/10.1177/002224299806200204.

Griffin, Ricky W. 1983. "Objective and Social Sources of Information in Task Redesign: A Field Experiment." *Administrative Science Quarterly* 28 (2): 184-200. https://doi.org/10.2307/2392617.

Griliches, Zvi. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25 (4): 501-22. https://doi.org/10.2307/1905380.

Grönroos, Christian. 1984. "A Service Quality Model and Its Marketing Implications." *European Journal of Marketing* 18 (4): 36-44. https://doi.org/10.1108/EUM0000000004784.

Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy* 94 (4): 691-719. https://doi.org/10.1086/261404.

Grossman, Sanford J., and Joseph E. Stiglitz. 1980. "On the Impossibility of Informationally Efficient

Markets." *The American Economic Review* 70 (3): 393-408.
Grubb, Edward L., and Harrison L. Grathwohl. 1967. "Consumer Self-Concept, Symbolism and Market Behavior: A Theoretical Approach." *Journal of Marketing* 31 (4): 22-27. https://doi.org/10.1177/002224296703100405.
Grunert, Klaus G., and Suzanne C. Grunert. 1995. "Measuring Subjective Meaning Structures by the Laddering Method: Theoretical Considerations and Methodological Problems." *International Journal of Research in Marketing* 12 (3): 209-25. https://doi.org/10.1016/0167-8116(95)00022-T.
Gul, Faruk. 1991. "A Theory of Disappointment Aversion." *Econometrica* 59 (3): 667-86. https://doi.org/10.2307/2938223.
Gulati, Ranjay. 1998. "Alliances and Networks." *Strategic Management Journal* 19 (4): 293-317.
Guler, Isin, Mauro F. Guillén, and John Muir Macpherson. 2002. "Global Competition, Institutions, and the Diffusion of Organizational Practices: The International Spread of ISO 9000 Quality Certificates." *Administrative Science Quarterly* 47 (2): 207-32. https://doi.org/10.2307/3094804.
Gupta, Anil K., Ken G. Smith, and Christina E. Shalley. 2006. "The Interplay Between Exploration and Exploitation." *Academy of Management Journal* 49 (4): 693-706.
Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart. 2004. "Valuing Customers." *Journal of Marketing Research* 41 (1): 7-18. https://doi.org/10.1509/jmkr.41.1.7.25084.
Güth, Werner, Rolf Schmittberger, and Bernd Schwarze. 1982. "An Experimental Analysis of Ultimatum Bargaining." *Journal of Economic Behavior & Organization* 3 (4): 367-88. https://doi.org/10.1016/0167-2681(82)90011-7.
Gutman, Jonathan. 1982. "A Means-End Chain Model Based on Consumer Categorization Processes." *Journal of Marketing* 46 (2): 60-72. https://doi.org/10.1177/002224298204600207.
Haans, Richard F. J., Constant Pieters, and Zi-Lin He. 2016. "Thinking about U: Theorizing and Testing U- and Inverted U-Shaped Relationships in Strategy Research." *Strategic Management Journal* 37 (7): 1177-95. https://doi.org/10.1002/smj.2399.
Hackman, J. Richard, and Greg R. Oldham. 1975. "Development of the Job Diagnostic Survey." *Journal of Applied Psychology* 60 (2): 159-70. https://doi.org/10.1037/h0076546.
Hackman, J. Richard, and Greg R. Oldham. 1976. "Motivation Through the Design of Work: Test of a Theory." *Organizational Behavior and Human Performance* 16 (2): 250-79. https://doi.org/10.1016/0030-5073(76)90016-7.
Hackman, J. Richard, and Greg R. Oldham. 1980. *Work Redesign*. Reading, MA: Addison-Wesley.
Hagger, Martin S., Nikos L. D. Chatzisarantis, Hugo Alberts, Calvin O. Anggono, Cedric Batailler, Angela R. Birt, et al. 2016. "A Multilab Preregistered Replication of the Ego-Depletion Effect." *Perspectives on Psychological Science* 11 (4): 546-73. https://doi.org/10.1177/1745691616652873.
Haigh, Michael S., and John A. List. 2005. "Do Professional Traders Exhibit Myopic Loss Aversion? An Experimental Analysis." *The Journal of Finance* 60 (1): 523-34. https://doi.org/10.1111/j.1540-6261.2005.00737.x.
Halberstadt, Jamin, and Gillian Rhodes. 2003. "It's Not Just Average Faces That Are Attractive: Computer-Manipulated Averageness Makes Birds, Fish, and Automobiles Attractive." *Psychonomic Bulletin and Review* 10 (1): 149-56. https://doi.org/10.3758/bf03196479.
Halbesleben, Jonathon R. B., Jean-Pierre Neveu, Samantha C. Paustian-Underdahl, and Mina Westman. 2014. "Getting to the 'COR': Understanding the Role of Resources in Conservation of Resources Theory." *Journal of Management* 40 (5): 1334-64. https://doi.org/10.1177/0149206314527130.
Hall, Edward T. 1976. *Beyond Culture*. Garden City, NY: Anchor Press/Doubleday.

Hallsworth, Michael, John A. List, Robert D. Metcalfe, and Ivo Vlaev. 2017. "The Behavioralist as Tax Collector: Using Natural Field Experiments to Enhance Tax Compliance." *Journal of Public Economics* 148: 14-31. https://doi.org/10.1016/j.jpubeco.2017.02.003.

Halpern, Scott D., Benjamin French, Dylan S. Small, Kathryn Saulsgiver, Michael O. Harhay, Janet Audrain-McGovern, George Loewenstein, Troyen A. Brennan, David A. Asch, and Kevin G. Volpp. 2015. "Randomized Trial of Four Financial-Incentive Programs for Smoking Cessation." *New England Journal of Medicine* 372 (22): 2108-17. https://doi.org/10.1056/NEJMoa1414293.

Hambrick, Donald C. 2007a. "The Field of Management's Devotion to Theory: Too Much of a Good Thing?" *Academy of Management Journal* 50 (6): 1346-52. https://doi.org/10.5465/amj.2007.28166119.

Hambrick, Donald C. 2007b. "Upper Echelons Theory: An Update." *Academy of Management Review* 32 (2): 334-43. https://doi.org/10.5465/amr.2007.24345254.

Hambrick, Donald C., and Sydney Finkelstein. 1987. "Managerial Discretion: A Bridge Between Polar Views of Organizational Outcomes." *Research in Organizational Behavior* 9: 369-406.

Hambrick, Donald C., and Phyllis A. Mason. 1984. "Upper Echelons: The Organization as a Reflection of Its Top Managers." *Academy of Management Review* 9 (2): 193-206. https://doi.org/10.5465/amr.1984.4277628.

Han, Sang-Pil, and Sharon Shavitt. 1994. "Persuasion and Culture: Advertising Appeals in Individualistic and Collectivistic Societies." *Journal of Experimental Social Psychology* 30 (4): 326-50.

Han, Seunghee, Jennifer S. Lerner, and Dacher Keltner. 2007. "Feelings and Consumer Decision Making: The Appraisal-Tendency Framework." *Journal of Consumer Psychology* 17 (3): 158-68.

Han, Young Jee, Joseph C. Nunes, and Xavier Drèze. 2010. "Signaling Status with Luxury Goods: The Role of Brand Prominence." *Journal of Marketing* 74 (4): 15-30. https://doi.org/10.1509/jmkg.74.4.015.

Hannan, Michael T., and Glenn R. Carroll. 1992. *Dynamics of Organizational Populations: Density, Legitimation, and Competition*. New York: Oxford University Press.

Hannan, Michael T., and John Freeman. 1977. "The Population Ecology of Organizations." *American Journal of Sociology* 82 (5): 929-64. https://doi.org/10.1086/226424.

Hannan, Michael T., and John Freeman. 1984. "Structural Inertia and Organizational Change." *American Sociological Review* 49 (2): 149-64. https://doi.org/10.2307/2095567.

Hannan, Michael T., and John Freeman. 1989. *Organizational Ecology*. Cambridge, MA: Harvard University Press.

Hansen, Morten T. 1999. "The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge Across Organization Subunits." *Administrative Science Quarterly* 44 (1): 82-111. https://doi.org/10.2307/2667032.

Hardie, Bruce G. S., Eric J. Johnson, and Peter S. Fader. 1993. "Modeling Loss Aversion and Reference Dependence Effects on Brand Choice." *Marketing Science* 12 (4): 378-94. https://doi.org/10.1287/mksc.12.4.378.

Hardin, Garrett. 1968. "The Tragedy of the Commons." *Science* 162 (3859): 1243-48. https://doi.org/10.1126/science.162.3859.1243.

Harford, Jarrad. 1999. "Corporate Cash Reserves and Acquisitions." *The Journal of Finance* 54 (6): 1969-97. https://doi.org/10.1111/0022-1082.00179.

Hargadon, Andrew, and Robert I. Sutton. 1997. "Technology Brokering and Innovation in a Product Development Firm." *Administrative Science Quarterly* 42 (4): 716-49.

Harinck, Fieke, Eric Van Dijk, Ilja Van Beest, and Paul Mersmann. 2007. "When Gains Loom Larger Than Losses: Reversed Loss Aversion for Small Amounts of Money." *Psychological Science* 18 (12): 1099-1105. https://doi.org/10.1111/j.1467-9280.2007.02031.x.

Harmon-Jones, Eddie, Cindy Harmon-Jones, and Nicholas Levy. 2015. "An Action-Based Model of

Cognitive-Dissonance Processes." *Current Directions in Psychological Science* 24 (3): 184-89. https://doi.org/10.1177/0963721414566449.

Harmon-Jones, Eddie, and Judson Mills, eds. 1999. *Cognitive Dissonance: Progress on a Pivotal Theory in Social Psychology*. Washington, DC: American Psychological Association.

Harreld, J. Bruce, III O'Reilly Charles A., and Michael L. Tushman. 2007. "Dynamic Capabilities at IBM: Driving Strategy into Action." *California Management Review* 49 (4): 21-43.

Harrington, Jesse R., and Michele J. Gelfand. 2014. "Tightness-Looseness Across the 50 United States." *Proceedings of the National Academy of Sciences* 111 (22): 7990-95.

Hart, Oliver. 1995. *Firms, Contracts, and Financial Structure*. Oxford: Clarendon Press.

Hart, Oliver, and John Moore. 1990. "Property Rights and the Nature of the Firm." *Journal of Political Economy* 98 (6): 1119-58. https://doi.org/10.1086/261729.

Harter, James K., Frank L. Schmidt, and Theodore L. Hayes. 2002. "Business-Unit-Level Relationship Between Employee Satisfaction, Employee Engagement, and Business Outcomes: A Meta-Analysis." *Journal of Applied Psychology* 87 (2): 268-79. https://doi.org/10.1037/0021-9010.87.2.268.

Harvey, Campbell R., Yan Liu, and Heqing Zhu. 2016"... And the Cross-Section of Expected Returns." *The Review of Financial Studies* 29 (1): 5-68. https://doi.org/10.1093/rfs/hhv059.

Haunschild, Pamela R., and Bilian Ni Sullivan. 2002. "Learning from Complexity: Effects of Prior Accidents and Incidents on Airlines' Learning." *Administrative Science Quarterly* 47 (4): 609-43. https://doi.org/10.2307/3094911.

Haveman, Heather A. 1993. "Follow the Leader: Mimetic Isomorphism and Entry into New Markets." *Administrative Science Quarterly* 38 (4): 593-627. https://doi.org/10.2307/2393338.

Hawawini, Gabriel, Venkat Subramanian, and Paul Verdin. 2003. "Is Performance Driven by Industry- or Firm-Specific Factors? A New Look at the Evidence." *Strategic Management Journal* 24 (1): 1-16. https://doi.org/10.1002/smj.278.

Hayes, Andrew F. 2018. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. 2nd ed. New York: Guilford Press.

Hayward, Mathew L. A., and Donald C. Hambrick. 1997. "Explaining the Premiums Paid for Large Acquisitions: Evidence of CEO Hubris." *Administrative Science Quarterly* 42 (1): 103-27. https://doi.org/10.2307/2393810.

He, Zi-Lin, and Poh-Kam Wong. 2004. "Exploration Vs. Exploitation: An Empirical Test of the Ambidexterity Hypothesis." *Organization Science* 15 (4): 481-94. https://doi.org/10.1287/orsc.1040.0078.

Heath, Chip, Richard P. Larrick, and George Wu. 1999. "Goals as Reference Points." *Cognitive Psychology* 38 (1): 79-109.

Heath, Chip, and Jack B. Soll. 1996. "Mental Budgeting and Consumer Decisions." *Journal of Consumer Research* 23 (1): 40-52.

Hedström, Peter, and Richard Swedberg, eds. 1998. *Social Mechanisms: An Analytical Approach to Social Theory*. Cambridge: Cambridge University Press.

Hedström, Peter, and Petri Ylikoski. 2010. "Causal Mechanisms in the Social Sciences." *Annual Review of Sociology* 36: 49-67.

Heffetz, Ori, and John A. List. 2014. "Is the Endowment Effect an Expectations Effect?" *Journal of the European Economic Association* 12 (5): 1396-1422. https://doi.org/10.1111/jeea.12084.

Heider, Fritz. 1946. "Attitudes and Cognitive Organization." *The Journal of Psychology* 21 (1): 107-12. https://doi.org/10.1080/00223980.1946.9917275.

Heider, Fritz. 1958. *The Psychology of Interpersonal Relations*. New York: Wiley.

Heidhues, Paul, and Botond Kőszegi. 2008. "Competition and Price Variation When Consumers Are Loss Averse." *American Economic Review* 98 (4): 1245-68. https://doi.org/10.1257/aer.98.4.1

245.
Heine, Steven J., Darrin R. Lehman, Kaiping Peng, and Joe Greenholtz. 2002. "What's Wrong with Cross-Cultural Comparisons of Subjective Likert Scales? The Reference-Group Effect." *Journal of Personality and Social Psychology* 82 (6): 903-18.
Hekkert, Paul, Dirk Snelders, and Piet C. W. van Wieringen. 2003. "'Most Advanced, yet Acceptable': Typicality and Novelty as Joint Predictors of Aesthetic Preference in Industrial Design." *British Journal of Psychology* 94 (1): 111-24.
Helfat, Constance E., Sydney Finkelstein, Will Mitchell, Margaret A. Peteraf, Harbir Singh, David J. Teece, and Sidney G. Winter. 2007. *Dynamic Capabilities: Understanding Strategic Change in Organizations*. Malden, MA: Blackwell Publishing.
Helfat, Constance E., and Margaret A. Peteraf. 2003. "The Dynamic Resource-Based View: Capability Lifecycles." *Strategic Management Journal* 24 (10): 997-1010. https://doi.org/10.1002/smj.332.
Helmreich, Robert L., and Ashleigh C. Merritt. 1998. *Culture at Work in Aviation and Medicine: National, Organizational and Professional Influences*. Aldershot, UK: Ashgate.
Helmreich, Robert L., Ashleigh C. Merritt, and John A. Wilhelm. 1999. "The Evolution of Crew Resource Management Training in Commercial Aviation." *The International Journal of Aviation Psychology* 9 (1): 19-32.
Helmreich, Robert, Elliot Aronson, and James LeFan. 1970. "To Err Is Humanizing: Sometimes: Effects of Self-Esteem, Competence, and a Pratfall on Interpersonal Attraction." *Journal of Personality and Social Psychology* 16 (2): 259-64.
Henderson, Rebecca. 1993. "Underinvestment and Incompetence as Responses to Radical Innovation: Evidence from the Photolithographic Alignment Equipment Industry." *The RAND Journal of Economics* 24 (2): 248-70. https://doi.org/10.2307/2555761.
Henderson, Rebecca M., and Kim B. Clark. 1990. "Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms." *Administrative Science Quarterly* 35 (1): 9-30. https://doi.org/10.2307/2393549.
Henrich, Joseph, Robert Boyd, Samuel Bowles, Colin Camerer, Ernst Fehr, Herbert Gintis, and Richard McElreath. 2001. "In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies." *American Economic Review* 91 (2): 73-78. https://doi.org/10.1257/aer.91.2.73.
Henrich, Joseph, Steven J. Heine, and Ara Norenzayan. 2010. "The Weirdest People in the World?" *Behavioral and Brain Sciences* 33 (2-3): 61-83.
Herrmann, Benedikt, Christian Thöni, and Simon Gächter. 2008. "Antisocial Punishment Across Societies." *Science* 319 (5868): 1362-67. https://doi.org/10.1126/science.1153808.
Hertwig, Ralph, Greg Barron, Elke U. Weber, and Ido Erev. 2004. "Decisions from Experience and the Effect of Rare Events in Risky Choice." *Psychological Science* 15 (8): 534-39. https://doi.org/10.1111/j.0956-7976.2004.00715.x.
Herzberg, Frederick, Bernard Mausner, and Barbara Bloch Snyderman. 1959. *The Motivation to Work*. New York: Wiley.
Heskett, James L., Thomas O. Jones, Gary W. Loveman, W. Earl Sasser, and Leonard A. Schlesinger. 1994. "Putting the Service-Profit Chain to Work." *Harvard Business Review* 72 (2): 164-74.
Higgins, E. Tory. 1987. "Self-Discrepancy: A Theory Relating Self and Affect." *Psychological Review* 94 (3): 319-40. https://doi.org/10.1037/0033-295X.94.3.319.
Higgins, E. Tory. 1997. "Beyond Pleasure and Pain." *American Psychologist* 52 (12): 1280-300. https://doi.org/10.1037/0003-066X.52.12.1280.
Higgins, E. Tory. 1998. "Promotion and Prevention: Regulatory Focus as a Motivational Principle." *Advances in Experimental Social Psychology* 30: 1-46. https://doi.org/10.1016/S0065-2601

(08)60381-0.
Higgins, E. Tory. 2000. "Making a Good Decision: Value from Fit." *American Psychologist* 55 (11): 1217-30. https://doi.org/10.1037/0003-066X.55.11.1217.
Higgins, E. Tory, Ronald S. Friedman, Robert E. Harlow, Lorraine Chen Idson, Ozlem N. Ayduk, and Amy Taylor. 2001. "Achievement Orientations from Subjective Histories of Success: Promotion Pride Versus Prevention Pride." *European Journal of Social Psychology* 31 (1): 3-23. https://doi.org/10.1002/ejsp.27.
Higgins, E. Tory, Lorraine Chen Idson, Antonio L. Freitas, Scott Spiegel, and Daniel C. Molden. 2003. "Transfer of Value from Fit." *Journal of Personality and Social Psychology* 84 (6): 1140-53. https://doi.org/10.1037/0022-3514.84.6.1140.
Higgins, E. Tory, and Nira Liberman. 2018. "The Loss of Loss Aversion: Paying Attention to Reference Points." *Journal of Consumer Psychology* 28 (3): 523-32. https://doi.org/10.1002/jcpy.1045.
Hillman, Amy J., and Thomas Dalziel. 2003. "Boards of Directors and Firm Performance: Integrating Agency and Resource Dependence Perspectives." *Academy of Management Review* 28 (3): 383-96. https://doi.org/10.5465/amr.2003.10196729.
Hillman, Amy J., Michael C. Withers, and Brian J. Collins. 2009. "Resource Dependence Theory: A Review." *Journal of Management* 35 (6): 1404-27. https://doi.org/10.1177/0149206309343469.
Hirschman, Elizabeth C., and Morris B. Holbrook. 1982. "Hedonic Consumption: Emerging Concepts, Methods and Propositions." *Journal of Marketing* 46 (3): 92-101. https://doi.org/10.1177/002224298204600314.
Hirshleifer, David, and Tyler Shumway. 2003. "Good Day Sunshine: Stock Returns and the Weather." *The Journal of Finance* 58 (3): 1009-32. https://doi.org/10.1111/1540-6261.00556.
Hobfoll, Stevan E. 1989. "Conservation of Resources: A New Attempt at Conceptualizing Stress." *American Psychologist* 44 (3): 513-24. https://doi.org/10.1037/0003-066X.44.3.513.
Hobfoll, Stevan E., Jonathon Halbesleben, Jean-Pierre Neveu, and Mina Westman. 2018. "Conservation of Resources in the Organizational Context: The Reality of Resources and Their Consequences." *Annual Review of Organizational Psychology and Organizational Behavior* 5: 103-28. https://doi.org/10.1146/annurev-orgpsych-032117-104640.
Hoffman, Elizabeth, and Matthew L. Spitzer. 1982. "The Coase Theorem: Some Experimental Tests." *The Journal of Law and Economics* 25 (1): 73-98. https://doi.org/10.1086/467008.
Hofmann, Wilhelm, Roy F. Baumeister, Georg Förster, and Kathleen D. Vohs. 2012. "Everyday Temptations: An Experience Sampling Study of Desire, Conflict, and Self-Control." *Journal of Personality and Social Psychology* 102 (6): 1318-35. https://doi.org/10.1037/a0026545.
Hofstede, Geert. 1980. *Culture's Consequences: International Differences in Work-Related Values*. Beverly Hills, CA: Sage.
Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations*. 2nd ed. Thousand Oaks, CA: Sage.
Hofstede, Geert. 2006. "What Did GLOBE Really Measure? Researchers' Minds Versus Respondents' Minds." *Journal of International Business Studies* 37 (6): 882-96.
Hofstede, Geert. 2011. "Dimensionalizing Cultures: The Hofstede Model in Context." *Online Readings in Psychology and Culture* 2 (1). https://doi.org/10.9707/2307-0919.1014.
Hofstede, Geert, and Michael Harris Bond. 1988. "The Confucius Connection: From Cultural Roots to Economic Growth." *Organizational Dynamics* 16 (4): 5-21.
Hofstede, Geert, Gert Jan Hofstede, and Michael Minkov. 2010. *Cultures and Organizations: Software of the Mind*. 3rd ed. New York: McGraw-Hill.
Holbrook, Morris B., and Elizabeth C. Hirschman. 1982. "The Experiential Aspects of Consumption:

Consumer Fantasies, Feelings, and Fun." *Journal of Consumer Research* 9 (2): 132-40. https://doi.org/10.1086/208906.

Holmström, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics* 10 (1): 74-91. https://doi.org/10.2307/3003320.

Holmström, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, and Organization* 7 (special issue): 24-52. https://doi.org/10.1093/jleo/7.special_issue.24.

Holt, Douglas B. 1998. "Does Cultural Capital Structure American Consumption?" *Journal of Consumer Research* 25 (1): 1-25. https://doi.org/10.1086/209523.

Holt, Douglas B. 2002. "Why Do Brands Cause Trouble? A Dialectical Theory of Consumer Culture and Branding." *Journal of Consumer Research* 29 (1): 70-90. https://doi.org/10.1086/339922.

Holt, Douglas B. 2004. *How Brands Become Icons: The Principles of Cultural Branding*. Boston, MA: Harvard Business School Press.

Homans, George C. 1958. "Social Behavior as Exchange." *American Journal of Sociology* 63 (6): 597-606. https://doi.org/10.1086/222355.

Hong, Lu, and Scott E. Page. 2004. "Groups of Diverse Problem Solvers Can Outperform Groups of High-Ability Problem Solvers." *Proceedings of the National Academy of Sciences* 101 (46): 16385-89. https://doi.org/10.1073/pnas.0403723101.

Hornsey, Matthew J. 2008. "Social Identity Theory and Self-Categorization Theory: A Historical Review." *Social and Personality Psychology Compass* 2 (1): 204-22. https://doi.org/10.1111/j.1751-9004.2007.00066.x.

Hossain, Tanjim, and John A. List. 2012. "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations." *Management Science* 58 (12): 2151-67. https://doi.org/10.1287/mnsc.1120.1544.

House, Robert J. 1971. "A Path Goal Theory of Leader Effectiveness." *Administrative Science Quarterly* 16 (3): 321-39. https://doi.org/10.2307/2391905.

House, Robert J. 1996. "Path-Goal Theory of Leadership: Lessons, Legacy, and a Reformulated Theory." *The Leadership Quarterly* 7 (3): 323-52. https://doi.org/10.1016/S1048-9843(96)90024-7.

House, Robert J., Paul J. Hanges, Mansour Javidan, Peter W. Dorfman, and Vipin Gupta, eds. 2004. *Culture, Leadership, and Organizations: The GLOBE Study of 62 Societies*. Thousand Oaks, CA: Sage.

Hovland, Carl I., Irving L. Janis, and Harold H. Kelley. 1953. *Communication and Persuasion: Psychological Studies of Opinion Change*. New Haven, CT: Yale University Press.

Huber, George P. 1991. "Organizational Learning: The Contributing Processes and the Literatures." *Organization Science* 2 (1): 88-115. https://doi.org/10.1287/orsc.2.1.88.

Huber, Joel, John W. Payne, and Christopher Puto. 1982. "Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis." *Journal of Consumer Research* 9 (1): 90-98. https://doi.org/10.1086/208899.

Huber, Joel, John W. Payne, and Christopher P. Puto. 2014. "Let's Be Honest about the Attraction Effect." *Journal of Marketing Research* 51 (4): 520-25.

Hull, C. L. 1932. "The Goal-Gradient Hypothesis and Maze Learning." *Psychological Review* 39 (1): 25-43. https://doi.org/10.1037/h0072640.

Hull, C. L. 1934. "The Rat's Speed-of-Locomotion Gradient in the Approach to Food." *Journal of Comparative Psychology* 17 (3): 393-422. https://doi.org/10.1037/h0071299.

Humphrey, Stephen E., Jennifer D. Nahrgang, and Frederick P. Morgeson. 2007. "Integrating Motivational, Social, and Contextual Work Design Features: A Meta-Analytic Summary and Theoretical Extension of the Work Design Literature." *Journal of Applied Psychology* 92 (5): 1332-56.

https://doi.org/10.1037/0021-9010.92.5.1332.
Hunt, Shelby D. 2000. *A General Theory of Competition: Resources, Competences, Productivity, Economic Growth*. Thousand Oaks, CA: Sage.
Hunt, Shelby D. 2007. "Resource-Advantage Theory: An Evolutionary, Process Theory of Competition." In *A General Theory of Competition: Resources, Competences, Productivity, Economic Growth*, 135-56. SAGE Publications, Inc. https://doi.org/10.4135/9781452220321.n6.
Hunt, Shelby D., and Robert M. Morgan. 1995. "The Comparative Advantage Theory of Competition." *Journal of Marketing* 59 (2): 1-15. https://doi.org/10.1177/002224299505900201.
Hunt, Shelby D., and Robert M. Morgan. 1996. "The Resource-Advantage Theory of Competition: Dynamics, Path Dependencies, and Evolutionary Dimensions." *Journal of Marketing* 60 (4): 107-14.
Huseman, Richard C., John D. Hatfield, and Edward W. Miles. 1987. "A New Perspective on Equity Theory: The Equity Sensitivity Construct." *Academy of Management Review* 12 (2): 222-34.
Huston, Larry, and Nabil Sakkab. 2006. "Connect and Develop: Inside Procter & Gamble's New Model for Innovation." *Harvard Business Review* 84 (3): 58-66.
Hymer, Stephen H. 1976. *The International Operations of National Firms: A Study of Direct Foreign Investment*. Cambridge, MA: MIT Press.
Iannone, Richard, and Olivier Roy. 2024. *DiagrammeR: Graph/Network Visualization*. https://rich-iannone.github.io/DiagrammeR/.
Ilies, Remus, and Timothy A. Judge. 2002. "Understanding the Dynamic Relationships Among Personality, Mood, and Job Satisfaction: A Field Experience Sampling Study." *Organizational Behavior and Human Decision Processes* 89 (2): 1119-39. https://doi.org/10.1016/S0749-5978(02)00018-3.
Imas, Alex. 2016. "The Realization Effect: Risk-Taking After Realized Versus Paper Losses." *American Economic Review* 106 (8): 2086-2109. https://doi.org/10.1257/aer.20140386.
Inman, J. Jeffrey. 2001. "The Role of Sensory-Specific Satiety in Attribute-Level Variety Seeking." *Journal of Consumer Research* 28 (1): 105-20. https://doi.org/10.1086/321950.
Inzlicht, Michael, and Brandon J. Schmeichel. 2012. "What Is Ego Depletion? Toward a Mechanistic Revision of the Resource Model of Self-Control." *Perspectives on Psychological Science* 7 (5): 450-63. https://doi.org/10.1177/1745691612454134.
Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Medicine* 2 (8): e124. https://doi.org/10.1371/journal.pmed.0020124.
Ireland, Molly E., Richard B. Slatcher, Paul W. Eastwick, Lauren E. Scissors, Eli J. Finkel, and James W. Pennebaker. 2011. "Language Style Matching Predicts Relationship Initiation and Stability." *Psychological Science* 22 (1): 39-44. https://doi.org/10.1177/0956797610392928.
Isen, Alice M., Kimberly A. Daubman, and Gary P. Nowicki. 1987. "Positive Affect Facilitates Creative Problem Solving." *Journal of Personality and Social Psychology* 52 (6): 1122-31. https://doi.org/10.1037/0022-3514.52.6.1122.
Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente. 2011. "Opinion Leadership and Social Contagion in New Product Diffusion." *Marketing Science* 30 (2): 195-212. https://doi.org/10.1287/mksc.1100.0566.
Iyengar, Shanto, and Donald R. Kinder. 1987. *News That Matters: Television and American Opinion*. Chicago: University of Chicago Press.
Iyengar, Sheena S., and Mark R. Lepper. 2000. "When Choice Is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology* 79 (6): 995-1006. https://doi.org/10.1037/0022-3514.79.6.995.
Iyer, Dinesh N., and Kent D. Miller. 2008. "Performance Feedback, Slack, and the Timing of Acquisitions." *Academy of Management Journal* 51 (4): 808-22. https://doi.org/10.5465/amj.2008.3

3666024.
Jachimowicz, Jon M., Shannon Duncan, Elke U. Weber, and Eric J. Johnson. 2019. "When and Why Defaults Influence Decisions: A Meta-Analysis of Default Effects." *Behavioural Public Policy* 3 (2): 159-86.
Jacobides, Michael G., Carmelo Cennamo, and Annabelle Gawer. 2018. "Towards a Theory of Ecosystems." *Strategic Management Journal* 39 (8): 2255-76. https://doi.org/10.1002/smj.2904.
Jacobson, Robert, and Natalie Mizik. 2009. "The Financial Markets and Customer Satisfaction: Reexamining Possible Financial Market Mispricing of Customer Satisfaction." *Marketing Science* 28 (5): 810-19. https://doi.org/10.1287/mksc.1090.0495.
Jacoby, Jacob. 2002. "Stimulus-Organism-Response Reconsidered: An Evolutionary Step in Modeling (Consumer) Behavior." *Journal of Consumer Psychology* 12 (1): 51-57.
Jacoby, Larry L., Colleen Kelley, Judith Brown, and Jennifer Jasechko. 1989. "Becoming Famous Overnight: Limits on the Ability to Avoid Unconscious Influences of the Past." *Journal of Personality and Social Psychology* 56 (3): 326-38. https://doi.org/10.1037/0022-3514.56.3.326.
Janis, Irving L. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Boston: Houghton Mifflin.
Janis, Irving L. 1982. *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. 2nd ed. Boston: Houghton Mifflin.
Jansson, David G., and Steven M. Smith. 1991. "Design Fixation." *Design Studies* 12 (1): 3-11.
Javidan, Mansour, Robert J. House, Peter W. Dorfman, Paul J. Hanges, and Mary Sully de Luque. 2006. "Conceptualizing and Measuring Cultures and Their Consequences: A Comparative Review of GLOBE's and Hofstede's Approaches." *Journal of International Business Studies* 37 (6): 897-914.
Jaworski, Bernard J., and Ajay K. Kohli. 1993. "Market Orientation: Antecedents and Consequences." *Journal of Marketing* 57 (3): 53-70. https://doi.org/10.1177/002224299305700304.
Jegadeesh, Narasimhan, and Sheridan Titman. 1993. "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency." *The Journal of Finance* 48 (1): 65-91. https://doi.org/10.1111/j.1540-6261.1993.tb04702.x.
Jensen, Michael C. 1968. "The Performance of Mutual Funds in the Period 1945-1964." *The Journal of Finance* 23 (2): 389-416. https://doi.org/10.1111/j.1540-6261.1968.tb00815.x.
Jensen, Michael C. 1986. "Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers." *The American Economic Review* 76 (2): 323-29.
Jensen, Michael C., and William H. Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics* 3 (4): 305-60. https://doi.org/10.1016/0304-405X(76)90026-X.
Jensen, Michael C., and Kevin J. Murphy. 1990. "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98 (2): 225-64. https://doi.org/10.1086/261677.
Jin, Ginger Zhe, and Phillip Leslie. 2003. "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards." *The Quarterly Journal of Economics* 118 (2): 409-51. https://doi.org/10.1162/003355303321675428.
Johanson, Jan, and Jan-Erik Vahlne. 1977. "The Internationalization Process of the Firm: A Model of Knowledge Development and Increasing Foreign Market Commitments." *Journal of International Business Studies* 8 (1): 23-32.
Johanson, Jan, and Jan-Erik Vahlne. 2009. "The Uppsala Internationalization Process Model Revisited: From Liability of Foreignness to Liability of Outsidership." *Journal of International Business Studies* 40 (9): 1411-31. https://doi.org/10.1057/jibs.2009.24.
Johanson, Jan, and Finn Wiedersheim-Paul. 1975. "The Internationalization of the Firm: Four Swedish Cases." *Journal of Management Studies* 12 (3): 305-23. https://doi.org/10.1111/j.1467-6486.1975.tb00514.x.

John, Leslie K., George Loewenstein, and Drazen Prelec. 2012. "Measuring the Prevalence of Questionable Research Practices with Incentives for Truth Telling." *Psychological Science* 23 (5): 524-32. https://doi.org/10.1177/0956797611430953.

Johnson, Blair T., and Alice H. Eagly. 1989. "Effects of Involvement on Persuasion: A Meta-Analysis." *Psychological Bulletin* 106 (2): 290-314. https://doi.org/10.1037/0033-2909.106.2.290.

Johnson, Eric J., and Daniel Goldstein. 2003. "Do Defaults Save Lives?" *Science* 302 (5649): 1338-39. https://doi.org/10.1126/science.1091721.

Johnson, Eric J., John Hershey, Jacqueline Meszaros, and Howard Kunreuther. 1993. "Framing, Probability Distortions, and Insurance Decisions." *Journal of Risk and Uncertainty* 7 (1): 35-51. https://doi.org/10.1007/BF01065313.

Johnson, Eric J., Suzanne B. Shu, Benedict G. C. Dellaert, Craig Fox, Daniel G. Goldstein, Gerald Häubl, Richard P. Larrick, et al. 2012. "Beyond Nudges: Tools of a Choice Architecture." *Marketing Letters* 23 (2): 487-504.

Johnson, Roger N. 1972. *Aggression in Man and Animals*. Wiley.

Jones, Edward E., and Victor A. Harris. 1967. "The Attribution of Attitudes." *Journal of Experimental Social Psychology* 3 (1): 1-24. https://doi.org/10.1016/0022-1031(67)90034-0.

Jones, Thomas M. 1995. "Instrumental Stakeholder Theory: A Synthesis of Ethics and Economics." *Academy of Management Review* 20 (2): 404-37. https://doi.org/10.2307/258852.

Joseph, John, and William Ocasio. 2012. "Architecture, Attention, and Adaptation in the Multibusiness Firm: General Electric from 1951 to 2001." *Strategic Management Journal* 33 (6): 633-60. https://doi.org/10.1002/smj.1971.

Judge, Timothy A., and Ronald F. Piccolo. 2004. "Transformational and Transactional Leadership: A Meta-Analytic Test of Their Relative Validity." *Journal of Applied Psychology* 89 (5): 755-68. https://doi.org/10.1037/0021-9010.89.5.755.

Jussila, Iiro, Anssi Tarkiainen, Marko Sarstedt, and Joseph F. Hair. 2015. "Individual Psychological Ownership: Concepts, Evidence, and Implications for Research in Marketing." *Journal of Marketing Theory and Practice* 23 (2): 121-39. https://doi.org/10.1080/10696679.2015.1002330.

Kagel, John H., and Dan Levin. 1986. "The Winner's Curse and Public Information in Common Value Auctions." *American Economic Review* 76 (5): 894-920.

Kahn, William A. 1990. "Psychological Conditions of Personal Engagement and Disengagement at Work." *Academy of Management Journal* 33 (4): 692-724. https://doi.org/10.2307/256287.

Kahneman, Daniel. 2003. "A Perspective on Judgment and Choice: Mapping Bounded Rationality." *American Psychologist* 58 (9): 697-720. https://doi.org/10.1037/0003-066X.58.9.697.

Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Kahneman, Daniel, and Angus Deaton. 2010. "High Income Improves Evaluation of Life but Not Emotional Well-Being." *Proceedings of the National Academy of Sciences* 107 (38): 16489-93. https://doi.org/10.1073/pnas.1011492107.

Kahneman, Daniel, Barbara L. Fredrickson, Charles A. Schreiber, and Donald A. Redelmeier. 1993. "When More Pain Is Preferred to Less: Adding a Better End." *Psychological Science* 4 (6): 401-5. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x.

Kahneman, Daniel, Jack L. Knetsch, and Richard Thaler. 1986. "Fairness as a Constraint on Profit Seeking: Entitlements in the Market." *American Economic Review* 76 (4): 728-41.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1990. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy* 98 (6): 1325-48. https://doi.org/10.1086/261737.

Kahneman, Daniel, Jack L. Knetsch, and Richard H. Thaler. 1991. "Anomalies: The Endowment Effect, Loss Aversion, and Status Quo Bias." *Journal of Economic Perspectives* 5 (1): 193-206. https://doi.org/10.1257/jep.5.1.193.

Kahneman, Daniel, Alan B. Krueger, David A. Schkade, Norbert Schwarz, and Arthur A. Stone. 2004. "A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method." *Science* 306 (5702): 1776-80. https://doi.org/10.1126/science.1103572.

Kahneman, Daniel, and Dan Lovallo. 1993. "Timid Choices and Bold Forecasts: A Cognitive Perspective on Risk Taking." *Management Science* 39 (1): 17-31. https://doi.org/10.1287/mnsc.39.1.17.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision Under Risk." *Econometrica* 47 (2): 263-91. https://doi.org/10.2307/1914185.

Kahneman, Daniel, and Amos Tversky. 1984. "Choices, Values, and Frames." *American Psychologist* 39 (4): 341-50. https://doi.org/10.1037/0003-066X.39.4.341.

Kahneman, Daniel, and Amos Tversky. 1996. "On the Reality of Cognitive Illusions." *Psychological Review* 103 (3): 582-91. https://doi.org/10.1037/0033-295X.103.3.582.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *The Quarterly Journal of Economics* 112 (2): 375-405. https://doi.org/10.1162/003355397555235.

Kale, Jayant R., Ebru Reis, and Anand Venkateswaran. 2009. "Rank-Order Tournaments and Incentive Alignment: The Effect on Firm Performance." *The Journal of Finance* 64 (3): 1479-1512. https://doi.org/10.1111/j.1540-6261.2009.01470.x.

Kalyanaram, Gurumurthy, William T. Robinson, and Glen L. Urban. 1995. "Order of Market Entry: Established Empirical Generalizations, Emerging Empirical Generalizations, and Future Research." *Marketing Science* 14 (3): G212-21. https://doi.org/10.1287/mksc.14.3.G212.

Kalyanaram, Gurumurthy, and Russell S. Winer. 1995. "Empirical Generalizations from Reference Price Research." *Marketing Science* 14 (3): G161-69. https://doi.org/10.1287/mksc.14.3.G161.

Kanfer, Ruth, Michael Frese, and Russell E. Johnson. 2017. "Motivation Related to Work: A Century of Progress." *Journal of Applied Psychology* 102 (3): 338-55.

Kaplan, Steven. 1989. "The Effects of Management Buyouts on Operating Performance and Value." *Journal of Financial Economics* 24 (2): 217-54. https://doi.org/10.1016/0304-405X(89)90047-0.

Karasek, Jr., Robert A. 1979. "Job Demands, Job Decision Latitude, and Mental Strain: Implications for Job Redesign." *Administrative Science Quarterly* 24 (2): 285-308. https://doi.org/10.2307/2392498.

Karau, Steven J., and Kipling D. Williams. 1993. "Social Loafing: A Meta-Analytic Review and Theoretical Integration." *Journal of Personality and Social Psychology* 65 (4): 681-706. https://doi.org/10.1037/0022-3514.65.4.681.

Karnani, Aneel, and Birger Wernerfelt. 1985. "Multiple Point Competition." *Strategic Management Journal* 6 (1): 87-96. https://doi.org/10.1002/smj.4250060107.

Katz, Elihu, Jay G. Blumler, and Michael Gurevitch. 1973. "Uses and Gratifications Research." *Public Opinion Quarterly* 37 (4): 509-23. https://doi.org/10.1086/268109.

Katz, Elihu, and Paul F. Lazarsfeld. 1955. *Personal Influence: The Part Played by People in the Flow of Mass Communications*. Glencoe, IL: Free Press.

Katz, Michael L., and Carl Shapiro. 1985. "Network Externalities, Competition, and Compatibility." *American Economic Review* 75 (3): 424-40.

Katz, Michael L., and Carl Shapiro. 1994. "Systems Competition and Network Effects." *Journal of Economic Perspectives* 8 (2): 93-115. https://doi.org/10.1257/jep.8.2.93.

Kaustia, Markku. 2010. "Prospect Theory and the Disposition Effect." *Journal of Financial and Quantitative Analysis* 45 (3): 791-812. https://doi.org/10.1017/S0022109010000244.

Keil, Mark, Joan Mann, and Arun Rai. 2000. "Why Software Projects Escalate: An Empirical Analysis and Test of Four Theoretical Models." *MIS Quarterly* 24 (4): 631-47. https://doi.org/10.2307/

3250950.
Keiningham, Timothy L., Bruce Cooil, Tor Wallin Andreassen, and Lerzan Aksoy. 2007. "A Longitudinal Examination of Net Promoter and Firm Revenue Growth." *Journal of Marketing* 71 (3): 39-51. https://doi.org/10.1509/jmkg.71.3.039.
Keller, Kevin Lane. 1993. "Conceptualizing, Measuring, and Managing Customer-Based Brand Equity." *Journal of Marketing* 57 (1): 1-22.
Keller, Kevin Lane, and Donald R. Lehmann. 2006. "Brands and Branding: Research Findings and Future Priorities." *Marketing Science* 25 (6): 740-59. https://doi.org/10.1287/mksc.1050.0153.
Kelley, Harold H. 1967. "Attribution Theory in Social Psychology." In *Nebraska Symposium on Motivation*, edited by David Levine, 15:192-238. Lincoln: University of Nebraska Press.
Kelley, Harold H. 1973. "The Processes of Causal Attribution." *American Psychologist* 28 (2): 107-28. https://doi.org/10.1037/h0034225.
Keren, Gideon, and Yaacov Schul. 2009. "Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories." *Perspectives on Psychological Science* 4 (6): 533-50.
Kerr, Norbert L. 1998. "HARKing: Hypothesizing After the Results Are Known." *Personality and Social Psychology Review* 2 (3): 196-217. https://doi.org/10.1207/s15327957pspr0203_4.
Kerr, Steven. 1975. "On the Folly of Rewarding A, While Hoping for B." *Academy of Management Journal* 18 (4): 769-83. https://doi.org/10.5465/255378.
Kerr, Steven, and John M. Jermier. 1978. "Substitutes for Leadership: Their Meaning and Measurement." *Organizational Behavior and Human Performance* 22 (3): 375-403. https://doi.org/10.1016/0030-5073(78)90023-5.
Khan, Uzma, and Ravi Dhar. 2006. "Licensing Effect in Consumer Choice." *Journal of Marketing Research* 43 (2): 259-66. https://doi.org/10.1509/jmkr.43.2.259.
Khanna, Tarun, and Krishna Palepu. 1997. "Why Focused Strategies May Be Wrong for Emerging Markets." *Harvard Business Review* 75 (4): 41-51.
Khanna, Tarun, and Krishna Palepu. 2000. "Is Group Affiliation Profitable in Emerging Markets? An Analysis of Diversified Indian Business Groups." *Journal of Finance* 55 (2): 867-91. https://doi.org/10.1111/0022-1082.00229.
Khanna, Tarun, and Krishna G. Palepu. 2010. *Winning in Emerging Markets: A Road Map for Strategy and Execution*. Boston, MA: Harvard Business Press.
Kidd, Celeste, Steven T. Piantadosi, and Richard N. Aslin. 2012. "The Goldilocks Effect: Human Infants Allocate Attention to Visual Sequences That Are Neither Too Simple nor Too Complex." *PLoS ONE* 7 (5): e36399. https://doi.org/10.1371/journal.pone.0036399.
Killingsworth, Matthew A. 2021. "Experienced Well-Being Rises with Income, Even Above $75,000 Per Year." *Proceedings of the National Academy of Sciences* 118 (4): e2016976118. https://doi.org/10.1073/pnas.2016976118.
Killingsworth, Matthew A., Daniel Kahneman, and Barbara Mellers. 2023. "Income and Emotional Well-Being: A Conflict Resolved." *Proceedings of the National Academy of Sciences* 120 (10): e2208661120. https://doi.org/10.1073/pnas.2208661120.
King, Andrew A., and Baljir Baatartogtokh. 2015. "How Useful Is the Theory of Disruptive Innovation?" *MIT Sloan Management Review* 57 (1): 77-90.
King, Andrew A., Michael J. Lenox, and Ann Terlaak. 2005. "The Strategic Use of Decentralized Institutions: Exploring Certification with the ISO 14001 Management Standard." *Academy of Management Journal* 48 (6): 1091-1106. https://doi.org/10.5465/amj.2005.19573111.
King, William R., and Jun He. 2006. "A Meta-Analysis of the Technology Acceptance Model." *Information & Management* 43 (6): 740-55. https://doi.org/10.1016/j.im.2006.05.003.
Kingdon, John W. 1984. *Agendas, Alternatives, and Public Policies*. Boston: Little, Brown.

Kirca, Ahmet H., Satish Jayachandran, and William O. Bearden. 2005. "Market Orientation: A Meta-Analytic Review and Assessment of Its Antecedents and Impact on Performance." *Journal of Marketing* 69 (2): 24-41. https://doi.org/10.1509/jmkg.69.2.24.60761.

Kirkman, Bradley L., Kevin B. Lowe, and Cristina B. Gibson. 2006. "A Quarter Century of Culture's Consequences: A Review of Empirical Research Incorporating Hofstede's Cultural Values Framework." *Journal of International Business Studies* 37 (3): 285-320. https://doi.org/10.1057/palgrave.jibs.8400202.

Kirmani, Amna, and Akshay R. Rao. 2000. "No Pain, No Gain: A Critical Review of the Literature on Signaling Unobservable Product Quality." *Journal of Marketing* 64 (2): 66-79. https://doi.org/10.1509/jmkg.64.2.66.18000.

Kirmani, Amna, and Peter Wright. 1989. "Money Talks: Perceived Advertising Expense and Expected Product Quality." *Journal of Consumer Research* 16 (3): 344-53. https://doi.org/10.1086/209220.

Kirzner, Israel M. 1973. *Competition and Entrepreneurship*. Chicago: University of Chicago Press.

Kivetz, Ran, and Itamar Simonson. 2002. "Earning the Right to Indulge: Effort as a Determinant of Customer Preferences Toward Frequency Program Rewards." *Journal of Marketing Research* 39 (2): 155-70. https://doi.org/10.1509/jmkr.39.2.155.19084.

Kivetz, Ran, Oleg Urminsky, and Yuhuang Zheng. 2006. "The Goal-Gradient Hypothesis Resurrected: Purchase Acceleration, Illusionary Goal Progress, and Customer Retention." *Journal of Marketing Research* 43 (1): 39-58. https://doi.org/10.1509/jmkr.43.1.39.

Klein, Benjamin. 2000. "Fisher-General Motors and the Nature of the Firm." *The Journal of Law and Economics* 43 (1): 105-42. https://doi.org/10.1086/467449.

Klein, Benjamin, Robert G. Crawford, and Armen A. Alchian. 1978. "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process." *The Journal of Law and Economics* 21 (2): 297-326. https://doi.org/10.1086/466922.

Klein, Gary. 2007. "Performing a Project Premortem." *Harvard Business Review* 85 (9): 18-19.

Klein, Katherine J., Fred Dansereau, and Rosalie J. Hall. 1994. "Levels Issues in Theory Development, Data Collection, and Analysis." *Academy of Management Review* 19 (2): 195-229. https://doi.org/10.5465/amr.1994.9410210745.

Klein, Richard A., Kate A. Ratliff, Michelangelo Vianello, Reginald B. Adams Jr., Štěpán Bahník, Michael J. Bernstein, et al. 2014. "Investigating Variation in Replicability: A 'Many Labs' Replication Project." *Social Psychology* 45 (3): 142-52. https://doi.org/10.1027/1864-9335/a000178.

Klemperer, Paul. 2002. "What Really Matters in Auction Design." *Journal of Economic Perspectives* 16 (1): 169-89. https://doi.org/10.1257/0895330027166.

Klepper, Steven. 1996. "Entry, Exit, Growth, and Innovation over the Product Life Cycle." *American Economic Review* 86 (3): 562-83.

Knight, Gary A., and S. Tamer Cavusgil. 2004. "Innovation, Organizational Capabilities, and the Born-Global Firm." *Journal of International Business Studies* 35 (2): 124-41. https://doi.org/10.1057/palgrave.jibs.8400071.

Knippenberg, Daan van, and Sim B. Sitkin. 2013. "A Critical Assessment of Charismatic-Transformational Leadership Research: Back to the Drawing Board?" *Academy of Management Annals* 7 (1): 1-60. https://doi.org/10.5465/19416520.2013.759433.

Kogut, Bruce. 1991. "Joint Ventures and the Option to Expand and Acquire." *Management Science* 37 (1): 19-33. https://doi.org/10.1287/mnsc.37.1.19.

Kogut, Bruce, and Harbir Singh. 1988. "The Effect of National Culture on the Choice of Entry Mode." *Journal of International Business Studies* 19 (3): 411-32. https://doi.org/10.1057/palgrave.jibs.8490394.

Kogut, Bruce, and Udo Zander. 1992. "Knowledge of the Firm, Combinative Capabilities, and the

Replication of Technology." *Organization Science* 3 (3): 383-97. https://doi.org/10.1287/orsc.3.3.383.

Kogut, Bruce, and Udo Zander. 1993. "Knowledge of the Firm and the Evolutionary Theory of the Multinational Corporation." *Journal of International Business Studies* 24 (4): 625-45. https://doi.org/10.1057/palgrave.jibs.8490248.

Kohavi, Ron, Diane Tang, and Ya Xu. 2020. *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge: Cambridge University Press.

Kohavi, Ron, and Stefan Thomke. 2017. "The Surprising Power of Online Experiments." *Harvard Business Review* 95 (5): 74-82.

Kohli, Ajay K., and Bernard J. Jaworski. 1990. "Market Orientation: The Construct, Research Propositions, and Managerial Implications." *Journal of Marketing* 54 (2): 1-18. https://doi.org/10.1177/002224299005400201.

Kohli, Ajay K., Bernard J. Jaworski, and Ajith Kumar. 1993. "MARKOR: A Measure of Market Orientation." *Journal of Marketing Research* 30 (4): 467-77. https://doi.org/10.1177/002224379303000406.

Koo, Minjung, and Ayelet Fishbach. 2008. "Dynamics of Self-Regulation: How (Un)accomplished Goal Actions Affect Motivation." *Journal of Personality and Social Psychology* 94 (2): 183-95.

Kossinets, Gueorgi, and Duncan J. Watts. 2009. "Origins of Homophily in an Evolving Social Network." *American Journal of Sociology* 115 (2): 405-50. https://doi.org/10.1086/599247.

Kostova, Tatiana, and Srilata Zaheer. 1999. "Organizational Legitimacy Under Conditions of Complexity: The Case of the Multinational Enterprise." *Academy of Management Review* 24 (1): 64-81. https://doi.org/10.5465/amr.1999.1580441.

Kőszegi, Botond, and Matthew Rabin. 2006. "A Model of Reference-Dependent Preferences." *The Quarterly Journal of Economics* 121 (4): 1133-65. https://doi.org/10.1093/qje/121.4.1133.

Kőszegi, Botond, and Matthew Rabin. 2007. "Reference-Dependent Risk Attitudes." *American Economic Review* 97 (4): 1047-73. https://doi.org/10.1257/aer.97.4.1047.

Kotler, Philip. 1973. "Atmospherics as a Marketing Tool." *Journal of Retailing* 49 (4): 48-64.

Kozinets, Robert V. 2002. "The Field Behind the Screen: Using Netnography for Marketing Research in Online Communities." *Journal of Marketing Research* 39 (1): 61-72. https://doi.org/10.1509/jmkr.39.1.61.18935.

Kozlenkova, Irina V., Stephen A. Samaha, and Robert W. Palmatier. 2013. "Resource-Based Theory in Marketing." *Journal of the Academy of Marketing Science* 42 (1): 1-21. https://doi.org/10.1007/s11747-013-0336-7.

Kozlowski, Steve W. J., and Katherine J. Klein. 2000. "A Multilevel Approach to Theory and Research in Organizations: Contextual, Temporal, and Emergent Processes." In *Multilevel Theory, Research, and Methods in Organizations: Foundations, Extensions, and New Directions*, edited by Katherine J. Klein and Steve W. J. Kozlowski, 3-90. San Francisco: Jossey-Bass.

Kramer, Adam D. I., Jamie E. Guillory, and Jeffrey T. Hancock. 2014. "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks." *Proceedings of the National Academy of Sciences* 111 (24): 8788-90. https://doi.org/10.1073/pnas.1320040111.

Kraus, Alan, and Robert H. Litzenberger. 1973. "A State-Preference Model of Optimal Financial Leverage." *The Journal of Finance* 28 (4): 911-22. https://doi.org/10.1111/j.1540-6261.1973.tb01415.x.

Kreps, David M., and Robert Wilson. 1982. "Reputation and Imperfect Information." *Journal of Economic Theory* 27 (2): 253-79. https://doi.org/10.1016/0022-0531(82)90030-8.

Krishna, Aradhna. 2012. "An Integrative Review of Sensory Marketing: Engaging the Senses to Affect Perception, Judgment and Behavior." *Journal of Consumer Psychology* 22 (3): 332-51.

Kristof, Amy L. 1996. "Person-Organization Fit: An Integrative Review of Its Conceptualizations,

Measurement, and Implications." *Personnel Psychology* 49 (1): 1-49. https://doi.org/10.1111/j.1744-6570.1996.tb01790.x.

Kristof-Brown, Amy L., Ryan D. Zimmerman, and Erin C. Johnson. 2005. "Consequences of Individuals' Fit at Work: A Meta-Analysis of Person-Job, Person-Organization, Person-Group, and Person-Supervisor Fit." *Personnel Psychology* 58 (2): 281-342. https://doi.org/10.1111/j.1744-6570.2005.00672.x.

Krueger, Alan B., and Lawrence H. Summers. 1988. "Efficiency Wages and the Inter-Industry Wage Structure." *Econometrica* 56 (2): 259-93. https://doi.org/10.2307/1911072.

Kruglanski, Arie W., and Gerd Gigerenzer. 2011. "Intuitive and Deliberate Judgments Are Based on Common Principles." *Psychological Review* 118 (1): 97-109.

Kruglanski, Arie W., and Erik P. Thompson. 1999. "Persuasion by a Single Route: A View from the Unimodel." *Psychological Inquiry* 10 (2): 83-109.

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kumar, Amit, Matthew A. Killingsworth, and Thomas Gilovich. 2014. "Waiting for Merlot: Anticipatory Consumption of Experiential and Material Purchases." *Psychological Science* 25 (10): 1924-31.

Kunst-Wilson, William R., and Robert B. Zajonc. 1980. "Affective Discrimination of Stimuli That Cannot Be Recognized." *Science* 207 (4430): 557-58. https://doi.org/10.1126/science.7352271.

Lafontaine, Francine. 1992. "Agency Theory and Franchising: Some Empirical Results." *The RAND Journal of Economics* 23 (2): 263-83. https://doi.org/10.2307/2555988.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *The Quarterly Journal of Economics* 112 (2): 443-78. https://doi.org/10.1162/003355397555253.

Lakatos, Imre. 1970. "Falsification and the Methodology of Scientific Research Programmes." In *Criticism and the Growth of Knowledge*, edited by Imre Lakatos and Alan Musgrave. Cambridge: Cambridge University Press.

Lanaj, Klodiana, Chu-Hsiang Daisy Chang, and Russell E. Johnson. 2012. "Regulatory Focus and Work-Related Outcomes: A Review and Meta-Analysis." *Psychological Bulletin* 138 (5): 998-1034. https://doi.org/10.1037/a0027723.

Landwehr, Jan R., Aparna A. Labroo, and Andreas Herrmann. 2011. "Gut Liking for the Ordinary: Incorporating Design Fluency Improves Automobile Sales Forecasts." *Marketing Science* 30 (3): 416-29. https://doi.org/10.1287/mksc.1110.0633.

Langley, Ann. 1999. "Strategies for Theorizing from Process Data." *Academy of Management Review* 24 (4): 691-710. https://doi.org/10.5465/amr.1999.2553248.

Langlois, Judith H., and Lori A. Roggman. 1990. "Attractive Faces Are Only Average." *Psychological Science* 1 (2): 115-21. https://doi.org/10.1111/j.1467-9280.1990.tb00079.x.

Larkin, Ian. 2014. "The Cost of High-Powered Incentives: Employee Gaming in Enterprise Software Sales." *Journal of Labor Economics* 32 (2): 199-227.

Larrick, Richard P., and Jack B. Soll. 2006. "Intuitions about Combining Opinions: Misappreciation of the Averaging Principle." *Management Science* 52 (1): 111-27. https://doi.org/10.1287/mnsc.1050.0459.

Latané, Bibb, and John M. Darley. 1970. *The Unresponsive Bystander: Why Doesn't He Help?* New York: Appleton-Century-Crofts.

Latané, Bibb, Kipling Williams, and Stephen Harkins. 1979. "Many Hands Make Light the Work: The Causes and Consequences of Social Loafing." *Journal of Personality and Social Psychology* 37 (6): 822-32. https://doi.org/10.1037/0022-3514.37.6.822.

Latham, Gary P., and Craig C. Pinder. 2005. "Work Motivation Theory and Research at the Dawn of the Twenty-First Century." *Annual Review of Psychology* 56: 485-516.

Laursen, Keld, and Ammon Salter. 2006. "Open for Innovation: The Role of Openness in Explaining Innovation Performance Among U.K. Manufacturing Firms." *Strategic Management Journal* 27 (2): 131-50. https://doi.org/10.1002/smj.507.

Lavidge, Robert J., and Gary A. Steiner. 1961. "A Model for Predictive Measurements of Advertising Effectiveness." *Journal of Marketing* 25 (6): 59-62. https://doi.org/10.1177/002224296102500611.

Lavie, Dovev, Uriel Stettner, and Michael L. Tushman. 2010. "Exploration and Exploitation Within and Across Organizations." *Academy of Management Annals* 4 (1): 109-55.

Lawler, Edward J. 2001. "An Affect Theory of Social Exchange." *American Journal of Sociology* 107 (2): 321-52. https://doi.org/10.1086/324071.

Lawrence, Paul R., and Jay W. Lorsch. 1967. *Organization and Environment: Managing Differentiation and Integration*. Boston: Division of Research, Graduate School of Business Administration, Harvard University.

Lazarsfeld, Paul F., Bernard Berelson, and Hazel Gaudet. 1944. *The People's Choice: How the Voter Makes up His Mind in a Presidential Campaign*. New York: Duell, Sloan and Pearce.

Lazarus, Richard S. 1991. *Emotion and Adaptation*. New York: Oxford University Press.

Lazear, Edward P. 1989. "Pay Equality and Industrial Politics." *Journal of Political Economy* 97 (3): 561-80. https://doi.org/10.1086/261616.

Lazear, Edward P. 2009. "Firm-Specific Human Capital: A Skill-Weights Approach." *Journal of Political Economy* 117 (5): 914-40. https://doi.org/10.1086/648671.

Lazear, Edward P., and Sherwin Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89 (5): 841-64. https://doi.org/10.1086/261010.

Lee, Angela Y., and Jennifer L. Aaker. 2004. "Bringing the Frame into Focus: The Influence of Regulatory Fit on Processing Fluency and Persuasion." *Journal of Personality and Social Psychology* 86 (2): 205-18. https://doi.org/10.1037/0022-3514.86.2.205.

Lee, Angela Y., Jennifer L. Aaker, and Wendi L. Gardner. 2000. "The Pleasures and Pains of Distinct Self-Construals: The Role of Interdependence in Regulatory Focus." *Journal of Personality and Social Psychology* 78 (6): 1122-34. https://doi.org/10.1037/0022-3514.78.6.1122.

Lee, Hau L., V. Padmanabhan, and Seungjin Whang. 1997a. "Information Distortion in a Supply Chain: The Bullwhip Effect." *Management Science* 43 (4): 546-58. https://doi.org/10.1287/mnsc.43.4.546.

Lee, Hau L., V. Padmanabhan, and Seungjin Whang. 1997b. "The Bullwhip Effect in Supply Chains." *Sloan Management Review* 38 (3): 93-102.

Legris, Paul, John Ingham, and Pierre Collerette. 2003. "Why Do People Use Information Technology? A Critical Review of the Technology Acceptance Model." *Information & Management* 40 (3): 191-204. https://doi.org/10.1016/S0378-7206(01)00143-4.

Leibenstein, H. 1950. "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand." *The Quarterly Journal of Economics* 64 (2): 183. https://doi.org/10.2307/1882692.

Lemmon, Michael L., Michael R. Roberts, and Jaime F. Zender. 2008. "Back to the Beginning: Persistence and the Cross-Section of Corporate Capital Structure." *The Journal of Finance* 63 (4): 1575-1608. https://doi.org/10.1111/j.1540-6261.2008.01369.x.

Leonard-Barton, Dorothy. 1992. "Core Capabilities and Core Rigidities: A Paradox in Managing New Product Development." *Strategic Management Journal* 13 (S1): 111-25. https://doi.org/10.1002/smj.4250131009.

Leonardelli, Geoffrey J., Cynthia L. Pickett, and Marilynn B. Brewer. 2010. "Optimal Distinctiveness Theory: A Framework for Social Identity, Social Cognition, and Intergroup Relations." *Advances in Experimental Social Psychology* 43: 63-113. https://doi.org/10.1016/S0065-2601(10)43002-6.

LePine, Jeffery A., Nathan P. Podsakoff, and Marcie A. LePine. 2005. "A Meta-Analytic Test of the Challenge Stressor-Hindrance Stressor Framework: An Explanation for Inconsistent Relationships Among Stressors and Performance." *Academy of Management Journal* 48 (5): 764-75. https://doi.org/10.5465/amj.2005.18803921.

Lepore, Jill. 2014. "The Disruption Machine: What the Gospel of Innovation Gets Wrong." *The New Yorker*, June.

Lepper, Mark R., David Greene, and Richard E. Nisbett. 1973. "Undermining Children's Intrinsic Interest with Extrinsic Reward: A Test of the 'Overjustification' Hypothesis." *Journal of Personality and Social Psychology* 28 (1): 129-37.

Lerner, Jennifer S., Roxana M. Gonzalez, Deborah A. Small, and Baruch Fischhoff. 2003. "Effects of Fear and Anger on Perceived Risks of Terrorism: A National Field Experiment." *Psychological Science* 14 (2): 144-50.

Lerner, Jennifer S., and Dacher Keltner. 2000. "Beyond Valence: Toward a Model of Emotion-Specific Influences on Judgement and Choice." *Cognition and Emotion* 14 (4): 473-93.

Lerner, Jennifer S., and Dacher Keltner. 2001. "Fear, Anger, and Risk." *Journal of Personality and Social Psychology* 81 (1): 146-59. https://doi.org/10.1037/0022-3514.81.1.146.

Lerner, Jennifer S., Ye Li, Piercarlo Valdesolo, and Karim S. Kassam. 2015. "Emotion and Decision Making." *Annual Review of Psychology* 66: 799-823. https://doi.org/10.1146/annurev-psych-010213-115043.

Lerner, Jennifer S., Deborah A. Small, and George Loewenstein. 2004. "Heart Strings and Purse Strings: Carryover Effects of Emotions on Economic Decisions." *Psychological Science* 15 (5): 337-41.

Leskovec, Jure, Daniel Huttenlocher, and Jon Kleinberg. 2010. "Signed Networks in Social Media." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1361-70. https://doi.org/10.1145/1753326.1753532.

Leventhal, Gerald S. 1980. "What Should Be Done with Equity Theory? New Approaches to the Study of Fairness in Social Relationships." In *Social Exchange: Advances in Theory and Research*, edited by Kenneth J. Gergen, Martin S. Greenberg, and Richard H. Willis, 27-55. New York: Plenum.

Levin, Irwin P., and Gary J. Gaeth. 1988. "How Consumers Are Affected by the Framing of Attribute Information Before and After Consuming the Product." *Journal of Consumer Research* 15 (3): 374-78. https://doi.org/10.1086/209174.

Levin, Irwin P., Sandra L. Schneider, and Gary J. Gaeth. 1998. "All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects." *Organizational Behavior and Human Decision Processes* 76 (2): 149-88.

Levinthal, Daniel A. 1991. "Random Walks and Organizational Mortality." *Administrative Science Quarterly* 36 (3): 397-420. https://doi.org/10.2307/2393202.

Levinthal, Daniel A., and James G. March. 1993. "The Myopia of Learning." *Strategic Management Journal* 14 (S2): 95-112. https://doi.org/10.1002/smj.4250141009.

Lévi-Strauss, Claude. 1966. *The Savage Mind*. Chicago: University of Chicago Press.

Levitt, Barbara, and James G. March. 1988. "Organizational Learning." *Annual Review of Sociology* 14: 319-40. https://doi.org/10.1146/annurev.so.14.080188.001535.

Levitt, Steven D., and John A. List. 2007. "What Do Laboratory Experiments Measuring Social Preferences Reveal about the Real World?" *Journal of Economic Perspectives* 21 (2): 153-74. https://doi.org/10.1257/jep.21.2.153.

Levitt, Theodore. 1965. "Exploit the Product Life Cycle." *Harvard Business Review* 43 (6): 81-94.

Levitt, Theodore. 1983. "The Globalization of Markets." *Harvard Business Review* 61 (3): 92-102.

Levy, Jack S. 1992. "Prospect Theory and International Relations: Theoretical Applications and

Analytical Problems." *Political Psychology* 13 (2): 283-310. https://doi.org/10.2307/3791682.
Lewandowsky, Stephan, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. 2012. "Misinformation and Its Correction: Continued Influence and Successful Debiasing." *Psychological Science in the Public Interest* 13 (3): 106-31. https://doi.org/10.1177/1529100612451018.
Lewis, Marianne W. 2000. "Exploring Paradox: Toward a More Comprehensive Guide." *Academy of Management Review* 25 (4): 760-76. https://doi.org/10.2307/259204.
Liberman, Nira, and Yaacov Trope. 1998. "The Role of Feasibility and Desirability Considerations in Near and Distant Future Decisions: A Test of Temporal Construal Theory." *Journal of Personality and Social Psychology* 75 (1): 5-18. https://doi.org/10.1037/0022-3514.75.1.5.
Liden, Robert C., and John M. Maslyn. 1998. "Multidimensionality of Leader-Member Exchange: An Empirical Assessment Through Scale Development." *Journal of Management* 24 (1): 43-72. https://doi.org/10.1016/S0149-2063(99)80053-1.
Lie, Erik. 2005. "On the Timing of CEO Stock Option Awards." *Management Science* 51 (5): 802-12. https://doi.org/10.1287/mnsc.1050.0365.
Lieberman, Marvin B., and David B. Montgomery. 1988. "First-Mover Advantages." *Strategic Management Journal* 9 (S1): 41-58. https://doi.org/10.1002/smj.4250090706.
Lieberman, Marvin B., and David B. Montgomery. 1998. "First-Mover (Dis)advantages: Retrospective and Link with the Resource-Based View." *Strategic Management Journal* 19 (12): 1111-25.
Lieberson, Stanley, and James F. O'Connor. 1972. "Leadership and Organizational Performance: A Study of Large Corporations." *American Sociological Review* 37 (2): 117-30. https://doi.org/10.2307/2094020.
Liebowitz, S. J., and Stephen E. Margolis. 1990. "The Fable of the Keys." *The Journal of Law and Economics* 33 (1): 1-25. https://doi.org/10.1086/467198.
Lin, Nan. 2001. *Social Capital: A Theory of Social Structure and Action*. Cambridge: Cambridge University Press.
Lind, E. Allan, Jerald Greenberg, Kimberly S. Scott, and Thomas D. Welchans. 2000. "The Winding Road from Employee to Complainant: Situational and Psychological Determinants of Wrongful-Termination Claims." *Administrative Science Quarterly* 45 (3): 557-90. https://doi.org/10.2307/2667109.
Lintner, John. 1956. "Distribution of Incomes of Corporations Among Dividends, Retained Earnings, and Taxes." *The American Economic Review* 46 (2): 97-113.
Lintner, John. 1965. "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets." *The Review of Economics and Statistics* 47 (1): 13-37. https://doi.org/10.2307/1924119.
Lippman, S. A., and R. P. Rumelt. 1982. "Uncertain Imitability: An Analysis of Interfirm Differences in Efficiency Under Competition." *The Bell Journal of Economics* 13 (2): 418-38.
List, John A. 2003. "Does Market Experience Eliminate Market Anomalies?" *The Quarterly Journal of Economics* 118 (1): 41-71. https://doi.org/10.1162/00335530360535144.
Locke, Edwin A., and Gary P. Latham. 1990. *A Theory of Goal Setting and Task Performance*. Englewood Cliffs, NJ: Prentice Hall.
Locke, Edwin A., and Gary P. Latham. 2002. "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey." *American Psychologist* 57 (9): 705-17. https://doi.org/10.1037/0003-066X.57.9.705.
Locke, Edwin A., and Gary P. Latham. 2009. "Has Goal Setting Gone Wild, or Have Its Attackers Abandoned Good Scholarship?" *Academy of Management Perspectives* 23 (1): 17-23.
Lockwood, Penelope, Christian H. Jordan, and Ziva Kunda. 2002. "Motivation by Positive or Negative Role Models: Regulatory Focus Determines Who Will Best Inspire Us." *Journal of Personality and*

Social Psychology 83 (4): 854-64. https://doi.org/10.1037/0022-3514.83.4.854.

Lockwood, Penelope, and Ziva Kunda. 1997. "Superstars and Me: Predicting the Impact of Role Models on the Self." *Journal of Personality and Social Psychology* 73 (1): 91-103. https://doi.org/10.1037/0022-3514.73.1.91.

Loewenstein, George F., Elke U. Weber, Christopher K. Hsee, and Ned Welch. 2001. "Risk as Feelings." *Psychological Bulletin* 127 (2): 267-86. https://doi.org/10.1037/0033-2909.127.2.267.

Loken, Barbara, and Deborah Roedder John. 1993. "Diluting Brand Beliefs: When Do Brand Extensions Have a Negative Impact?" *Journal of Marketing* 57 (3): 71-84. https://doi.org/10.1177/002224299305700305.

Loomes, Graham, and Robert Sugden. 1982. "Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty." *The Economic Journal* 92 (368): 805-24. https://doi.org/10.2307/2232669.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. "How Social Influence Can Undermine the Wisdom of Crowd Effect." *Proceedings of the National Academy of Sciences* 108 (22): 9020-25. https://doi.org/10.1073/pnas.1008636108.

Lowenstein, Roger. 2000. *When Genius Failed: The Rise and Fall of Long-Term Capital Management*. New York: Random House.

Lucas, Henry C., and Jie Mein Goh. 2009. "Disruptive Technology: How Kodak Missed the Digital Photography Revolution." *The Journal of Strategic Information Systems* 18 (1): 46-55. https://doi.org/10.1016/j.jsis.2009.01.002.

Lucas, Richard E. 2007. "Adaptation and the Set-Point Model of Subjective Well-Being: Does Happiness Change After Major Life Events?" *Current Directions in Psychological Science* 16 (2): 75-79. https://doi.org/10.1111/j.1467-8721.2007.00479.x.

Lucas, Richard E., Andrew E. Clark, Yannis Georgellis, and Ed Diener. 2003. "Reexamining Adaptation and the Set Point Model of Happiness: Reactions to Changes in Marital Status." *Journal of Personality and Social Psychology* 84 (3): 527-39. https://doi.org/10.1037/0022-3514.84.3.527.

Luhmann, Maike, Wilhelm Hofmann, Michael Eid, and Richard E. Lucas. 2012. "Subjective Well-Being and Adaptation to Life Events: A Meta-Analysis." *Journal of Personality and Social Psychology* 102 (3): 592-615.

Lynn, Michael. 1991. "Scarcity Effects on Value: A Quantitative Review of the Commodity Theory Literature." *Psychology and Marketing* 8 (1): 43-57. https://doi.org/10.1002/mar.4220080105.

Lynn, Michael, and C. R. Snyder. 2002. "Uniqueness Seeking." In *Handbook of Positive Psychology*.

Lyubomirsky, Sonja, Laura King, and Ed Diener. 2005. "The Benefits of Frequent Positive Affect: Does Happiness Lead to Success?" *Psychological Bulletin* 131 (6): 803-55. https://doi.org/10.1037/0033-2909.131.6.803.

Lyubomirsky, Sonja, Kennon M. Sheldon, and David Schkade. 2005. "Pursuing Happiness: The Architecture of Sustainable Change." *Review of General Psychology* 9 (2): 111-31. https://doi.org/10.1037/1089-2680.9.2.111.

MacKenzie, Scott B., Philip M. Podsakoff, and Nathan P. Podsakoff. 2011. "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques." *MIS Quarterly* 35 (2): 293-334. https://doi.org/10.2307/23044045.

Madrian, Brigitte C., and Dennis F. Shea. 2001. "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior." *The Quarterly Journal of Economics* 116 (4): 1149-87. https://doi.org/10.1162/003355301753265543.

Madsen, Peter M., and Vinit Desai. 2010. "Failing to Learn? The Effects of Failure and Success on Organizational Learning in the Global Orbital Launch Vehicle Industry." *Academy of Management*

Journal 53 (3): 451-76. https://doi.org/10.5465/amj.2010.51467631.
Mahajan, Vijay, Eitan Muller, and Frank M. Bass. 1990. "New Product Diffusion Models in Marketing: A Review and Directions for Research." *Journal of Marketing* 54 (1): 1-26. https://doi.org/10.1177/002224299005400101.
Mahoney, Joseph T., and J. Rajendran Pandian. 1992. "The Resource-Based View Within the Conversation of Strategic Management." *Strategic Management Journal* 13 (5): 363-80. https://doi.org/10.1002/smj.4250130505.
Maier, Maximilian, František Bartoš, T. D. Stanley, David R. Shanks, Adam J. L. Harris, and Eric-Jan Wagenmakers. 2022. "No Evidence for Nudging After Adjusting for Publication Bias." *Proceedings of the National Academy of Sciences* 119 (31): e2200300119. https://doi.org/10.1073/pnas.2200300119.
Main, Brian G. M., III O'Reilly Charles A., and James Wade. 1993. "Top Executive Pay: Tournament or Teamwork?" *Journal of Labor Economics* 11 (4): 606-28. https://doi.org/10.1086/298309.
Maitlis, Sally, and Marlys Christianson. 2014. "Sensemaking in Organizations: Taking Stock and Moving Forward." *Academy of Management Annals* 8 (1): 57-125. https://doi.org/10.5465/19416520.2014.873177.
Makadok, Richard. 2001. "Toward a Synthesis of the Resource-Based and Dynamic-Capability Views of Rent Creation." *Strategic Management Journal* 22 (5): 387-401. https://doi.org/10.1002/smj.158.
Malär, Lucia, Harley Krohmer, Wayne D. Hoyer, and Bettina Nyffenegger. 2011. "Emotional Brand Attachment and Brand Personality: The Relative Importance of the Actual and the Ideal Self." *Journal of Marketing* 75 (4): 35-52. https://doi.org/10.1509/jmkg.75.4.35.
Malkiel, Burton G. 2003. "The Efficient Market Hypothesis and Its Critics." *Journal of Economic Perspectives* 17 (1): 59-82. https://doi.org/10.1257/089533003321164958.
Malle, Bertram F. 2006. "The Actor-Observer Asymmetry in Attribution: A (Surprising) Meta-Analysis." *Psychological Bulletin* 132 (6): 895-919. https://doi.org/10.1037/0033-2909.132.6.895.
Malmendier, Ulrike, and Geoffrey Tate. 2005. "CEO Overconfidence and Corporate Investment." *The Journal of Finance* 60 (6): 2661-2700. https://doi.org/10.1111/j.1540-6261.2005.00813.x.
Mandler, George. 1982. "The Structure of Value: Accounting for Taste." In *Affect and Cognition: The Seventeenth Annual Carnegie Symposium on Cognition*, edited by Margaret S. Clark and Susan T. Fiske. Hillsdale, NJ: Lawrence Erlbaum Associates.
Manning, Rachel, Mark Levine, and Alan Collins. 2007. "The Kitty Genovese Murder and the Social Psychology of Helping: The Parable of the 38 Witnesses." *American Psychologist* 62 (6): 555-62. https://doi.org/10.1037/0003-066X.62.6.555.
March, James G. 1988. "Variable Risk Preferences and Adaptive Aspirations." *Journal of Economic Behavior and Organization* 9 (1): 5-24. https://doi.org/10.1016/0167-2681(88)90004-2.
March, James G. 1991. "Exploration and Exploitation in Organizational Learning." *Organization Science* 2 (1): 71-87. https://doi.org/10.1287/orsc.2.1.71.
March, James G., and Zur Shapira. 1992. "Variable Risk Preferences and the Focus of Attention." *Psychological Review* 99 (1): 172-83. https://doi.org/10.1037/0033-295X.99.1.172.
Margolis, Joshua D., and James P. Walsh. 2003. "Misery Loves Companies: Rethinking Social Initiatives by Business." *Administrative Science Quarterly* 48 (2): 268-305. https://doi.org/10.2307/3556659.
Markovsky, Barry, David Willer, and Travis Patton. 1988. "Power Relations in Exchange Networks." *American Sociological Review* 53 (2): 220-36. https://doi.org/10.2307/2095689.
Markowitz, Harry. 1952. "Portfolio Selection." *The Journal of Finance* 7 (1): 77-91. https://doi.org/10.1111/j.1540-6261.1952.tb01525.x.

Markus, Hazel. 1978. "The Effect of Mere Presence on Social Facilitation: An Unobtrusive Test." *Journal of Experimental Social Psychology* 14 (4): 389-97. https://doi.org/10.1016/0022-1031(78)90034-3.

Markus, Hazel R., and Shinobu Kitayama. 1991. "Culture and the Self: Implications for Cognition, Emotion, and Motivation." *Psychological Review* 98 (2): 224-53. https://doi.org/10.1037/0033-295X.98.2.224.

Martindale, Colin, Kathleen Moore, and Jonathan Borkum. 1990. "Aesthetic Preference: Anomalous Findings for Berlyne's Psychobiological Theory." *The American Journal of Psychology* 103 (1): 53. https://doi.org/10.2307/1423259.

Martindale, Colin, Kathleen Moore, and Alan West. 1988. "Relationship of Preference Judgments to Typicality, Novelty, and Mere Exposure." *Empirical Studies of the Arts* 6 (1): 79-96. https://doi.org/10.2190/mcaj-0gqt-djtl-lnqd.

Mas, Alexandre. 2006. "Pay, Reference Points, and Police Performance." *The Quarterly Journal of Economics* 121 (3): 783-821. https://doi.org/10.1162/qjec.121.3.783.

Mas, Alexandre, and Enrico Moretti. 2009. "Peers at Work." *American Economic Review* 99 (1): 112-45. https://doi.org/10.1257/aer.99.1.112.

Maskin, Eric, and Jean Tirole. 1999. "Unforeseen Contingencies and Incomplete Contracts." *The Review of Economic Studies* 66 (1): 83-114. https://doi.org/10.1111/1467-937X.00079.

Maslach, Christina, Wilmar B. Schaufeli, and Michael P. Leiter. 2001. "Job Burnout." *Annual Review of Psychology* 52: 397-422. https://doi.org/10.1146/annurev.psych.52.1.397.

Maslow, Abraham H. 1943. "A Theory of Human Motivation." *Psychological Review* 50 (4): 370-96. https://doi.org/10.1037/h0054346.

Mason, Edward S. 1939. "Price and Production Policies of Large-Scale Enterprise." *The American Economic Review* 29 (1): 61-74.

Mason, Winter A., Frederica R. Conrey, and Eliot R. Smith. 2007. "Situating Social Influence Processes: Dynamic, Multidirectional Flows of Influence Within Social Networks." *Personality and Social Psychology Review* 11 (3): 279-300. https://doi.org/10.1177/1088868307301032.

Matten, Dirk, and Jeremy Moon. 2008. "'Implicit' and 'Explicit' CSR: A Conceptual Framework for a Comparative Understanding of Corporate Social Responsibility." *Academy of Management Review* 33 (2): 404-24. https://doi.org/10.5465/amr.2008.31193458.

Mayer, Richard E., and Roxana Moreno. 2003. "Nine Ways to Reduce Cognitive Load in Multimedia Learning." *Educational Psychologist* 38 (1): 43-52.

Mayhew, Glenn E., and Russell S. Winer. 1992. "An Empirical Analysis of Internal and External Reference Prices Using Scanner Data." *Journal of Consumer Research* 19 (1): 62-70. https://doi.org/10.1086/209286.

Mazumdar, Tridib, S. P. Raj, and Indrajit Sinha. 2005. "Reference Price Research: Review and Propositions." *Journal of Marketing* 69 (4): 84-102. https://doi.org/10.1509/jmkg.2005.69.4.84.

McAlister, Leigh, and Edgar Pessemier. 1982. "Variety Seeking Behavior: An Interdisciplinary Review." *Journal of Consumer Research* 9 (3): 311-22. https://doi.org/10.1086/208926.

McCall, J. J. 1970. "Economics of Information and Job Search." *The Quarterly Journal of Economics* 84 (1): 113-26. https://doi.org/10.2307/1879403.

McCall, Robert B., and Paul E. McGhee. 1977. "The Discrepancy Hypothesis of Attention and Affect in Infants." In *The Structuring of Experience*, 179-210. Springer US. https://doi.org/10.1007/978-1-4615-8786-6_7.

McCarthy, E. Jerome. 1960. *Basic Marketing: A Managerial Approach*. Homewood, IL: Richard D. Irwin.

McCombs, Maxwell E., and Donald L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *Public Opinion Quarterly* 36 (2): 176-87. https://doi.org/10.1086/267990.

McCracken, Grant. 1986. "Culture and Consumption: A Theoretical Account of the Structure and Movement of the Cultural Meaning of Consumer Goods." *Journal of Consumer Research* 13 (1): 71-84. https://doi.org/10.1086/209048.

McCracken, Grant. 1989. "Who Is the Celebrity Endorser? Cultural Foundations of the Endorsement Process." *Journal of Consumer Research* 16 (3): 310-21. https://doi.org/10.1086/209217.

McGahan, Anita M., and Michael E. Porter. 1997. "How Much Does Industry Matter, Really?" *Strategic Management Journal* 18 (S1): 15-30.

McGrath, Rita Gunther. 1997. "A Real Options Logic for Initiating Technology Positioning Investments." *The Academy of Management Review* 22 (4): 974-96. https://doi.org/10.2307/259251.

McGrath, Rita Gunther. 1999. "Falling Forward: Real Options Reasoning and Entrepreneurial Failure." *The Academy of Management Review* 24 (1): 13-30. https://doi.org/10.2307/259034.

McGregor, Douglas. 1960. *The Human Side of Enterprise*. New York: McGraw-Hill.

McLean, R. David, and Jeffrey Pontiff. 2016. "Does Academic Research Destroy Stock Return Predictability?" *The Journal of Finance* 71 (1): 5-32. https://doi.org/10.1111/jofi.12365.

McMillan, John. 1994. "Selling Spectrum Rights." *Journal of Economic Perspectives* 8 (3): 145-62. https://doi.org/10.1257/jep.8.3.145.

McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. 2001. "Birds of a Feather: Homophily in Social Networks." *Annual Review of Sociology* 27: 415-44. https://doi.org/10.1146/annurev.soc.27.1.415.

McSweeney, Brendan. 2002. "Hofstede's Model of National Cultural Differences and Their Consequences: A Triumph of Faith: A Failure of Analysis." *Human Relations* 55 (1): 89-118. https://doi.org/10.1177/0018726702551004.

Mehra, Rajnish, and Edward C. Prescott. 1985. "The Equity Premium: A Puzzle." *Journal of Monetary Economics* 15 (2): 145-61. https://doi.org/10.1016/0304-3932(85)90061-3.

Mehrabian, Albert, and James A. Russell. 1974. *An Approach to Environmental Psychology*. Cambridge, MA: MIT Press.

Mertens, Stephanie, Mario Herberz, Ulf J. J. Hahnel, and Tobias Brosch. 2022. "The Effectiveness of Nudging: A Meta-Analysis of Choice Architecture Interventions Across Behavioral Domains." *Proceedings of the National Academy of Sciences* 119 (1): e2107346118. https://doi.org/10.1073/pnas.2107346118.

Merton, Robert K. 1968a. *Social Theory and Social Structure*. Enlarged. New York: Free Press.

Merton, Robert K. 1968b. "The Matthew Effect in Science." *Science* 159 (3810): 56-63. https://doi.org/10.1126/science.159.3810.56.

Meyer, John P., and Natalie J. Allen. 1991. "A Three-Component Conceptualization of Organizational Commitment." *Human Resource Management Review* 1 (1): 61-89. https://doi.org/10.1016/1053-4822(91)90011-Z.

Meyer, John P., David J. Stanley, Lynne Herscovitch, and Laryssa Topolnytsky. 2002. "Affective, Continuance, and Normative Commitment to the Organization: A Meta-Analysis of Antecedents, Correlates, and Consequences." *Journal of Vocational Behavior* 61 (1): 20-52. https://doi.org/10.1006/jvbe.2001.1842.

Meyer, John W., and Brian Rowan. 1977. "Institutionalized Organizations: Formal Structure as Myth and Ceremony." *American Journal of Sociology* 83 (2): 340-63. https://doi.org/10.1086/226550.

Meyerowitz, Beth E., and Shelly Chaiken. 1987. "The Effect of Message Framing on Breast Self-Examination Attitudes, Intentions, and Behavior." *Journal of Personality and Social Psychology* 52 (3): 500-510. https://doi.org/10.1037/0022-3514.52.3.500.

Meyers-Levy, Joan, Therese A. Louie, and Mary T. Curren. 1994. "How Does the Congruity of Brand Names Affect Evaluations of Brand Name Extensions?" *Journal of Applied Psychology* 79 (1):

46-53. https://doi.org/10.1037/0021-9010.79.1.46.
Meyers-Levy, Joan, and Alice M. Tybout. 1989. "Schema Congruity as a Basis for Product Evaluation." *Journal of Consumer Research* 16 (1): 39-54. https://doi.org/10.1086/209192.
Mezias, John M. 2002. "Identifying Liabilities of Foreignness and Strategies to Minimize Their Effects: The Case of Labor Lawsuit Judgments in the United States." *Strategic Management Journal* 23 (3): 229-44. https://doi.org/10.1002/smj.220.
Milgrom, Paul. 2004. *Putting Auction Theory to Work*. Cambridge: Cambridge University Press.
Milgrom, Paul R., and Robert J. Weber. 1982. "A Theory of Auctions and Competitive Bidding." *Econometrica* 50 (5): 1089-1122. https://doi.org/10.2307/1911865.
Milgrom, Paul, and John Roberts. 1982. "Predation, Reputation, and Entry Deterrence." *Journal of Economic Theory* 27 (2): 280-312. https://doi.org/10.1016/0022-0531(82)90031-X.
Milgrom, Paul, and John Roberts. 1986. "Price and Advertising Signals of Product Quality." *Journal of Political Economy* 94 (4): 796-821. https://doi.org/10.1086/261408.
Milgrom, Paul, and John Roberts. 1995. "Complementarities and Fit: Strategy, Structure, and Organizational Change in Manufacturing." *Journal of Accounting and Economics* 19 (2-3): 179-208. https://doi.org/10.1016/0165-4101(94)00382-F.
Milkman, Katherine L., and John Beshears. 2009. "Mental Accounting and Small Windfalls: Evidence from an Online Grocer." *Journal of Economic Behavior & Organization* 71 (2): 384-94.
Miller, Danny, and Ming-Jer Chen. 1994. "Sources and Consequences of Competitive Inertia: A Study of the U.S. Airline Industry." *Administrative Science Quarterly* 39 (1): 1-23. https://doi.org/10.2307/2393492.
Miller, George A. 1956. "The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information." *Psychological Review* 63 (2): 81-97. https://doi.org/10.1037/h0043158.
Miller, Merton H., and Franco Modigliani. 1961. "Dividend Policy, Growth, and the Valuation of Shares." *The Journal of Business* 34 (4): 411-33. https://doi.org/10.1086/294442.
Milliken, Frances J., Elizabeth W. Morrison, and Patricia F. Hewlin. 2003. "An Exploratory Study of Employee Silence: Issues That Employees Don't Communicate Upward and Why." *Journal of Management Studies* 40 (6): 1453-76. https://doi.org/10.1111/1467-6486.00387.
Milliman, Ronald E. 1982. "Using Background Music to Affect the Behavior of Supermarket Shoppers." *Journal of Marketing* 46 (3): 86-91.
Mincer, Jacob. 1974. *Schooling, Experience, and Earnings*. New York: National Bureau of Economic Research.
Mischel, Walter, Yuichi Shoda, and Monica L. Rodriguez. 1989. "Delay of Gratification in Children." *Science* 244 (4907): 933-38. https://doi.org/10.1126/science.2658056.
Mitchell, Deborah J., Barbara E. Kahn, and Susan C. Knasko. 1995. "There's Something in the Air: Effects of Congruent or Incongruent Ambient Odor on Consumer Decision Making." *Journal of Consumer Research* 22 (2): 229-38.
Mitchell, Ronald K., Bradley R. Agle, and Donna J. Wood. 1997. "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of Who and What Really Counts." *Academy of Management Review* 22 (4): 853-86. https://doi.org/10.2307/259247.
Mittal, Vikas, and Wagner A. Kamakura. 2001. "Satisfaction, Repurchase Intent, and Repurchase Behavior: Investigating the Moderating Effect of Customer Characteristics." *Journal of Marketing Research* 38 (1): 131-42. https://doi.org/10.1509/jmkr.38.1.131.18832.
Modigliani, Franco, and Merton H. Miller. 1958. "The Cost of Capital, Corporation Finance and the Theory of Investment." *The American Economic Review* 48 (3): 261-97.
Modigliani, Franco, and Merton H. Miller. 1963. "Corporate Income Taxes and the Cost of Capital: A Correction." *The American Economic Review* 53 (3): 433-43.

Mohr, Lawrence B. 1982. *Explaining Organizational Behavior*. San Francisco: Jossey-Bass.

Molm, Linda D., Nobuyuki Takahashi, and Gretchen Peterson. 2000. "Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition." *American Journal of Sociology* 105 (5): 1396-1427. https://doi.org/10.1086/210434.

Monin, Benoît. 2003. "The Warm Glow Heuristic: When Liking Leads to Familiarity." *Journal of Personality and Social Psychology* 85 (6): 1035-48. https://doi.org/10.1037/0022-3514.85.6.1035.

Monteverde, Kirk, and David J. Teece. 1982. "Supplier Switching Costs and Vertical Integration in the Automobile Industry." *The Bell Journal of Economics* 13 (1): 206-13. https://doi.org/10.2307/3003441.

Montgomery, W. David. 1972. "Markets in Licenses and Efficient Pollution Control Programs." *Journal of Economic Theory* 5 (3): 395-418. https://doi.org/10.1016/0022-0531(72)90049-X.

Montoya, R. Matthew, Robert S. Horton, Jack L. Vevea, Martyna Citkowicz, and Elissa A. Lauber. 2017. "A Re-Examination of the Mere Exposure Effect: The Influence of Repeated Exposure on Recognition, Familiarity, and Liking." *Psychological Bulletin* 143 (5): 459-98.

Moore, Gary C., and Izak Benbasat. 1991. "Development of an Instrument to Measure the Perceptions of Adopting an Information Technology Innovation." *Information Systems Research* 2 (3): 192-222. https://doi.org/10.1287/isre.2.3.192.

Moore, Geoffrey A. 1991. *Crossing the Chasm: Marketing and Selling Technology Products to Mainstream Customers*. New York: HarperBusiness.

Moors, Agnes, Phoebe C. Ellsworth, Klaus R. Scherer, and Nico H. Frijda. 2013. "Appraisal Theories of Emotion: State of the Art and Future Development." *Emotion Review* 5 (2): 119-24.

Morewedge, Carey K., and Colleen E. Giblin. 2015. "Explanations of the Endowment Effect: An Integrative Review." *Trends in Cognitive Sciences* 19 (6): 339-48.

Morewedge, Carey K., Ashwani Monga, Robert W. Palmatier, Suzanne B. Shu, and Deborah A. Small. 2021. "Evolution of Consumption: A Psychological Ownership Framework." *Journal of Marketing* 85 (1): 196-218. https://doi.org/10.1177/0022242920957007.

Morgan, Neil A., Rebecca J. Slotegraaf, and Douglas W. Vorhies. 2009. "Linking Marketing Capabilities with Profit Growth." *International Journal of Research in Marketing* 26 (4): 284-93. https://doi.org/10.1016/j.ijresmar.2009.06.005.

Morgan, Robert M., and Shelby D. Hunt. 1994. "The Commitment-Trust Theory of Relationship Marketing." *Journal of Marketing* 58 (3): 20-38. https://doi.org/10.1177/002224299405800302.

Morgenroth, Thekla, Teri A. Kirby, Michelle K. Ryan, and Antonia Sudkämper. 2020. "The Who, When, and Why of the Glass Cliff Phenomenon: A Meta-Analysis of Appointments to Precarious Leadership Positions." *Psychological Bulletin* 146 (9): 797-829.

Morgeson, Frederick P., and Stephen E. Humphrey. 2006. "The Work Design Questionnaire (WDQ): Developing and Validating a Comprehensive Measure for Assessing Job Design and the Nature of Work." *Journal of Applied Psychology* 91 (6): 1321-39. https://doi.org/10.1037/0021-9010.91.6.1321.

Morris, Michael W., and Kaiping Peng. 1994. "Culture and Cause: American and Chinese Attributions for Social and Physical Events." *Journal of Personality and Social Psychology* 67 (6): 949-71. https://doi.org/10.1037/0022-3514.67.6.949.

Morrison, Elizabeth Wolfe. 2011. "Employee Voice Behavior: Integration and Directions for Future Research." *Academy of Management Annals* 5 (1): 373-412.

Morrison, Elizabeth Wolfe, and Frances J. Milliken. 2000. "Organizational Silence: A Barrier to Change and Development in a Pluralistic World." *Academy of Management Review* 25 (4): 706-25. https://doi.org/10.2307/259200.

Morrison, Elizabeth Wolfe, and Sandra L. Robinson. 1997. "When Employees Feel Betrayed: A Model of How Psychological Contract Violation Develops." *Academy of Management Review* 22 (1): 226-56. https://doi.org/10.2307/259230.

Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *The Review of Economic Studies* 61 (3): 397-415. https://doi.org/10.2307/2297896.

Morwitz, Vicki G., Eric A. Greenleaf, and Eric J. Johnson. 1998. "Divide and Prosper: Consumers' Reactions to Partitioned Prices." *Journal of Marketing Research* 35 (4): 453-63.

Mrkva, Kellen, Eric J. Johnson, Simon Gächter, and Andreas Herrmann. 2020. "Moderating Loss Aversion: Loss Aversion Has Moderators, but Reports of Its Death Are Greatly Exaggerated." *Journal of Consumer Psychology* 30 (3): 407-28.

Muchnik, Lev, Sinan Aral, and Sean J. Taylor. 2013. "Social Influence Bias: A Randomized Experiment." *Science* 341 (6146): 647-51. https://doi.org/10.1126/science.1240466.

Mulder, Mauk. 1977. *The Daily Power Game*. Leiden: Martinus Nijhoff.

Munafò, Marcus R., Brian A. Nosek, Dorothy V. M. Bishop, Katherine S. Button, Christopher D. Chambers, Nathalie Percie du Sert, Uri Simonsohn, Eric-Jan Wagenmakers, Jennifer J. Ware, and John P. A. Ioannidis. 2017. "A Manifesto for Reproducible Science." *Nature Human Behaviour* 1 (1): 0021. https://doi.org/10.1038/s41562-016-0021.

Muniz, Jr., Albert M., and Thomas C. O'Guinn. 2001. "Brand Community." *Journal of Consumer Research* 27 (4): 412-32. https://doi.org/10.1086/319618.

Muraven, Mark, and Roy F. Baumeister. 2000. "Self-Regulation and Depletion of Limited Resources: Does Self-Control Resemble a Muscle?" *Psychological Bulletin* 126 (2): 247-59. https://doi.org/10.1037/0033-2909.126.2.247.

Murdock, Jr., Bennet B. 1962. "The Serial Position Effect of Free Recall." *Journal of Experimental Psychology* 64 (5): 482-88.

Myers, Stewart C. 1977. "Determinants of Corporate Borrowing." *Journal of Financial Economics* 5 (2): 147-75. https://doi.org/10.1016/0304-405X(77)90015-0.

Myers, Stewart C. 1984. "The Capital Structure Puzzle." *The Journal of Finance* 39 (3): 575-92. https://doi.org/10.1111/j.1540-6261.1984.tb03646.x.

Myers, Stewart C. 2001. "Capital Structure." *Journal of Economic Perspectives* 15 (2): 81-102. https://doi.org/10.1257/jep.15.2.81.

Myers, Stewart C., and Nicholas S. Majluf. 1984. "Corporate Financing and Investment Decisions When Firms Have Information That Investors Do Not Have." *Journal of Financial Economics* 13 (2): 187-221. https://doi.org/10.1016/0304-405X(84)90023-0.

Myerson, Roger B. 1981. "Optimal Auction Design." *Mathematics of Operations Research* 6 (1): 58-73. https://doi.org/10.1287/moor.6.1.58.

Nadella, Satya, Greg Shaw, and Jill Tracie Nichols. 2017. *Hit Refresh: The Quest to Rediscover Microsoft's Soul and Imagine a Better Future for Everyone*. New York: Harper Business.

Nahapiet, Janine, and Sumantra Ghoshal. 1998. "Social Capital, Intellectual Capital, and the Organizational Advantage." *Academy of Management Review* 23 (2): 242-66. https://doi.org/10.2307/259373.

Narver, John C., and Stanley F. Slater. 1990. "The Effect of a Market Orientation on Business Profitability." *Journal of Marketing* 54 (4): 20-35. https://doi.org/10.1177/002224299005400403.

Narver, John C., Stanley F. Slater, and Douglas L. MacLachlan. 2004. "Responsive and Proactive Market Orientation and New-Product Success." *Journal of Product Innovation Management* 21 (5): 334-47. https://doi.org/10.1111/j.0737-6782.2004.00086.x.

Nash, John. 1951. "Non-Cooperative Games." *Annals of Mathematics* 54 (2): 286-95. https://doi.org/10.2307/1969529.

Nash, John F. 1950. "Equilibrium Points in n-Person Games." *Proceedings of the National Academy of Sciences* 36 (1): 48-49. https://doi.org/10.1073/pnas.36.1.48.

Neale, Margaret A., and Max H. Bazerman. 1985. "The Effects of Framing and Negotiator Overconfidence on Bargaining Behaviors and Outcomes." *Academy of Management Journal* 28 (1): 34-49. https://doi.org/10.2307/256060.

Nelson, Leif D., and Tom Meyvis. 2008. "Interrupted Consumption: Disrupting Adaptation to Hedonic Experiences." *Journal of Marketing Research* 45 (6): 654-64.

Nelson, Leif D., Tom Meyvis, and Jeff Galak. 2009. "Enhancing the Television-Viewing Experience Through Commercial Interruptions." *Journal of Consumer Research* 36 (2): 160-72.

Nelson, Phillip. 1974. "Advertising as Information." *Journal of Political Economy* 82 (4): 729-54. https://doi.org/10.1086/260231.

Nelson, Richard R., and Sidney G. Winter. 1982. *An Evolutionary Theory of Economic Change*. Cambridge, MA: Belknap Press of Harvard University Press.

Nembhard, Ingrid M., and Amy C. Edmondson. 2006. "Making It Safe: The Effects of Leader Inclusiveness and Professional Status on Psychological Safety and Improvement Efforts in Health Care Teams." *Journal of Organizational Behavior* 27 (7): 941-66. https://doi.org/10.1002/job.413.

Neumann, John von, and Oskar Morgenstern. 1944. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press.

Newbert, Scott L. 2007. "Empirical Research on the Resource-Based View of the Firm: An Assessment and Suggestions for Future Research." *Strategic Management Journal* 28 (2): 121-46. https://doi.org/10.1002/smj.573.

Noelle-Neumann, Elisabeth. 1974. "The Spiral of Silence: A Theory of Public Opinion." *Journal of Communication* 24 (2): 43-51. https://doi.org/10.1111/j.1460-2466.1974.tb00367.x.

Nohria, Nitin, and Ranjay Gulati. 1996. "Is Slack Good or Bad for Innovation?" *Academy of Management Journal* 39 (5): 1245-64. https://doi.org/10.2307/256998.

Nolan, Jessica M., P. Wesley Schultz, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2008. "Normative Social Influence Is Underdetected." *Personality and Social Psychology Bulletin* 34 (7): 913-23. https://doi.org/10.1177/0146167208316691.

Nonaka, Ikujiro. 1994. "A Dynamic Theory of Organizational Knowledge Creation." *Organization Science* 5 (1): 14-37. https://doi.org/10.1287/orsc.5.1.14.

Nonaka, Ikujiro, and Georg von Krogh. 2009. "Tacit Knowledge and Knowledge Conversion: Controversy and Advancement in Organizational Knowledge Creation Theory." *Organization Science* 20 (3): 635-52. https://doi.org/10.1287/orsc.1080.0412.

Nonaka, Ikujiro, and Hirotaka Takeuchi. 1995. *The Knowledge-Creating Company: How Japanese Companies Create the Dynamics of Innovation*. New York: Oxford University Press.

North, Douglass C. 1990. *Institutions, Institutional Change and Economic Performance*. Cambridge: Cambridge University Press.

North, Douglass C. 1991. "Institutions." *Journal of Economic Perspectives* 5 (1): 97-112. https://doi.org/10.1257/jep.5.1.97.

North, Douglass C., and Barry R. Weingast. 1989. "Constitutions and Commitment: The Evolution of Institutions Governing Public Choice in Seventeenth-Century England." *The Journal of Economic History* 49 (4): 803-32. https://doi.org/10.1017/S0022050700009451.

Northcraft, Gregory B., and Margaret A. Neale. 1987. "Experts, Amateurs, and Real Estate: An Anchoring-and-Adjustment Perspective on Property Pricing Decisions." *Organizational Behavior and Human Decision Processes* 39 (1): 84-97. https://doi.org/10.1016/0749-5978(87)90046-X.

Norton, Michael I., Daniel Mochon, and Dan Ariely. 2012. "The IKEA Effect: When Labor Leads to

Love." *Journal of Consumer Psychology* 22 (3): 453-60. https://doi.org/10.1016/j.jcps.2011.08.002.

Nosek, Brian A., Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor. 2018. "The Preregistration Revolution." *Proceedings of the National Academy of Sciences* 115 (11): 2600-2606. https://doi.org/10.1073/pnas.1708274114.

Novemsky, Nathan, Ravi Dhar, Norbert Schwarz, and Itamar Simonson. 2007. "Preference Fluency in Choice." *Journal of Marketing Research* 44 (3): 347-56. https://doi.org/10.1509/jmkr.44.3.347.

Novemsky, Nathan, and Daniel Kahneman. 2005. "The Boundaries of Loss Aversion." *Journal of Marketing Research* 42 (2): 119-28.

Nunes, Joseph C., and Xavier Drèze. 2006. "The Endowed Progress Effect: How Artificial Advancement Increases Effort." *Journal of Consumer Research* 32 (4): 504-12. https://doi.org/10.1086/500480.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89 (1): 103-24. https://doi.org/10.1257/aer.89.1.103.

O'Donoghue, Ted, and Charles Sprenger. 2018. "Reference-Dependent Preferences." In *Handbook of Behavioral Economics: Applications and Foundations 1*, edited by B. Douglas Bernheim, Stefano DellaVigna, and David Laibson, 1:1-77. North-Holland. https://doi.org/10.1016/bs.hesbe.2018.07.003.

O'Keefe, Daniel J., and Jakob D. Jensen. 2007. "The Relative Persuasiveness of Gain-Framed and Loss-Framed Messages for Encouraging Disease Prevention Behaviors: A Meta-Analytic Review." *Journal of Health Communication* 12 (7): 623-44.

O'Reilly, Charles A., Jennifer Chatman, and David F. Caldwell. 1991. "People and Organizational Culture: A Profile Comparison Approach to Assessing Person-Organization Fit." *Academy of Management Journal* 34 (3): 487-516. https://doi.org/10.2307/256404.

O'Reilly, III, Charles A., and Michael L. Tushman. 2004. "The Ambidextrous Organization." *Harvard Business Review* 82 (4): 74-81.

O'Reilly, III, Charles A., and Michael L. Tushman. 2013. "Organizational Ambidexterity: Past, Present, and Future." *Academy of Management Perspectives* 27 (4): 324-38. https://doi.org/10.5465/amp.2013.0025.

Obstfeld, David. 2005. "Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation." *Administrative Science Quarterly* 50 (1): 100-130. https://doi.org/10.2189/asqu.2005.50.1.100.

Ocasio, William. 1997. "Towards an Attention-Based View of the Firm." *Strategic Management Journal* 18 (S1): 187-206.

Ocasio, William. 2011. "Attention to Attention." *Organization Science* 22 (5): 1286-96. https://doi.org/10.1287/orsc.1100.0602.

Odean, Terrance. 1998. "Are Investors Reluctant to Realize Their Losses?" *The Journal of Finance* 53 (5): 1775-98. https://doi.org/10.1111/0022-1082.00072.

Ohno, Taiichi. 1988. *Toyota Production System: Beyond Large-Scale Production*. Cambridge, MA: Productivity Press.

Okada, Erica Mina. 2005. "Justification Effects on Consumer Choice of Hedonic and Utilitarian Goods." *Journal of Marketing Research* 42 (1): 43-53. https://doi.org/10.1509/jmkr.42.1.43.56889.

Oliver, Christine. 1991. "Strategic Responses to Institutional Processes." *Academy of Management Review* 16 (1): 145-79. https://doi.org/10.2307/258610.

Oliver, Richard L. 1980. "A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions." *Journal of Marketing Research* 17 (4): 460-69. https://doi.org/10.2307/3150499.

Oliver, Richard L. 1993. "Cognitive, Affective, and Attribute Bases of the Satisfaction Response." *Journal of Consumer Research* 20 (3): 418-30. https://doi.org/10.1086/209358.

Olson, Mancur. 1965. *The Logic of Collective Action: Public Goods and the Theory of Groups*. Cambridge, MA: Harvard University Press.

Oosterbeek, Hessel, Randolph Sloof, and Gijs van de Kuilen. 2004. "Cultural Differences in Ultimatum Game Experiments: Evidence from a Meta-Analysis." *Experimental Economics* 7 (2): 171-88. https://doi.org/10.1023/B:EXEC.0000026978.14316.74.

Open Science Collaboration. 2015. "Estimating the Reproducibility of Psychological Science." *Science* 349 (6251): aac4716. https://doi.org/10.1126/science.aac4716.

Ordóñez, Lisa D., Maurice E. Schweitzer, Adam D. Galinsky, and Max H. Bazerman. 2009. "Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting." *Academy of Management Perspectives* 23 (1): 6-16.

Ostrom, Elinor. 1990. *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.

Ostrom, Elinor. 2000. "Collective Action and the Evolution of Social Norms." *Journal of Economic Perspectives* 14 (3): 137-58. https://doi.org/10.1257/jep.14.3.137.

Oswald, Andrew J., Eugenio Proto, and Daniel Sgroi. 2015. "Happiness and Productivity." *Journal of Labor Economics* 33 (4): 789-822.

Oviatt, Benjamin M., and Patricia Phillips McDougall. 1994. "Toward a Theory of International New Ventures." *Journal of International Business Studies* 25 (1): 45-64. https://doi.org/10.1057/palgrave.jibs.8490193.

Oyer, Paul. 1998. "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality." *The Quarterly Journal of Economics* 113 (1): 149-85.

Oyserman, Daphna, Heather M. Coon, and Markus Kemmelmeier. 2002. "Rethinking Individualism and Collectivism: Evaluation of Theoretical Assumptions and Meta-Analyses." *Psychological Bulletin* 128 (1): 3-72.

Page, Scott E. 2007. *The Difference: How the Power of Diversity Creates Better Groups, Firms, Schools, and Societies*. Princeton, NJ: Princeton University Press.

Palmatier, Robert W., Rajiv P. Dant, Dhruv Grewal, and Kenneth R. Evans. 2006. "Factors Influencing the Effectiveness of Relationship Marketing: A Meta-Analysis." *Journal of Marketing* 70 (4): 136-53. https://doi.org/10.1509/jmkg.70.4.136.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry. 1985. "A Conceptual Model of Service Quality and Its Implications for Future Research." *Journal of Marketing* 49 (4): 41-50. https://doi.org/10.1177/002224298504900403.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry. 1988. "SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality." *Journal of Retailing* 64 (1): 12-40.

Parasuraman, A., Valarie A. Zeithaml, and Leonard L. Berry. 1994. "Reassessment of Expectations as a Comparison Standard in Measuring Service Quality: Implications for Further Research." *Journal of Marketing* 58 (1): 111-24. https://doi.org/10.1177/002224299405800109.

Park, C. Whan, Deborah J. MacInnis, Joseph Priester, Andreas B. Eisingerich, and Dawn Iacobucci. 2010. "Brand Attachment and Brand Attitude Strength: Conceptual and Empirical Differentiation of Two Critical Brand Equity Drivers." *Journal of Marketing* 74 (6): 1-17. https://doi.org/10.1509/jmkg.74.6.1.

Parker, Geoffrey G., and Marshall W. Van Alstyne. 2005. "Two-Sided Network Effects: A Theory of Information Product Design." *Management Science* 51 (10): 1494-504. https://doi.org/10.1287/mnsc.1050.0400.

Parker, Sharon K. 2014. "Beyond Motivation: Job and Work Design for Development, Health, Am-

bidexterity, and More." *Annual Review of Psychology* 65: 661-91. https://doi.org/10.1146/annurev-psych-010213-115208.

Parmar, Bidhan L., R. Edward Freeman, Jeffrey S. Harrison, Andrew C. Wicks, Lauren Purnell, and Simone de Colle. 2010. "Stakeholder Theory: The State of the Art." *Academy of Management Annals* 4 (1): 403-45. https://doi.org/10.5465/19416520.2010.495581.

Patel, Mitesh S., David A. Asch, Roy Rosin, Dylan S. Small, Scarlett L. Bellamy, Jack Heuer, Susan Sproat, et al. 2016. "Framing Financial Incentives to Increase Physical Activity Among Overweight and Obese Adults: A Randomized, Controlled Trial." *Annals of Internal Medicine* 164 (6): 385-94. https://doi.org/10.7326/M15-1635.

Payne, John W., James R. Bettman, and Eric J. Johnson. 1993. *The Adaptive Decision Maker*. Cambridge: Cambridge University Press.

Pearl, Judea. 2009. *Causality: Models, Reasoning, and Inference*. 2nd ed. Cambridge: Cambridge University Press.

Peck, Joann, and Suzanne B. Shu. 2009. "The Effect of Mere Touch on Perceived Ownership." *Journal of Consumer Research* 36 (3): 434-47.

Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles, and David G. Rand. 2021. "Shifting Attention to Accuracy Can Reduce Misinformation Online." *Nature* 592: 590-95. https://doi.org/10.1038/s41586-021-03344-2.

Penrose, Edith T. 1959. *The Theory of the Growth of the Firm*. Oxford: Basil Blackwell.

Peracchio, Laura A., and Alice M. Tybout. 1996. "The Moderating Role of Prior Knowledge in Schema-Based Product Evaluation." *Journal of Consumer Research* 23 (3): 177-92. https://doi.org/10.1086/209476.

Peres, Renana, Eitan Muller, and Vijay Mahajan. 2010. "Innovation Diffusion and New Product Growth Models: A Critical Review and Research Directions." *International Journal of Research in Marketing* 27 (2): 91-106. https://doi.org/10.1016/j.ijresmar.2009.12.012.

Pessin, Joseph. 1933. "The Comparative Effects of Social and Mechanical Stimulation on Memorizing." *The American Journal of Psychology* 45 (2): 263. https://doi.org/10.2307/1414277.

Peteraf, Margaret A. 1993. "The Cornerstones of Competitive Advantage: A Resource-Based View." *Strategic Management Journal* 14 (3): 179-91. https://doi.org/10.1002/smj.4250140303.

Peteraf, Margaret A., and Jay B. Barney. 2003. "Unraveling the Resource-Based Tangle." *Managerial and Decision Economics* 24 (4): 309-23. https://doi.org/10.1002/mde.1126.

Peteraf, Margaret, Giada Di Stefano, and Gianmario Verona. 2013. "The Elephant in the Room of Dynamic Capabilities: Bringing Two Diverging Conversations Together." *Strategic Management Journal* 34 (12): 1389-1410.

Petter, Stacie, William DeLone, and Ephraim McLean. 2008. "Measuring Information Systems Success: Models, Dimensions, Measures, and Interrelationships." *European Journal of Information Systems* 17 (3): 236-63. https://doi.org/10.1057/ejis.2008.15.

Petty, Richard E., and John T. Cacioppo. 1986. "The Elaboration Likelihood Model of Persuasion." *Advances in Experimental Social Psychology* 19: 123-205. https://doi.org/10.1016/S0065-2601(08)60214-2.

Petty, Richard E., John T. Cacioppo, and Rachel Goldman. 1981. "Personal Involvement as a Determinant of Argument-Based Persuasion." *Journal of Personality and Social Psychology* 41 (5): 847-55. https://doi.org/10.1037/0022-3514.41.5.847.

Petty, Richard E., John T. Cacioppo, and David Schumann. 1983. "Central and Peripheral Routes to Advertising Effectiveness: The Moderating Role of Involvement." *Journal of Consumer Research* 10 (2): 135-46. https://doi.org/10.1086/208954.

Pfeffer, Jeffrey. 1972. "Size and Composition of Corporate Boards of Directors: The Organization and Its Environment." *Administrative Science Quarterly* 17 (2): 218-28. https://doi.org/10.230

7/2393956.
Pfeffer, Jeffrey. 1993. "Barriers to the Advance of Organizational Science: Paradigm Development as a Dependent Variable." *Academy of Management Review* 18 (4): 599-620. https://doi.org/10.2307/258592.
Pfeffer, Jeffrey, and Gerald R. Salancik. 1978. *The External Control of Organizations: A Resource Dependence Perspective*. New York: Harper & Row.
Pfeffer, Jeffrey, and Robert I. Sutton. 2006. *Hard Facts, Dangerous Half-Truths, and Total Nonsense: Profiting from Evidence-Based Management*. Boston: Harvard Business School Press.
Pham, Michel Tuan. 1998. "Representativeness, Relevance, and the Use of Feelings in Decision Making." *Journal of Consumer Research* 25 (2): 144-59.
Pham, Michel Tuan. 2004. "The Logic of Feeling." *Journal of Consumer Psychology* 14 (4): 360-69.
Pham, Michel Tuan, Maggie Geuens, and Patrick De Pelsmacker. 2013. "The Influence of Ad-Evoked Feelings on Brand Evaluations: Empirical Generalizations from Consumers' Responses to More Than 1000 TV Commercials." *International Journal of Research in Marketing* 30 (4): 383-94.
Pierce, Jon L., Tatiana Kostova, and Kurt T. Dirks. 2001. "Toward a Theory of Psychological Ownership in Organizations." *Academy of Management Review* 26 (2): 298-310. https://doi.org/10.2307/259124.
Pierce, Jon L., Tatiana Kostova, and Kurt T. Dirks. 2003. "The State of Psychological Ownership: Integrating and Extending a Century of Research." *Review of General Psychology* 7 (1): 84-107. https://doi.org/10.1037/1089-2680.7.1.84.
Pigou, A. C. 1920. *The Economics of Welfare*. London: Macmillan.
Pisano, Gary P., Richard M. J. Bohmer, and Amy C. Edmondson. 2001. "Organizational Differences in Rates of Learning: Evidence from the Adoption of Minimally Invasive Cardiac Surgery." *Management Science* 47 (6): 752-68. https://doi.org/10.1287/mnsc.47.6.752.9811.
Platt, Jerome J., Tony Yaksh, and Charles L. Darby. 1967. "Social Facilitation of Eating Behavior in Armadillos." *Psychological Reports* 20 (3_suppl): 1136-36. https://doi.org/10.2466/pr0.1967.20.3c.1136.
Platt, John R. 1964. "Strong Inference." *Science* 146 (3642): 347-53. https://doi.org/10.1126/science.146.3642.347.
Plott, Charles R., and Kathryn Zeiler. 2005. "The Willingness to Pay-Willingness to Accept Gap, the 'Endowment Effect,' Subject Misconceptions, and Experimental Procedures for Eliciting Valuations." *American Economic Review* 95 (3): 530-45.
Polanyi, Michael. 1966. *The Tacit Dimension*. Garden City, NY: Doubleday.
Pope, Devin G., and Maurice E. Schweitzer. 2011. "Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes." *American Economic Review* 101 (1): 129-57.
Popper, Karl R. 1959. *The Logic of Scientific Discovery*. London: Hutchinson.
Porac, Joseph F., Howard Thomas, and Charles Baden-Fuller. 1989. "Competitive Groups as Cognitive Communities: The Case of Scottish Knitwear Manufacturers." *Journal of Management Studies* 26 (4): 397-416. https://doi.org/10.1111/j.1467-6486.1989.tb00736.x.
Porter, Lyman W., and Edward E. Lawler. 1968. *Managerial Attitudes and Performance*. Homewood, IL: Richard D. Irwin.
Porter, Michael E. 1980a. "Industry Structure and Competitive Strategy: Keys to Profitability." *Financial Analysts Journal* 36 (4): 30-41.
Porter, Michael E. 1979. "How Competitive Forces Shape Strategy." *Harvard Business Review* 57 (2): 137-45.
Porter, Michael E. 1980b. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.

Porter, Michael E. 1981. "The Contributions of Industrial Organization to Strategic Management." *The Academy of Management Review* 6 (4): 609-20. https://doi.org/10.2307/257639.

Porter, Michael E. 1996. "What Is Strategy?" *Harvard Business Review* 74 (6): 61-78.

Porter, Michael E. 2008. "The Five Competitive Forces That Shape Strategy." *Harvard Business Review* 86 (1): 78-93.

Porter, Michael E., and Mark R. Kramer. 2006. "Strategy and Society: The Link Between Competitive Advantage and Corporate Social Responsibility." *Harvard Business Review* 84 (12): 78-92.

Porter, Michael E., and Mark R. Kramer. 2011. "Creating Shared Value." *Harvard Business Review* 89 (1/2): 62-77.

Portes, Alejandro. 1998. "Social Capital: Its Origins and Applications in Modern Sociology." *Annual Review of Sociology* 24: 1-24. https://doi.org/10.1146/annurev.soc.24.1.1.

Post, Thierry, Martijn J. van den Assem, Guido Baltussen, and Richard H. Thaler. 2008. "Deal or No Deal? Decision Making Under Risk in a Large-Payoff Game Show." *American Economic Review* 98 (1): 38-71. https://doi.org/10.1257/aer.98.1.38.

Prahalad, C. K., and Gary Hamel. 1990. "The Core Competence of the Corporation." *Harvard Business Review* 68 (3): 79-91.

Preacher, Kristopher J., Derek D. Rucker, and Andrew F. Hayes. 2007. "Addressing Moderated Mediation Hypotheses: Theory, Methods, and Prescriptions." *Multivariate Behavioral Research* 42 (1): 185-227. https://doi.org/10.1080/00273170701341316.

Prelec, Drazen. 1998. "The Probability Weighting Function." *Econometrica* 66 (3): 497-527. https://doi.org/10.2307/2998573.

Prelec, Drazen, and George Loewenstein. 1998. "The Red and the Black: Mental Accounting of Savings and Debt." *Marketing Science* 17 (1): 4-28. https://doi.org/10.1287/mksc.17.1.4.

Prentice, Deborah A., and Dale T. Miller. 1993. "Pluralistic Ignorance and Alcohol Use on Campus: Some Consequences of Misperceiving the Social Norm." *Journal of Personality and Social Psychology* 64 (2): 243-56. https://doi.org/10.1037/0022-3514.64.2.243.

President's Working Group on Financial Markets. 1999. "Hedge Funds, Leverage, and the Lessons of Long-Term Capital Management." Washington, DC: U.S. Department of the Treasury.

Priem, Richard L., and John E. Butler. 2001. "Is the Resource-Based 'View' a Useful Perspective for Strategic Management Research?" *Academy of Management Review* 26 (1): 22-40.

Putnam, Robert D. 2000. *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon & Schuster.

Quiggin, John. 1982. "A Theory of Anticipated Utility." *Journal of Economic Behavior & Organization* 3 (4): 323-43. https://doi.org/10.1016/0167-2681(82)90008-7.

Quigley, Timothy J., and Donald C. Hambrick. 2015. "Has the 'CEO Effect' Increased in Recent Decades? A New Explanation for the Great Rise in America's Attention to Corporate Leaders." *Strategic Management Journal* 36 (6): 821-30. https://doi.org/10.1002/smj.2258.

Rabin, Matthew. 1993. "Incorporating Fairness into Game Theory and Economics." *American Economic Review* 83 (5): 1281-1302.

Rabin, Matthew. 2000. "Risk Aversion and Expected-Utility Theory: A Calibration Theorem." *Econometrica* 68 (5): 1281-92. https://doi.org/10.1111/1468-0262.00158.

Rabin, Matthew, and Richard H. Thaler. 2001. "Anomalies: Risk Aversion." *Journal of Economic Perspectives* 15 (1): 219-32. https://doi.org/10.1257/jep.15.1.219.

Raff, Daniel M. G., and Lawrence H. Summers. 1987. "Did Henry Ford Pay Efficiency Wages?" *Journal of Labor Economics* 5 (4): S57-86. https://doi.org/10.1086/298165.

Rains, Stephen A. 2013. "The Nature of Psychological Reactance Revisited: A Meta-Analytic Review." *Human Communication Research* 39 (1): 47-73. https://doi.org/10.1111/j.1468-2958.2012.01443.x.

Raisch, Sebastian, and Julian Birkinshaw. 2008. "Organizational Ambidexterity: Antecedents, Outcomes, and Moderators." *Journal of Management* 34 (3): 375-409. https://doi.org/10.1177/0149206308316058.

Rajkumar, Karthik, Guillaume Saint-Jacques, Iavor Bojinov, Erik Brynjolfsson, and Sinan Aral. 2022. "A Causal Test of the Strength of Weak Ties." *Science* 377 (6612): 1304-10. https://doi.org/10.1126/science.abl4476.

Raju, P. S. 1980. "Optimum Stimulation Level: Its Relationship to Personality, Demographics, and Exploratory Behavior." *Journal of Consumer Research* 7 (3): 272-82. https://doi.org/10.1086/208815.

Ramoglou, Stratos, and Eric W. K. Tsang. 2016. "A Realist Perspective of Entrepreneurship: Opportunities as Propensities." *Academy of Management Review* 41 (3): 410-34. https://doi.org/10.5465/amr.2014.0281.

Ratner, Rebecca K., Barbara E. Kahn, and Daniel Kahneman. 1999. "Choosing Less-Preferred Experiences for the Sake of Variety." *Journal of Consumer Research* 26 (1): 1-15. https://doi.org/10.1086/209547.

Ray, Gautam, Jay B. Barney, and Waleed A. Muhanna. 2004. "Capabilities, Business Processes, and Competitive Advantage: Choosing the Dependent Variable in Empirical Tests of the Resource-Based View." *Strategic Management Journal* 25 (1): 23-37.

Read, Daniel, and George Loewenstein. 1995. "Diversification Bias: Explaining the Discrepancy in Variety Seeking Between Combined and Separated Choices." *Journal of Experimental Psychology: Applied* 1 (1): 34-49. https://doi.org/10.1037/1076-898X.1.1.34.

Read, Stuart, Michael Song, and Willem Smit. 2009. "A Meta-Analytic Review of Effectuation and Venture Performance." *Journal of Business Venturing* 24 (6): 573-87. https://doi.org/10.1016/j.jbusvent.2008.02.005.

Reagans, Ray, and Bill McEvily. 2003. "Network Structure and Knowledge Transfer: The Effects of Cohesion and Range." *Administrative Science Quarterly* 48 (2): 240-67. https://doi.org/10.2307/3556658.

Reb, Jochen, and Terry Connolly. 2007. "Possession, Feelings of Ownership and the Endowment Effect." *Judgment and Decision Making* 2 (2): 107-14. https://doi.org/10.1017/S1930297500000085.

Reber, Rolf. 2011. "Processing Fluency, Aesthetic Pleasure, and Culturally Shared Taste." In, 223-42. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199732142.003.0055.

Reber, Rolf, Norbert Schwarz, and Piotr Winkielman. 2004. "Processing Fluency and Aesthetic Pleasure: Is Beauty in the Perceiver's Processing Experience?" *Personality and Social Psychology Review* 8 (4): 364-82. https://doi.org/10.1207/s15327957pspr0804_3.

Redelmeier, Donald A., and Daniel Kahneman. 1996. "Patients' Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures." *Pain* 66 (1): 3-8.

Redelmeier, Donald A., Joel Katz, and Daniel Kahneman. 2003. "Memories of Colonoscopy: A Randomized Trial." *Pain* 104 (1-2): 187-94.

Reed, Richard, and Robert J. DeFillippi. 1990. "Causal Ambiguity, Barriers to Imitation, and Sustainable Competitive Advantage." *Academy of Management Review* 15 (1): 88-102.

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *The Review of Economic Studies* 85 (2): 1251-78. https://doi.org/10.1093/restud/rdx038.

Reger, Rhonda K., and Anne Sigismund Huff. 1993. "Strategic Groups: A Cognitive Perspective." *Strategic Management Journal* 14 (2): 103-23. https://doi.org/10.1002/smj.4250140203.

Reichheld, Frederick F. 2003. "The One Number You Need to Grow." *Harvard Business Review* 81 (12): 46-54.

Reinartz, Werner J., and V. Kumar. 2000. "On the Profitability of Long-Life Customers in a Noncontractual Setting: An Empirical Investigation and Implications for Marketing." *Journal of Marketing* 64 (4): 17-35. https://doi.org/10.1509/jmkg.64.4.17.18077.

Repp, Bruno H. 1997. "The Aesthetic Quality of a Quantitatively Average Music Performance: Two Preliminary Experiments." *Music Perception* 14 (4): 419-44. https://doi.org/10.2307/40285732.

Reynolds, Thomas J., and Jonathan Gutman. 1988. "Laddering Theory, Method, Analysis, and Interpretation." *Journal of Advertising Research* 28 (1): 11-31. https://doi.org/10.1080/00218499.1988.12467766.

Rhoades, Linda, and Robert Eisenberger. 2002. "Perceived Organizational Support: A Review of the Literature." *Journal of Applied Psychology* 87 (4): 698-714. https://doi.org/10.1037/0021-9010.87.4.698.

Rhodes, Gillian. 2006. "The Evolutionary Psychology of Facial Beauty." *Annual Review of Psychology* 57: 199-226. https://doi.org/10.1146/annurev.psych.57.102904.190208.

Ries, Eric. 2011. *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*. New York: Crown Business.

Riley, John G., and William F. Samuelson. 1981. "Optimal Auctions." *American Economic Review* 71 (3): 381-92.

Rivera, Lauren A. 2012. "Hiring as Cultural Matching: The Case of Elite Professional Service Firms." *American Sociological Review* 77 (6): 999-1022. https://doi.org/10.1177/0003122412463213.

Robinson, Sandra L., and Elizabeth Wolfe Morrison. 2000. "The Development of Psychological Contract Breach and Violation: A Longitudinal Study." *Journal of Organizational Behavior* 21 (5): 525-46.

Robinson, Sandra L., and Denise M. Rousseau. 1994. "Violating the Psychological Contract: Not the Exception but the Norm." *Journal of Organizational Behavior* 15 (3): 245-59. https://doi.org/10.1002/job.4030150306.

Robinson, W. S. 1950. "Ecological Correlations and the Behavior of Individuals." *American Sociological Review* 15 (3): 351-57. https://doi.org/10.2307/2087176.

Robinson, William T., and Claes Fornell. 1985. "Sources of Market Pioneer Advantages in Consumer Goods Industries." *Journal of Marketing Research* 22 (3): 305-17. https://doi.org/10.1177/002224378502200306.

Rochet, Jean-Charles, and Jean Tirole. 2003. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association* 1 (4): 990-1029. https://doi.org/10.1162/154247603322493212.

Rochet, Jean-Charles, and Jean Tirole. 2006. "Two-Sided Markets: A Progress Report." *The RAND Journal of Economics* 37 (3): 645-67. https://doi.org/10.1111/j.1756-2171.2006.tb00036.x.

Rodrik, Dani, Arvind Subramanian, and Francesco Trebbi. 2004. "Institutions Rule: The Primacy of Institutions over Geography and Integration in Economic Development." *Journal of Economic Growth* 9 (2): 131-65.

Roediger, III, Henry L., and Jeffrey D. Karpicke. 2006. "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention." *Psychological Science* 17 (3): 249-55.

Rogers, Everett M. 2003. *Diffusion of Innovations*. 5th ed. New York: Free Press.

Roll, Richard. 1977. "A Critique of the Asset Pricing Theory's Tests Part i: On Past and Potential Testability of the Theory." *Journal of Financial Economics* 4 (2): 129-76. https://doi.org/10.1016/0304-405X(77)90009-5.

Roll, Richard. 1986. "The Hubris Hypothesis of Corporate Takeovers." *The Journal of Business* 59 (2): 197-216. https://doi.org/10.1086/296325.

Rolls, Barbara J., Edmund T. Rolls, Edward A. Rowe, and Kevin Sweeney. 1981. "Sensory Specific Satiety in Man." *Physiology & Behavior* 27 (1): 137-42. https://doi.org/10.1016/0031-9384(81)90310-3.

Rolls, Edmund T. 2004. "The Functions of the Orbitofrontal Cortex." *Brain and Cognition* 55 (1): 11-29. https://doi.org/10.1016/s0278-2626(03)00277-x.

Rosen, Sherwin. 1981. "The Economics of Superstars." *The American Economic Review* 71 (5): 845-58.

Ross, Jerry, and Barry M. Staw. 1993. "Organizational Escalation and Exit: Lessons from the Shoreham Nuclear Power Plant." *Academy of Management Journal* 36 (4): 701-32. https://doi.org/10.2307/256756.

Ross, Lee. 1977. "The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process." In *Advances in Experimental Social Psychology*, edited by Leonard Berkowitz, 10:173-220. New York: Academic Press. https://doi.org/10.1016/S0065-2601(08)60357-3.

Ross, Stephen A. 1973. "The Economic Theory of Agency: The Principal's Problem." *The American Economic Review* 63 (2): 134-39.

Ross, Stephen A. 1977. "The Determination of Financial Structure: The Incentive-Signalling Approach." *The Bell Journal of Economics* 8 (1): 23-40. https://doi.org/10.2307/3003485.

Rothman, Alexander J., and Peter Salovey. 1997. "Shaping Perceptions to Motivate Healthy Behavior: The Role of Message Framing." *Psychological Bulletin* 121 (1): 3-19. https://doi.org/10.1037/0033-2909.121.1.3.

Rothschild, Michael, and Joseph Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *The Quarterly Journal of Economics* 90 (4): 629-49. https://doi.org/10.2307/1885326.

Rousseau, Denise M. 1985. "Issues of Level in Organizational Research: Multi-Level and Cross-Level Perspectives." *Research in Organizational Behavior* 7: 1-37.

Rousseau, Denise M. 1989. "Psychological and Implied Contracts in Organizations." *Employee Responsibilities and Rights Journal* 2 (2): 121-39. https://doi.org/10.1007/BF01384942.

Rousseau, Denise M. 1995. *Psychological Contracts in Organizations: Understanding Written and Unwritten Agreements*. Thousand Oaks, CA: Sage.

Rousseau, Denise M. 2006. "Is There Such a Thing as "Evidence-Based Management"?" *Academy of Management Review* 31 (2): 256-69. https://doi.org/10.5465/amr.2006.20208679.

Rubin, Paul H. 1978. "The Theory of the Firm and the Structure of the Franchise Contract." *The Journal of Law and Economics* 21 (1): 223-33. https://doi.org/10.1086/466918.

Ruggeri, Kai, Sonia Alí, Mari Louise Berge, Giulia Bertoldo, Ludvig D. Bjørndal, Anna Cortijos-Bernabeu, Clair Davison, et al. 2020. "Replicating Patterns of Prospect Theory for Decision Under Risk." *Nature Human Behaviour* 4 (6): 622-33. https://doi.org/10.1038/s41562-020-0886-x.

Ruggiero, Thomas E. 2000. "Uses and Gratifications Theory in the 21st Century." *Mass Communication and Society* 3 (1): 3-37. https://doi.org/10.1207/S15327825MCS0301_02.

Rugman, Alan M. 1981. *Inside the Multinationals: The Economics of Internal Markets*. London: Croom Helm.

Rugman, Alan M., and Alain Verbeke. 2004. "A Perspective on Regional and Global Strategies of Multinational Enterprises." *Journal of International Business Studies* 35 (1): 3-18. https://doi.org/10.1057/palgrave.jibs.8400073.

Rumelt, Richard P. 1984. "Towards a Strategic Theory of the Firm." In *Competitive Strategic Management*, edited by Robert Boyden Lamb, 556-70. Englewood Cliffs, NJ: Prentice-Hall.

Rumelt, Richard P. 1991. "How Much Does Industry Matter?" *Strategic Management Journal* 12 (3): 167-85.

Russell, James A. 1980. "A Circumplex Model of Affect." *Journal of Personality and Social Psychology* 39 (6): 1161-78. https://doi.org/10.1037/h0077714.

Russell, James A. 2003. "Core Affect and the Psychological Construction of Emotion." *Psychological Review* 110 (1): 145-72. https://doi.org/10.1037/0033-295X.110.1.145.

Rust, Roland T., Katherine N. Lemon, and Valarie A. Zeithaml. 2004. "Return on Marketing: Using Customer Equity to Focus Marketing Strategy." *Journal of Marketing* 68 (1): 109-27. https://doi.org/10.1509/jmkg.68.1.109.24030.

Ryan, Bryce, and Neal C. Gross. 1943. "The Diffusion of Hybrid Seed Corn in Two Iowa Communities." *Rural Sociology* 8 (1): 15-24.

Ryan, Michelle K., and S. Alexander Haslam. 2005. "The Glass Cliff: Evidence That Women Are over-Represented in Precarious Leadership Positions." *British Journal of Management* 16 (2): 81-90. https://doi.org/10.1111/j.1467-8551.2005.00433.x.

Ryan, Michelle K., S. Alexander Haslam, Mette D. Hersby, and Renata Bongiorno. 2011. "Think Crisis-Think Female: The Glass Cliff and Contextual Variation in the Think Manager-Think Male Stereotype." *Journal of Applied Psychology* 96 (3): 470-84. https://doi.org/10.1037/a0022133.

Ryan, Michelle K., S. Alexander Haslam, Thekla Morgenroth, Floor Rink, Janka Stoker, and Kim Peters. 2016. "Getting on Top of the Glass Cliff: Reviewing a Decade of Evidence, Explanations, and Impact." *The Leadership Quarterly* 27 (3): 446-55. https://doi.org/10.1016/j.leaqua.2015.10.008.

Ryan, Richard M., and Edward L. Deci. 2000. "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being." *American Psychologist* 55 (1): 68-78. https://doi.org/10.1037/0003-066X.55.1.68.

Rysman, Marc. 2009. "The Economics of Two-Sided Markets." *Journal of Economic Perspectives* 23 (3): 125-43. https://doi.org/10.1257/jep.23.3.125.

Salancik, Gerald R., and Jeffrey Pfeffer. 1978. "A Social Information Processing Approach to Job Attitudes and Task Design." *Administrative Science Quarterly* 23 (2): 224-53. https://doi.org/10.2307/2392563.

Salganik, Matthew J., Peter Sheridan Dodds, and Duncan J. Watts. 2006. "Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market." *Science* 311 (5762): 854-56. https://doi.org/10.1126/science.1121066.

Saloner, Garth. 1991. "Modeling, Game Theory, and Strategic Management." *Strategic Management Journal* 12 (S2): 119-36. https://doi.org/10.1002/smj.4250121009.

Samuelson, Paul A. 1954. "The Pure Theory of Public Expenditure." *The Review of Economics and Statistics* 36 (4): 387-89. https://doi.org/10.2307/1925895.

Samuelson, Paul A. 1974. "Challenge to Judgment." *The Journal of Portfolio Management* 1 (1): 17-19. https://doi.org/10.3905/jpm.1974.408496.

Samuelson, William, and Richard Zeckhauser. 1988. "Status Quo Bias in Decision Making." *Journal of Risk and Uncertainty* 1 (1): 7-59. https://doi.org/10.1007/BF00055564.

Sanders, Glenn S., Robert S. Baron, and Danny L. Moore. 1978. "Distraction and Social Comparison as Mediators of Social Facilitation Effects." *Journal of Experimental Social Psychology* 14 (3): 291-303. https://doi.org/10.1016/0022-1031(78)90017-3.

Sarasvathy, Saras D. 2001. "Causation and Effectuation: Toward a Theoretical Shift from Economic Inevitability to Entrepreneurial Contingency." *Academy of Management Review* 26 (2): 243-63. https://doi.org/10.2307/259121.

Sarasvathy, Saras D. 2008. *Effectuation: Elements of Entrepreneurial Expertise*. Cheltenham, UK: Edward Elgar.

Savage, Leonard J. 1954. *The Foundations of Statistics*. New York: Wiley.

Schad, Jonathan, Marianne W. Lewis, Sebastian Raisch, and Wendy K. Smith. 2016. "Paradox Re-

search in Management Science: Looking Back to Move Forward." *Academy of Management Annals* 10 (1): 5-64. https://doi.org/10.5465/19416520.2016.1162422.

Schaufeli, Wilmar B., and Arnold B. Bakker. 2004. "Job Demands, Job Resources, and Their Relationship with Burnout and Engagement: A Multi-Sample Study." *Journal of Organizational Behavior* 25 (3): 293-315. https://doi.org/10.1002/job.248.

Schaufeli, Wilmar B., Arnold B. Bakker, and Marisa Salanova. 2006. "The Measurement of Work Engagement with a Short Questionnaire: A Cross-National Study." *Educational and Psychological Measurement* 66 (4): 701-16. https://doi.org/10.1177/0013164405282471.

Scheibehenne, Benjamin, Rainer Greifeneder, and Peter M. Todd. 2010. "Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload." *Journal of Consumer Research* 37 (3): 409-25. https://doi.org/10.1086/651235.

Schein, Edgar H. 1990. "Organizational Culture." *American Psychologist* 45 (2): 109-19. https://doi.org/10.1037/0003-066X.45.2.109.

Schein, Edgar H. 2010. *Organizational Culture and Leadership*. 4th ed. San Francisco: Jossey-Bass.

Schelling, Thomas C. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Scheufele, Dietram A., and David Tewksbury. 2007. "Framing, Agenda Setting, and Priming: The Evolution of Three Media Effects Models." *Journal of Communication* 57 (1): 9-20. https://doi.org/10.1111/j.0021-9916.2007.00326.x.

Schilke, Oliver, Songcui Hu, and Constance E. Helfat. 2018. "Quo Vadis, Dynamic Capabilities? A Content-Analytic Review of the Current State of Knowledge and Recommendations for Future Research." *Academy of Management Annals* 12 (1): 390-439. https://doi.org/10.5465/annals.2016.0014.

Schmalensee, Richard. 1985. "Do Markets Differ Much?" *The American Economic Review* 75 (3): 341-51.

Schmalensee, Richard, and Robert N. Stavins. 2013. "The SO2 Allowance Trading System: The Ironic History of a Grand Policy Experiment." *Journal of Economic Perspectives* 27 (1): 103-22. https://doi.org/10.1257/jep.27.1.103.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo. 1987. "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science* 33 (1): 1-24. https://doi.org/10.1287/mnsc.33.1.1.

Schneider, Benjamin. 1987. "The People Make the Place." *Personnel Psychology* 40 (3): 437-53. https://doi.org/10.1111/j.1744-6570.1987.tb00609.x.

Schneider, Benjamin, Harold W. Goldstein, and D. Brent Smith. 1995. "The ASA Framework: An Update." *Personnel Psychology* 48 (4): 747-73. https://doi.org/10.1111/j.1744-6570.1995.tb01780.x.

Schouten, John W., and James H. McAlexander. 1995. "Subcultures of Consumption: An Ethnography of the New Bikers." *Journal of Consumer Research* 22 (1): 43-61. https://doi.org/10.1086/209434.

Schultz, P. Wesley, Jessica M. Nolan, Robert B. Cialdini, Noah J. Goldstein, and Vladas Griskevicius. 2007. "The Constructive, Destructive, and Reconstructive Power of Social Norms." *Psychological Science* 18 (5): 429-34. https://doi.org/10.1111/j.1467-9280.2007.01917.x.

Schumpeter, Joseph A. 1934. *The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle*. Cambridge, MA: Harvard University Press.

Schumpeter, Joseph A. 1942. *Capitalism, Socialism and Democracy*. New York: Harper & Brothers.

Schwartz, Christine R., and Robert D. Mare. 2005. "Trends in Educational Assortative Marriage from 1940 to 2003." *Demography* 42 (4): 621-46. https://doi.org/10.1353/dem.2005.0036.

Schwartz, Mark S., and Archie B. Carroll. 2003. "Corporate Social Responsibility: A Three-Domain Approach." *Business Ethics Quarterly* 13 (4): 503-30. https://doi.org/10.5840/beq200313435.

Schwartz, Shalom H. 1992. "Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries." *Advances in Experimental Social Psychology* 25: 1-65. https://doi.org/10.1016/S0065-2601(08)60281-6.

Schwartz, Shalom H. 1999. "A Theory of Cultural Values and Some Implications for Work." *Applied Psychology: An International Review* 48 (1): 23-47.

Schwarz, Norbert. 2012. "Feelings-as-Information Theory." In *Handbook of Theories of Social Psychology*, edited by Paul A. M. Van Lange, Arie W. Kruglanski, and E. Tory Higgins, 1:289-308. London: SAGE. https://doi.org/10.4135/9781446249215.n15.

Schwarz, Norbert, Herbert Bless, Fritz Strack, Gisela Klumpp, Helga Rittenauer-Schatka, and Annette Simons. 1991. "Ease of Retrieval as Information: Another Look at the Availability Heuristic." *Journal of Personality and Social Psychology* 61 (2): 195-202. https://doi.org/10.1037/0022-3514.61.2.195.

Schwarz, Norbert, and Gerald L. Clore. 1983. "Mood, Misattribution, and Judgments of Well-Being: Informative and Directive Functions of Affective States." *Journal of Personality and Social Psychology* 45 (3): 513-23. https://doi.org/10.1037/0022-3514.45.3.513.

Schweitzer, Maurice E., Lisa Ordóñez, and Bambi Douma. 2004. "Goal Setting as a Motivator of Unethical Behavior." *Academy of Management Journal* 47 (3): 422-32.

Scott, W. Richard. 2014. *Institutions and Organizations: Ideas, Interests, and Identities*. 4th ed. Thousand Oaks, CA: Sage.

Seddon, Peter B. 1997. "A Respecification and Extension of the DeLone and McLean Model of IS Success." *Information Systems Research* 8 (3): 240-53. https://doi.org/10.1287/isre.8.3.240.

Selznick, Philip. 1949. *TVA and the Grass Roots: A Study in the Sociology of Formal Organization*. Berkeley: University of California Press.

Senyard, Julienne, Ted Baker, Paul Steffens, and Per Davidsson. 2014. "Bricolage as a Path to Innovativeness for Resource-Constrained New Firms." *Journal of Product Innovation Management* 31 (2): 211-30. https://doi.org/10.1111/jpim.12091.

Settoon, Randall P., Nathan Bennett, and Robert C. Liden. 1996. "Social Exchange in Organizations: Perceived Organizational Support, Leader-Member Exchange, and Employee Reciprocity." *Journal of Applied Psychology* 81 (3): 219-27. https://doi.org/10.1037/0021-9010.81.3.219.

Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.

Shalizi, Cosma Rohilla, and Andrew C. Thomas. 2011. "Homophily and Contagion Are Generically Confounded in Observational Social Network Studies." *Sociological Methods & Research* 40 (2): 211-39. https://doi.org/10.1177/0049124111404820.

Shane, Scott. 2000. "Prior Knowledge and the Discovery of Entrepreneurial Opportunities." *Organization Science* 11 (4): 448-69. https://doi.org/10.1287/orsc.11.4.448.14602.

Shane, Scott, and S. Venkataraman. 2000. "The Promise of Entrepreneurship as a Field of Research." *Academy of Management Review* 25 (1): 217-26. https://doi.org/10.5465/amr.2000.2791611.

Shapiro, Carl. 1989. "The Theory of Business Strategy." *The Rand Journal of Economics* 20 (1): 125-37.

Shapiro, Carl, and Joseph E. Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review* 74 (3): 433-44.

Sharpe, William F. 1964. "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk." *The Journal of Finance* 19 (3): 425-42. https://doi.org/10.1111/j.1540-6261.1964.tb02865.x.

Sharpe, William F. 1991. "The Arithmetic of Active Management." *Financial Analysts Journal* 47 (1): 7-9. https://doi.org/10.2469/faj.v47.n1.7.

Shefrin, Hersh, and Meir Statman. 1985. "The Disposition to Sell Winners Too Early and Ride Losers

Too Long: Theory and Evidence." *The Journal of Finance* 40 (3): 777-90. https://doi.org/10.1111/j.1540-6261.1985.tb05002.x.

Shelanski, Howard A., and Peter G. Klein. 1995. "Empirical Research in Transaction Cost Economics: A Review and Assessment." *Journal of Law, Economics, and Organization* 11 (2): 335-61.

Sheldon, Kennon M., and Sonja Lyubomirsky. 2012. "The Challenge of Staying Happier: Testing the Hedonic Adaptation Prevention Model." *Personality and Social Psychology Bulletin* 38 (5): 670-80.

Shenkar, Oded. 2001. "Cultural Distance Revisited: Towards a More Rigorous Conceptualization and Measurement of Cultural Differences." *Journal of International Business Studies* 32 (3): 519-35.

Sheppard, Blair H., Jon Hartwick, and Paul R. Warshaw. 1988. "The Theory of Reasoned Action: A Meta-Analysis of Past Research with Recommendations for Modifications and Future Research." *Journal of Consumer Research* 15 (3): 325-43. https://doi.org/10.1086/209170.

Sherif, Muzafer. 1936. *The Psychology of Social Norms*. New York: Harper.

Shiller, Robert J. 1981. "Do Stock Prices Move Too Much to Be Justified by Subsequent Changes in Dividends?" *The American Economic Review* 71 (3): 421-36.

Shimer, Robert. 2005. "The Cyclical Behavior of Equilibrium Unemployment and Vacancies." *American Economic Review* 95 (1): 25-49. https://doi.org/10.1257/0002828053828572.

Shinkle, George A. 2012. "Organizational Aspirations, Reference Points, and Goals: Building on the Past and Aiming for the Future." *Journal of Management* 38 (1): 415-55. https://doi.org/10.1177/0149206311419856.

Shiv, Baba, and Alexander Fedorikhin. 1999. "Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making." *Journal of Consumer Research* 26 (3): 278-92.

Shleifer, Andrei. 2000. *Inefficient Markets: An Introduction to Behavioral Finance*. Oxford: Oxford University Press.

Shleifer, Andrei, and Robert W. Vishny. 1997. "The Limits of Arbitrage." *The Journal of Finance* 52 (1): 35-55. https://doi.org/10.1111/j.1540-6261.1997.tb03807.x.

Shmueli, Galit. 2010. "To Explain or to Predict?" *Statistical Science* 25 (3): 289-310. https://doi.org/10.1214/10-STS330.

Shostack, G. Lynn. 1977. "Breaking Free from Product Marketing." *Journal of Marketing* 41 (2): 73-80. https://doi.org/10.1177/002224297704100219.

Shu, Suzanne B., and Joann Peck. 2011. "Psychological Ownership and Affective Reaction: Emotional Attachment Process Variables and the Endowment Effect." *Journal of Consumer Psychology* 21 (4): 439-52. https://doi.org/10.1016/j.jcps.2011.01.002.

Shyam-Sunder, Lakshmi, and Stewart C. Myers. 1999. "Testing Static Tradeoff Against Pecking Order Models of Capital Structure." *Journal of Financial Economics* 51 (2): 219-44. https://doi.org/10.1016/S0304-405X(98)00051-8.

Simmons, Joseph P., Leif D. Nelson, and Uri Simonsohn. 2011. "False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant." *Psychological Science* 22 (11): 1359-66. https://doi.org/10.1177/0956797611417632.

Simon, Herbert A. 1955. "A Behavioral Model of Rational Choice." *The Quarterly Journal of Economics* 69 (1): 99-118. https://doi.org/10.2307/1884852.

Simon, Herbert A. 1956. "Rational Choice and the Structure of the Environment." *Psychological Review* 63 (2): 129-38. https://doi.org/10.1037/h0042769.

Simonson, Itamar. 1989. "Choice Based on Reasons: The Case of Attraction and Compromise Effects." *Journal of Consumer Research* 16 (2): 158-74. https://doi.org/10.1086/209205.

Simonson, Itamar. 1990. "The Effect of Purchase Quantity and Timing on Variety-Seeking Behavior." *Journal of Marketing Research* 27 (2): 150-62. https://doi.org/10.2307/3172842.

Simonson, Itamar, and Ran Kivetz. 2018. "Bringing (Contingent) Loss Aversion down to Earth: A Comment on Gal and Rucker's Rejection of 'Losses Loom Larger Than Gains'." *Journal of Consumer Psychology* 28 (3): 517-22. https://doi.org/10.1002/jcpy.1046.

Simonson, Itamar, and Amos Tversky. 1992. "Choice in Context: Tradeoff Contrast and Extremeness Aversion." *Journal of Marketing Research* 29 (3): 281-95.

Singelis, Theodore M. 1994. "The Measurement of Independent and Interdependent Self-Construals." *Personality and Social Psychology Bulletin* 20 (5): 580-91.

Sirgy, M. Joseph. 1982. "Self-Concept in Consumer Behavior: A Critical Review." *Journal of Consumer Research* 9 (3): 287-300. https://doi.org/10.1086/208924.

Sirgy, M. Joseph, Dhruv Grewal, Tamara F. Mangleburg, Jae-ok Park, Kye-Sung Chon, C. B. Claiborne, J. S. Johar, and Harold Berkman. 1997. "Assessing the Predictive Validity of Two Methods of Measuring Self-Image Congruence." *Journal of the Academy of Marketing Science* 25 (3): 229-41. https://doi.org/10.1177/0092070397253004.

Sirmon, David G., Michael A. Hitt, and R. Duane Ireland. 2007. "Managing Firm Resources in Dynamic Environments to Create Value: Looking Inside the Black Box." *Academy of Management Review* 32 (1): 273-92.

Sleesman, Dustin J., Donald E. Conlon, Gerry McNamara, and Jonathan E. Miles. 2012. "Cleaning up the Big Muddy: A Meta-Analytic Review of the Determinants of Escalation of Commitment." *Academy of Management Journal* 55 (3): 541-62. https://doi.org/10.5465/amj.2010.0696.

Sleesman, Dustin J., Anna C. Lennard, Gerry McNamara, and Donald E. Conlon. 2018. "Putting Escalation of Commitment in Context: A Multilevel Review and Analysis." *Academy of Management Annals* 12 (1): 178-207. https://doi.org/10.5465/annals.2016.0046.

Slovic, Paul, Melissa L. Finucane, Ellen Peters, and Donald G. MacGregor. 2007. "The Affect Heuristic." *European Journal of Operational Research* 177 (3): 1333-52.

Smith, Amy K., Ruth N. Bolton, and Janet Wagner. 1999. "A Model of Customer Satisfaction with Service Encounters Involving Failure and Recovery." *Journal of Marketing Research* 36 (3): 356-72. https://doi.org/10.1177/002224379903600305.

Smith, Craig A., and Phoebe C. Ellsworth. 1985. "Patterns of Cognitive Appraisal in Emotion." *Journal of Personality and Social Psychology* 48 (4): 813-38. https://doi.org/10.1037/0022-3514.48.4.813.

Smith, Steven M., Thomas B. Ward, and Jay S. Schumacher. 1993. "Constraining Effects of Examples in a Creative Generation Task." *Memory & Cognition* 21 (6): 837-45.

Smith, Wendy K., and Marianne W. Lewis. 2011. "Toward a Theory of Paradox: A Dynamic Equilibrium Model of Organizing." *Academy of Management Review* 36 (2): 381-403. https://doi.org/10.5465/amr.2011.59330958.

Sniehotta, Falko F., Justin Presseau, and Vera Araújo-Soares. 2014. "Time to Retire the Theory of Planned Behaviour." *Health Psychology Review* 8 (1): 1-7. https://doi.org/10.1080/17437199.2013.869710.

Snyder, C. R., and Howard L. Fromkin. 1980. *Uniqueness: The Human Pursuit of Difference*. New York: Plenum Press. https://doi.org/10.1007/978-1-4684-3659-4.

Soderberg, Courtney K., Shannon P. Callahan, Annie O. Kochersberger, Elinor Amit, and Alison Ledgerwood. 2015. "The Effects of Psychological Distance on Abstraction: Two Meta-Analyses." *Psychological Bulletin* 141 (3): 525-48. https://doi.org/10.1037/bul0000005.

Solow, Robert M. 1979. "Another Possible Source of Wage Stickiness." *Journal of Macroeconomics* 1 (1): 79-82. https://doi.org/10.1016/0164-0704(79)90022-3.

Song, Hyunjin, and Norbert Schwarz. 2008. "If It's Hard to Read, It's Hard to Do: Processing Fluency Affects Effort Prediction and Motivation." *Psychological Science* 19 (10): 986-88. https://doi.org/10.1111/j.1467-9280.2008.02189.x.

Sørensen, Jesper B. 2002. "The Strength of Corporate Culture and the Reliability of Firm Performance." *Administrative Science Quarterly* 47 (1): 70-91. https://doi.org/10.2307/3094891.

Spangenberg, Eric R., Ayn E. Crowley, and Pamela W. Henderson. 1996. "Improving the Store Environment: Do Olfactory Cues Affect Evaluations and Behaviors?" *Journal of Marketing* 60 (2): 67-80.

Spear, Steven, and H. Kent Bowen. 1999. "Decoding the DNA of the Toyota Production System." *Harvard Business Review* 77 (5): 96-106.

Spence, Michael. 1973. "Job Market Signaling." *The Quarterly Journal of Economics* 87 (3): 355-74. https://doi.org/10.2307/1882010.

Spence, Michael. 2002. "Signaling in Retrospect and the Informational Structure of Markets." *American Economic Review* 92 (3): 434-59. https://doi.org/10.1257/00028280260136200.

Spencer, Steven J., Mark P. Zanna, and Geoffrey T. Fong. 2005. "Establishing a Causal Chain: Why Experiments Are Often More Effective Than Mediational Analyses in Examining Psychological Processes." *Journal of Personality and Social Psychology* 89 (6): 845-51. https://doi.org/10.1037/0022-3514.89.6.845.

Sperling, George. 1960. "The Information Available in Brief Visual Presentations." *Psychological Monographs: General and Applied* 74 (11): 1-29.

Spiller, Stephen A., Gavan J. Fitzsimons, Jr. Lynch John G., and Gary H. McClelland. 2013. "Spotlights, Floodlights, and the Magic Number Zero: Simple Effects Tests in Moderated Regression." *Journal of Marketing Research* 50 (2): 277-88. https://doi.org/10.1509/jmr.12.0420.

Spreng, Richard A., Scott B. MacKenzie, and Richard W. Olshavsky. 1996. "A Reexamination of the Determinants of Consumer Satisfaction." *Journal of Marketing* 60 (3): 15-32. https://doi.org/10.1177/002224299606000302.

Sprenger, Charles. 2015. "An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points." *Journal of Political Economy* 123 (6): 1456-99. https://doi.org/10.1086/683836.

Srinivasan, V., and Charlotte H. Mason. 1986. "Nonlinear Least Squares Estimation of New Product Diffusion Models." *Marketing Science* 5 (2): 169-78. https://doi.org/10.1287/mksc.5.2.169.

Srivastava, Rajendra K., Liam Fahey, and H. Kurt Christensen. 2001. "The Resource-Based View and Marketing: The Role of Market-Based Assets in Gaining Competitive Advantage." *Journal of Management* 27 (6): 777-802.

Srivastava, Rajendra K., Tasadduq A. Shervani, and Liam Fahey. 1998. "Market-Based Assets and Shareholder Value: A Framework for Analysis." *Journal of Marketing* 62 (1): 2-18. https://doi.org/10.1177/002224299806200102.

Stahl, Günter K., Martha L. Maznevski, Andreas Voigt, and Karsten Jonsen. 2010. "Unraveling the Effects of Cultural Diversity in Teams: A Meta-Analysis of Research on Multicultural Work Groups." *Journal of International Business Studies* 41 (4): 690-709.

Stajkovic, Alexander D., and Fred Luthans. 1998. "Self-Efficacy and Work-Related Performance: A Meta-Analysis." *Psychological Bulletin* 124 (2): 240-61. https://doi.org/10.1037/0033-2909.124.2.240.

Stalk, George, Philip Evans, and Lawrence E. Shulman. 1992. "Competing on Capabilities: The New Rules of Corporate Strategy." *Harvard Business Review* 70 (2): 57-69.

Stanovich, Keith E., and Richard F. West. 2000. "Individual Differences in Reasoning: Implications for the Rationality Debate?" *Behavioral and Brain Sciences* 23 (5): 645-65. https://doi.org/10.1017/S0140525X00003435.

Starbuck, William H., and Frances J. Milliken. 1988. "Challenger: Fine-Tuning the Odds Until Something Breaks." *Journal of Management Studies* 25 (4): 319-40. https://doi.org/10.1111/j.1467-6486.1988.tb00040.x.

Starmer, Chris. 2000. "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive

Theory of Choice Under Risk." *Journal of Economic Literature* 38 (2): 332-82. https://doi.org/10.1257/jel.38.2.332.

Stavins, Robert N. 1998. "What Can We Learn from the Grand Policy Experiment? Lessons from SO2 Allowance Trading." *Journal of Economic Perspectives* 12 (3): 69-88. https://doi.org/10.1257/jep.12.3.69.

Staw, Barry M. 1976. "Knee-Deep in the Big Muddy: A Study of Escalating Commitment to a Chosen Course of Action." *Organizational Behavior and Human Performance* 16 (1): 27-44. https://doi.org/10.1016/0030-5073(76)90005-2.

Staw, Barry M. 1981. "The Escalation of Commitment to a Course of Action." *Academy of Management Review* 6 (4): 577-87. https://doi.org/10.2307/257636.

Staw, Barry M., Pamela I. McKechnie, and Sheila M. Puffer. 1983. "The Justification of Organizational Performance." *Administrative Science Quarterly* 28 (4): 582-600. https://doi.org/10.2307/2393010.

Staw, Barry M., Robert I. Sutton, and Lisa H. Pelled. 1994. "Employee Positive Emotion and Favorable Outcomes at the Workplace." *Organization Science* 5 (1): 51-71. https://doi.org/10.1287/orsc.5.1.51.

Steel, Piers, and Cornelius J. König. 2006. "Integrating Theories of Motivation." *Academy of Management Review* 31 (4): 889-913.

Steele, Claude M. 1988. "The Psychology of Self-Affirmation: Sustaining the Integrity of the Self." In *Advances in Experimental Social Psychology*, edited by Leonard Berkowitz, 21:261-302. San Diego, CA: Academic Press. https://doi.org/10.1016/S0065-2601(08)60229-4.

Steenkamp, Jan-Benedict E. M., and Hans Baumgartner. 1992. "The Role of Optimum Stimulation Level in Exploratory Consumer Behavior." *Journal of Consumer Research* 19 (3): 434-48. https://doi.org/10.1086/209313.

Steindl, Christina, Eva Jonas, Sandra Sittenthaler, Eva Traut-Mattausch, and Jeff Greenberg. 2015. "Understanding Psychological Reactance: New Developments and Findings." *Zeitschrift für Psychologie* 223 (4): 205-14. https://doi.org/10.1027/2151-2604/a000222.

Stephen, Andrew T., Peter Pal Zubcsek, and Jacob Goldenberg. 2016. "Lower Connectivity Is Better: The Effects of Network Structure on Redundancy of Ideas and Customer Innovativeness in Interdependent Ideation Tasks." *Journal of Marketing Research* 53 (2): 263-79. https://doi.org/10.1509/jmr.13.0127.

Sterman, John D. 1989. "Modeling Managerial Behavior: Misperceptions of Feedback in a Dynamic Decision Making Experiment." *Management Science* 35 (3): 321-39. https://doi.org/10.1287/mnsc.35.3.321.

Stern, Robert N. 1979. "Review of The External Control of Organizations: A Resource Dependence Perspective, by Jeffrey Pfeffer and Gerald r. Salancik." *Contemporary Sociology* 8 (4): 612-13. https://doi.org/10.2307/2065200.

Stewart, Neil, Nick Chater, and Gordon D. A. Brown. 2006. "Decision by Sampling." *Cognitive Psychology* 53 (1): 1-26. https://doi.org/10.1016/j.cogpsych.2005.10.003.

Stigler, George J. 1961. "The Economics of Information." *Journal of Political Economy* 69 (3): 213-25. https://doi.org/10.1086/258464.

Stiglitz, J. E. 2000. "The Contributions of the Economics of Information to Twentieth Century Economics." *The Quarterly Journal of Economics* 115 (4): 1441-78. https://doi.org/10.1162/003355300555015.

Stinchcombe, Arthur L. 1965. "Social Structure and Organizations." In *Handbook of Organizations*, edited by James G. March, 142-93. Chicago: Rand McNally.

Stone, Jeff, Elliot Aronson, A. Lauren Crain, Matthew P. Winslow, and Carrie B. Fried. 1994. "Inducing Hypocrisy as a Means of Encouraging Young Adults to Use Condoms." *Personality and Social*

Psychology Bulletin 20 (1): 116-28. https://doi.org/10.1177/0146167294201012.

Strack, Fritz, and Thomas Mussweiler. 1997. "Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility." *Journal of Personality and Social Psychology* 73 (3): 437-46. https://doi.org/10.1037/0022-3514.73.3.437.

Strahilevitz, Michal, and John G. Myers. 1998. "Donations to Charity as Purchase Incentives: How Well They Work May Depend on What You Are Trying to Sell." *Journal of Consumer Research* 24 (4): 434-46. https://doi.org/10.1086/209519.

Strang, David, and Sarah A. Soule. 1998. "Diffusion in Organizations and Social Movements: From Hybrid Corn to Poison Pills." *Annual Review of Sociology* 24: 265-90. https://doi.org/10.1146/annurev.soc.24.1.265.

Strube, Michael J. 2005. "What Did Triplett Really Find? A Contemporary Analysis of the First Experiment in Social Psychology." *The American Journal of Psychology* 118 (2): 271-86. https://doi.org/10.2307/30039059.

Strzalko, Jan, and Katarzyna A. Kaszycka. 1992. "Physical Attractiveness: Interpersonal and Intrapersonal Variability of Assessments." *Biodemography and Social Biology* 39 (1-2): 170-76. https://doi.org/10.1080/19485565.1992.9988813.

Suarez, Fernando F., and Gianvito Lanzolla. 2007. "The Role of Environmental Dynamics in Building a First Mover Advantage Theory." *Academy of Management Review* 32 (2): 377-92. https://doi.org/10.5465/amr.2007.24349587.

Suárez, Fernando F., and James M. Utterback. 1995. "Dominant Designs and the Survival of Firms." *Strategic Management Journal* 16 (6): 415-30. https://doi.org/10.1002/smj.4250160602.

Suchman, Mark C. 1995. "Managing Legitimacy: Strategic and Institutional Approaches." *Academy of Management Review* 20 (3): 571-610. https://doi.org/10.2307/258788.

Suddaby, Roy. 2010. "Construct Clarity in Theories of Management and Organization." *Academy of Management Review* 35 (3): 346-57. https://doi.org/10.5465/amr.2010.51141319.

Suls, Jerry, René Martin, and Ladd Wheeler. 2002. "Social Comparison: Why, with Whom, and with What Effect?" *Current Directions in Psychological Science* 11 (5): 159-63. https://doi.org/10.1111/1467-8721.00191.

Sultan, Fareena, John U. Farley, and Donald R. Lehmann. 1990. "A Meta-Analysis of Applications of Diffusion Models." *Journal of Marketing Research* 27 (1): 70-77. https://doi.org/10.1177/002224379002700107.

Surowiecki, James. 2005. *The Wisdom of Crowds*. Anchor.

Sutton, Richard S., and Andrew G. Barto. 2018. *Reinforcement Learning: An Introduction*. 2nd ed. Cambridge, MA: MIT Press.

Sutton, Robert I., and Barry M. Staw. 1995. "What Theory Is Not." *Administrative Science Quarterly* 40 (3): 371-84. https://doi.org/10.2307/2393788.

Sweeney, Jillian C., Douglas Hausknecht, and Geoffrey N. Soutar. 2000. "Cognitive Dissonance After Purchase: A Multidimensional Scale." *Psychology and Marketing* 17 (5): 369-85.

Sweller, John. 1988. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science* 12 (2): 257-85.

Sy, Thomas, Stéphane Côté, and Richard Saavedra. 2005. "The Contagious Leader: Impact of the Leader's Mood on the Mood of Group Members, Group Affective Tone, and Group Processes." *Journal of Applied Psychology* 90 (2): 295-305. https://doi.org/10.1037/0021-9010.90.2.295.

Sydnor, Justin. 2010. "(Over)insuring Modest Risks." *American Economic Journal: Applied Economics* 2 (4): 177-99. https://doi.org/10.1257/app.2.4.177.

Szulanski, Gabriel. 1996. "Exploring Internal Stickiness: Impediments to the Transfer of Best Practice Within the Firm." *Strategic Management Journal* 17 (S2): 27-43. https://doi.org/10.1002/smj.4250171105.

Szymanski, David M., and David H. Henard. 2001. "Customer Satisfaction: A Meta-Analysis of the Empirical Evidence." *Journal of the Academy of Marketing Science* 29 (1): 16-35. https://doi.org/10.1177/009207030102900102.

Tajfel, Henri, M. G. Billig, R. P. Bundy, and Claude Flament. 1971. "Social Categorization and Intergroup Behaviour." *European Journal of Social Psychology* 1 (2): 149-78. https://doi.org/10.1002/ejsp.2420010202.

Tajfel, Henri, and John C. Turner. 1979. "An Integrative Theory of Intergroup Conflict." In *The Social Psychology of Intergroup Relations*, edited by William G. Austin and Stephen Worchel, 33-47. Monterey, CA: Brooks/Cole.

Tamir, Diana I., and Jason P. Mitchell. 2012. "Disclosing Information about the Self Is Intrinsically Rewarding." *Proceedings of the National Academy of Sciences* 109 (21): 8038-43. https://doi.org/10.1073/pnas.1202129109.

Taras, Vas, Bradley L. Kirkman, and Piers Steel. 2010. "Examining the Impact of Culture's Consequences: A Three-Decade, Multilevel, Meta-Analytic Review of Hofstede's Cultural Value Dimensions." *Journal of Applied Psychology* 95 (3): 405-39. https://doi.org/10.1037/a0018938.

Taylor, Paul J., Darlene F. Russ-Eft, and Daniel W. L. Chan. 2005. "A Meta-Analytic Review of Behavior Modeling Training." *Journal of Applied Psychology* 90 (4): 692-709. https://doi.org/10.1037/0021-9010.90.4.692.

Taylor, Shirley, and Peter A. Todd. 1995. "Understanding Information Technology Usage: A Test of Competing Models." *Information Systems Research* 6 (2): 144-76. https://doi.org/10.1287/isre.6.2.144.

Teas, R. Kenneth. 1993. "Expectations, Performance Evaluation, and Consumers' Perceptions of Quality." *Journal of Marketing* 57 (4): 18-34. https://doi.org/10.1177/002224299305700402.

Teece, David J. 1986. "Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing and Public Policy." *Research Policy* 15 (6): 285-305. https://doi.org/10.1016/0048-7333(86)90027-2.

Teece, David J. 2006. "Reflections on 'Profiting from Innovation'." *Research Policy* 35 (8): 1131-46. https://doi.org/10.1016/j.respol.2006.09.009.

Teece, David J. 2007. "Explicating Dynamic Capabilities: The Nature and Microfoundations of (Sustainable) Enterprise Performance." *Strategic Management Journal* 28 (13): 1319-50. https://doi.org/10.1002/smj.640.

Teece, David J. 2014. "The Foundations of Enterprise Performance: Dynamic and Ordinary Capabilities in an (Economic) Theory of Firms." *Academy of Management Perspectives* 28 (4): 328-52.

Teece, David J., Gary Pisano, and Amy Shuen. 1997. "Dynamic Capabilities and Strategic Management." *Strategic Management Journal* 18 (7): 509-33. https://doi.org/10.1002/(sici)1097-0266(199708)18:7%3C509::aid-smj882%3E3.0.co;2-z.

Teixeira, Thales, Michel Wedel, and Rik Pieters. 2012. "Emotion-Induced Engagement in Internet Video Advertisements." *Journal of Marketing Research* 49 (2): 144-59.

ter Hofstede, Frenkel, Anke Audenaert, Jan-Benedict E. M. Steenkamp, and Michel Wedel. 1998. "An Investigation into the Association Pattern Technique as a Quantitative Approach to Measuring Means-End Chains." *International Journal of Research in Marketing* 15 (1): 37-50. https://doi.org/10.1016/S0167-8116(97)00029-3.

Tesser, Abraham. 1988. "Toward a Self-Evaluation Maintenance Model of Social Behavior." *Advances in Experimental Social Psychology* 21: 181-227. https://doi.org/10.1016/S0065-2601(08)60227-0.

Thakral, Neil, and Linh T. Tô. 2021. "Daily Labor Supply and Adaptive Reference Points." *American Economic Review* 111 (8): 2417-43. https://doi.org/10.1257/aer.20170768.

Thaler, Richard. 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behav-*

ior & Organization 1 (1): 39-60. https://doi.org/10.1016/0167-2681(80)90051-7.
Thaler, Richard. 1985. "Mental Accounting and Consumer Choice." *Marketing Science* 4 (3): 199-214. https://doi.org/10.1287/mksc.4.3.199.
Thaler, Richard H. 1988. "Anomalies: The Winner's Curse." *Journal of Economic Perspectives* 2 (1): 191-202. https://doi.org/10.1257/jep.2.1.191.
Thaler, Richard H. 1999. "Mental Accounting Matters." *Journal of Behavioral Decision Making* 12 (3): 183-206.
Thaler, Richard H., and Shlomo Benartzi. 2004. "Save More Tomorrow: Using Behavioral Economics to Increase Employee Saving." *Journal of Political Economy* 112 (S1): S164-87. https://doi.org/10.1086/380085.
Thaler, Richard H., and Eric J. Johnson. 1990. "Gambling with the House Money and Trying to Break Even: The Effects of Prior Outcomes on Risky Choice." *Management Science* 36 (6): 643-60. https://doi.org/10.1287/mnsc.36.6.643.
Thaler, Richard H., and H. M. Shefrin. 1981. "An Economic Theory of Self-Control." *Journal of Political Economy* 89 (2): 392-406. https://doi.org/10.1086/260971.
Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.
Thaler, Richard H., Amos Tversky, Daniel Kahneman, and Alan Schwartz. 1997. "The Effect of Myopia and Loss Aversion on Risk Taking: An Experimental Test." *The Quarterly Journal of Economics* 112 (2): 647-61. https://doi.org/10.1162/003355397555226.
Thibaut, John W., and Harold H. Kelley. 1959. *The Social Psychology of Groups*. New York: John Wiley & Sons.
Thibaut, John, and Laurens Walker. 1975. *Procedural Justice: A Psychological Analysis*. Hillsdale, NJ: Lawrence Erlbaum.
Thomas, Joe, and Ricky Griffin. 1983. "The Social Information Processing Model of Task Design: A Review of the Literature." *Academy of Management Review* 8 (4): 672-82. https://doi.org/10.2307/258268.
Thompson, Abigail. 2014. "Does Diversity Trump Ability? An Example of the Misuse of Mathematics in the Social Sciences." *Notices of the American Mathematical Society* 61 (9): 1024-30.
Thompson, Craig J., Aric Rindfleisch, and Zeynep Arsel. 2006. "Emotional Branding and the Strategic Value of the Doppelgänger Brand Image." *Journal of Marketing* 70 (1): 50-64. https://doi.org/10.1509/jmkg.70.1.050.qxd.
Thompson, James D. 1967. *Organizations in Action: Social Science Bases of Administrative Theory*. New York: McGraw-Hill.
Thomson, Matthew, Deborah J. MacInnis, and C. Whan Park. 2005. "The Ties That Bind: Measuring the Strength of Consumers' Emotional Attachments to Brands." *Journal of Consumer Psychology* 15 (1): 77-91. https://doi.org/10.1207/s15327663jcp1501_10.
Thornhill, Randy, and Steven W. Gangestad. 1993. "Human Facial Beauty." *Human Nature* 4 (3): 237-69. https://doi.org/10.1007/bf02692201.
Tian, Kelly Tepper, William O. Bearden, and Gary L. Hunter. 2001. "Consumers' Need for Uniqueness: Scale Development and Validation." *Journal of Consumer Research* 28 (1): 50-66. https://doi.org/10.1086/321947.
Tiedens, Larissa Z., and Susan Linton. 2001. "Judgment Under Emotional Certainty and Uncertainty: The Effects of Specific Emotions on Information Processing." *Journal of Personality and Social Psychology* 81 (6): 973-88. https://doi.org/10.1037/0022-3514.81.6.973.
Till, Brian D., and Terence A. Shimp. 1998. "Endorsers in Advertising: The Case of Negative Celebrity Information." *Journal of Advertising* 27 (1): 67-82. https://doi.org/10.1080/00913367.1998.10673543.

Timmermans, Stefan, and Iddo Tavory. 2012. "Theory Construction in Qualitative Research: From Grounded Theory to Abductive Analysis." *Sociological Theory* 30 (3): 167-86. https://doi.org/10.1177/0735275112457914.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press.

Tobin, James. 1958. "Liquidity Preference as Behavior Towards Risk." *The Review of Economic Studies* 25 (2): 65-86. https://doi.org/10.2307/2296205.

Tolbert, Pamela S., and Lynne G. Zucker. 1983. "Institutional Sources of Change in the Formal Structure of Organizations: The Diffusion of Civil Service Reform, 1880-1935." *Administrative Science Quarterly* 28 (1): 22-39. https://doi.org/10.2307/2392383.

Tom, Sabrina M., Craig R. Fox, Christopher Trepel, and Russell A. Poldrack. 2007. "The Neural Basis of Loss Aversion in Decision-Making Under Risk." *Science* 315 (5811): 515-18. https://doi.org/10.1126/science.1134239.

Tornatzky, Louis G., and Katherine J. Klein. 1982. "Innovation Characteristics and Innovation Adoption-Implementation: A Meta-Analysis of Findings." *IEEE Transactions on Engineering Management* EM-29 (1): 28-45. https://doi.org/10.1109/TEM.1982.6447463.

Toubia, Olivier, and Andrew T. Stephen. 2013. "Intrinsic Vs. Image-Related Utility in Social Media: Why Do People Contribute Content to Twitter?" *Marketing Science* 32 (3): 368-92. https://doi.org/10.1287/mksc.2013.0773.

Travis, L. E. 1925. "The Effect of a Small Audience Upon Eye-Hand Coordination." *The Journal of Abnormal and Social Psychology* 20 (2): 142-46. https://doi.org/10.1037/h0071311.

Triandis, Harry C. 1995. *Individualism and Collectivism*. Boulder, CO: Westview Press.

Triandis, Harry C., and Michele J. Gelfand. 1998. "Converging Measurement of Horizontal and Vertical Individualism and Collectivism." *Journal of Personality and Social Psychology* 74 (1): 118-28.

Trigeorgis, Lenos. 1996. *Real Options: Managerial Flexibility and Strategy in Resource Allocation*. Cambridge, MA: MIT Press.

Trigeorgis, Lenos, and Jeffrey J. Reuer. 2017. "Real Options Theory in Strategic Management." *Strategic Management Journal* 38 (1): 42-63. https://doi.org/10.1002/smj.2593.

Triplett, Norman. 1898. "The Dynamogenic Factors in Pacemaking and Competition." *The American Journal of Psychology* 9 (4): 507. https://doi.org/10.2307/1412188.

Tripsas, Mary. 1997. "Unraveling the Process of Creative Destruction: Complementary Assets and Incumbent Survival in the Typesetter Industry." *Strategic Management Journal* 18 (S1): 119-42.

Tripsas, Mary. 2009. "Technology, Identity, and Inertia Through the Lens of 'The Digital Photography Company'." *Organization Science* 20 (2): 441-60. https://doi.org/10.1287/orsc.1080.0419.

Tripsas, Mary, and Giovanni Gavetti. 2000. "Capabilities, Cognition, and Inertia: Evidence from Digital Imaging." *Strategic Management Journal* 21 (10-11): 1147-61.

Trope, Yaacov, and Nira Liberman. 2003. "Temporal Construal." *Psychological Review* 110 (3): 403-21. https://doi.org/10.1037/0033-295X.110.3.403.

Trope, Yaacov, and Nira Liberman. 2010. "Construal-Level Theory of Psychological Distance." *Psychological Review* 117 (2): 440-63. https://doi.org/10.1037/a0018963.

Trope, Yaacov, Nira Liberman, and Cheryl Wakslak. 2007. "Construal Levels and Psychological Distance: Effects on Representation, Prediction, Evaluation, and Behavior." *Journal of Consumer Psychology* 17 (2): 83-95. https://doi.org/10.1016/S1057-7408(07)70013-X.

Tucker, Catherine. 2008. "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities." *Management Science* 54 (12): 2024-38. https://doi.org/10.1287/mnsc.1080.0897.

Tugade, Michele M., and Barbara L. Fredrickson. 2004. "Resilient Individuals Use Positive Emotions to Bounce Back from Negative Emotional Experiences." *Journal of Personality and Social Psychology* 86 (2): 320-33. https://doi.org/10.1037/0022-3514.86.2.320.

Tung, Rosalie L., and Alain Verbeke. 2010. "Beyond Hofstede and GLOBE: Improving the Quality of Cross-Cultural Research." *Journal of International Business Studies* 41 (8): 1259-74.

Turner, John C., Michael A. Hogg, Penelope J. Oakes, Stephen D. Reicher, and Margaret S. Wetherell. 1987. *Rediscovering the Social Group: A Self-Categorization Theory*. Oxford: Basil Blackwell.

Tushman, Michael L., and Philip Anderson. 1986. "Technological Discontinuities and Organizational Environments." *Administrative Science Quarterly* 31 (3): 439-65. https://doi.org/10.2307/2392832.

Tushman, Michael L., and III O'Reilly Charles A. 1996. "Ambidextrous Organizations: Managing Evolutionary and Revolutionary Change." *California Management Review* 38 (4): 8-30.

Tversky, Amos. 1977. "Features of Similarity." *Psychological Review* 84 (4): 327-52. https://doi.org/10.1037/0033-295x.84.4.327.

Tversky, Amos, and Daniel Kahneman. 1973. "Availability: A Heuristic for Judging Frequency and Probability." *Cognitive Psychology* 5 (2): 207-32. https://doi.org/10.1016/0010-0285(73)90033-9.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185 (4157): 1124-31. https://doi.org/10.1126/science.185.4157.1124.

Tversky, Amos, and Daniel Kahneman. 1981. "The Framing of Decisions and the Psychology of Choice." *Science* 211 (4481): 453-58. https://doi.org/10.1126/science.7455683.

Tversky, Amos, and Daniel Kahneman. 1983. "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment." *Psychological Review* 90 (4): 293-315. https://doi.org/10.1037/0033-295X.90.4.293.

Tversky, Amos, and Daniel Kahneman. 1986. "Rational Choice and the Framing of Decisions." *The Journal of Business* 59 (4): S251-78. https://doi.org/10.1086/296365.

Tversky, Amos, and Daniel Kahneman. 1991. "Loss Aversion in Riskless Choice: A Reference-Dependent Model." *The Quarterly Journal of Economics* 106 (4): 1039-61. https://doi.org/10.2307/2937956.

Tversky, Amos, and Daniel Kahneman. 1992. "Advances in Prospect Theory: Cumulative Representation of Uncertainty." *Journal of Risk and Uncertainty* 5 (4): 297-323. https://doi.org/10.1007/BF00122574.

Urbany, Joel E., William O. Bearden, and Dan C. Weilbaker. 1988. "The Effect of Plausible and Exaggerated Reference Prices on Consumer Perceptions and Price Search." *Journal of Consumer Research* 15 (1): 95-110. https://doi.org/10.1086/209148.

U.S. Securities and Exchange Commission. 2021. "Staff Report on Equity and Options Market Structure Conditions in Early 2021." Washington, DC: U.S. Securities and Exchange Commission.

Utterback, James M. 1994. *Mastering the Dynamics of Innovation: How Companies Can Seize Opportunities in the Face of Technological Change*. Boston, MA: Harvard Business School Press.

Utterback, James M., and William J. Abernathy. 1975. "A Dynamic Model of Process and Product Innovation." *Omega* 3 (6): 639-56. https://doi.org/10.1016/0305-0483(75)90068-7.

Uzzi, Brian. 1997. "Social Structure and Competition in Interfirm Networks: The Paradox of Embeddedness." *Administrative Science Quarterly* 42 (1): 35-67. https://doi.org/10.2307/2393808.

Uzzi, Brian, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. "Atypical Combinations and Scientific Impact." *Science* 342 (6157): 468-72. https://doi.org/10.1126/science.1240474.

Vaidis, David C., Willem W. A. Sleegers, Florian van Leeuwen, Kenneth G. DeMarree, Bjørn Sætrevik, Robert M. Ross, Kathleen Schmidt, John Protzko, et al. 2024. "A Multilab Replication of the Induced-Compliance Paradigm of Cognitive Dissonance." *Advances in Methods and Practices in Psychological Science* 7 (1). https://doi.org/10.1177/25152459231213375.

Vakratsas, Demetrios, and Tim Ambler. 1999. "How Advertising Works: What Do We Really Know?" *Journal of Marketing* 63 (1): 26-43. https://doi.org/10.1177/002224299906300103.

Vallacher, Robin R., and Daniel M. Wegner. 1987. "What Do People Think They're Doing? Action Identification and Human Behavior." *Psychological Review* 94 (1): 3-15. https://doi.org/10.1037/0033-295X.94.1.3.

Van Boven, Leaf, and Thomas Gilovich. 2003. "To Do or to Have? That Is the Question." *Journal of Personality and Social Psychology* 85 (6): 1193-1202. https://doi.org/10.1037/0022-3514.85.6.1193.

Van den Bulte, Christophe, and Gary L. Lilien. 2001. "Medical Innovation Revisited: Social Contagion Versus Marketing Effort." *American Journal of Sociology* 106 (5): 1409-35. https://doi.org/10.1086/320819.

Van den Bulte, Christophe, and Stefan Stremersch. 2004. "Social Contagion and Income Heterogeneity in New Product Diffusion: A Meta-Analytic Test." *Marketing Science* 23 (4): 530-44. https://doi.org/10.1287/mksc.1040.0054.

Van Dyne, Linn, and Jon L. Pierce. 2004. "Psychological Ownership and Feelings of Possession: Three Field Studies Predicting Employee Attitudes and Organizational Citizenship Behavior." *Journal of Organizational Behavior* 25 (4): 439-59. https://doi.org/10.1002/job.249.

Van Eerde, Wendelien, and Henk Thierry. 1996. "Vroom's Expectancy Models and Work-Related Criteria: A Meta-Analysis." *Journal of Applied Psychology* 81 (5): 575-86. https://doi.org/10.1037/0021-9010.81.5.575.

van Trijp, Hans C. M., Wayne D. Hoyer, and J. Jeffrey Inman. 1996. "Why Switch? Product Category-Level Explanations for True Variety-Seeking Behavior." *Journal of Marketing Research* 33 (3): 281-92. https://doi.org/10.1177/002224379603300303.

Vancouver, Jeffrey B., Charles M. Thompson, E. Casey Tischner, and Dan J. Putka. 2002. "Two Studies Examining the Negative Effect of Self-Efficacy on Performance." *Journal of Applied Psychology* 87 (3): 506-16. https://doi.org/10.1037/0021-9010.87.3.506.

Vandermerwe, Sandra, and Juan Rada. 1988. "Servitization of Business: Adding Value by Adding Services." *European Management Journal* 6 (4): 314-24. https://doi.org/10.1016/0263-2373(88)90033-3.

VanderWerf, Pieter A., and John F. Mahon. 1997. "Meta-Analysis of the Impact of Research Methods on Findings of First-Mover Advantage." *Management Science* 43 (11): 1510-19. https://doi.org/10.1287/mnsc.43.11.1510.

Vargo, Stephen L., and Robert F. Lusch. 2004. "Evolving to a New Dominant Logic for Marketing." *Journal of Marketing* 68 (1): 1-17. https://doi.org/10.1509/jmkg.68.1.1.24036.

Vargo, Stephen L., and Robert F. Lusch. 2008. "Service-Dominant Logic: Continuing the Evolution." *Journal of the Academy of Marketing Science* 36 (1): 1-10. https://doi.org/10.1007/s11747-007-0069-6.

Vargo, Stephen L., and Robert F. Lusch. 2016. "Institutions and Axioms: An Extension and Update of Service-Dominant Logic." *Journal of the Academy of Marketing Science* 44 (1): 5-23. https://doi.org/10.1007/s11747-015-0456-3.

Varian, Hal R. 1980. "A Model of Sales." *American Economic Review* 70 (4): 651-59.

Vaughan, Diane. 1996. *The Challenger Launch Decision: Risky Technology, Culture, and Deviance at NASA*. Chicago: University of Chicago Press.

Vaughn, Richard. 1980. "How Advertising Works: A Planning Model." *Journal of Advertising Research* 20 (5): 27-33. https://doi.org/10.1080/00218499.1980.12518790.

Veblen, Thorstein. 1899. *The Theory of the Leisure Class: An Economic Study of Institutions*. New York: Macmillan.

Venkatesh, Viswanath, and Hillol Bala. 2008. "Technology Acceptance Model 3 and a Research Agenda on Interventions." *Decision Sciences* 39 (2): 273-315. https://doi.org/10.1111/j.1540-5915.2008.00192.x.

Venkatesh, Viswanath, and Fred D. Davis. 2000. "A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies." *Management Science* 46 (2): 186-204. https://doi.org/10.1287/mnsc.46.2.186.11926.

Venkatesh, Viswanath, Michael G. Morris, Gordon B. Davis, and Fred D. Davis. 2003. "User Acceptance of Information Technology: Toward a Unified View." *MIS Quarterly* 27 (3): 425-78. https://doi.org/10.2307/30036540.

Venkatesh, Viswanath, James Y. L. Thong, and Xin Xu. 2012. "Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology." *MIS Quarterly* 36 (1): 157-78. https://doi.org/10.2307/41410412.

Vernon, Raymond. 1966. "International Investment and International Trade in the Product Cycle." *Quarterly Journal of Economics* 80 (2): 190-207. https://doi.org/10.2307/1880689.

Veryzer, Jr., Robert W., and J. Wesley Hutchinson. 1998. "The Influence of Unity and Prototypicality on Aesthetic Responses to New Product Designs." *Journal of Consumer Research* 24 (4): 374-85. https://doi.org/10.1086/209516.

Vickrey, William. 1961. "Counterspeculation, Auctions, and Competitive Sealed Tenders." *The Journal of Finance* 16 (1): 8-37. https://doi.org/10.1111/j.1540-6261.1961.tb02789.x.

Völckner, Franziska, and Henrik Sattler. 2006. "Drivers of Brand Extension Success." *Journal of Marketing* 70 (2): 18-34. https://doi.org/10.1509/jmkg.70.2.018.

Volpp, Kevin G., Leslie K. John, Andrea B. Troxel, Laurie Norton, Jennifer Fassbender, and George Loewenstein. 2008. "Financial Incentive-Based Approaches for Weight Loss: A Randomized Trial." *JAMA* 300 (22): 2631-37. https://doi.org/10.1001/jama.2008.804.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146-51. https://doi.org/10.1126/science.aap9559.

Voss, Kevin E., Eric R. Spangenberg, and Bianca Grohmann. 2003. "Measuring the Hedonic and Utilitarian Dimensions of Consumer Attitude." *Journal of Marketing Research* 40 (3): 310-20. https://doi.org/10.1509/jmkr.40.3.310.19238.

Vroom, Victor H. 1964. *Work and Motivation*. New York: Wiley.

Wahba, Mahmoud A., and Lawrence G. Bridwell. 1976. "Maslow Reconsidered: A Review of Research on the Need Hierarchy Theory." *Organizational Behavior and Human Performance* 15 (2): 212-40.

Wakker, Peter P. 2010. *Prospect Theory: For Risk and Ambiguity*. Cambridge: Cambridge University Press.

Walasek, Lukasz, and Neil Stewart. 2015. "How to Make Loss Aversion Disappear and Reverse: Tests of the Decision by Sampling Origin of Loss Aversion." *Journal of Experimental Psychology: General* 144 (1): 7-11. https://doi.org/10.1037/xge0000039.

Walker, Gordon, and David Weber. 1984. "A Transaction Cost Approach to Make-or-Buy Decisions." *Administrative Science Quarterly* 29 (3): 373-91. https://doi.org/10.2307/2393030.

Wänke, Michaela, Gerd Bohner, and Andreas Jurkowitsch. 1997. "There Are Many Reasons to Drive a BMW: Does Imagined Ease of Argument Generation Influence Attitudes?" *Journal of Consumer Research* 24 (2): 170-77.

Ward, Thomas B. 1994. "Structured Imagination: The Role of Category Structure in Exemplar Generation." *Cognitive Psychology* 27 (1): 1-40.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Wasserstein, Ronald L., and Nicole A. Lazar. 2016. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician* 70 (2): 129-33. https://doi.org/10.1080/00031305.2016.1154108.

Watson, David, Lee Anna Clark, and Auke Tellegen. 1988. "Development and Validation of Brief

Measures of Positive and Negative Affect: The PANAS Scales." *Journal of Personality and Social Psychology* 54 (6): 1063-70. https://doi.org/10.1037/0022-3514.54.6.1063.

Watson, David, and Auke Tellegen. 1985. "Toward a Consensual Structure of Mood." *Psychological Bulletin* 98 (2): 219-35. https://doi.org/10.1037/0033-2909.98.2.219.

Watts, Duncan J., and Peter Sheridan Dodds. 2007. "Influentials, Networks, and Public Opinion Formation." *Journal of Consumer Research* 34 (4): 441-58. https://doi.org/10.1086/518527.

Watts, Tyler W., Greg J. Duncan, and Haonan Quan. 2018. "Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Delay of Gratification and Later Outcomes." *Psychological Science* 29 (7): 1159-77. https://doi.org/10.1177/0956797618761661.

Weaver, Ray, and Shane Frederick. 2012. "A Reference Price Theory of the Endowment Effect." *Journal of Marketing Research* 49 (5): 696-707. https://doi.org/10.1509/jmr.09.0103.

Webb, Thomas L., and Paschal Sheeran. 2006. "Does Changing Behavioral Intentions Engender Behavior Change? A Meta-Analysis of the Experimental Evidence." *Psychological Bulletin* 132 (2): 249-68. https://doi.org/10.1037/0033-2909.132.2.249.

Weber, Martin, and Colin F. Camerer. 1998. "The Disposition Effect in Securities Trading: An Experimental Analysis." *Journal of Economic Behavior & Organization* 33 (2): 167-84. https://doi.org/10.1016/S0167-2681(97)00089-9.

Weber, Roberto A., and Colin F. Camerer. 2003. "Cultural Conflict and Merger Failure: An Experimental Approach." *Management Science* 49 (4): 400-415. https://doi.org/10.1287/mnsc.49.4.400.14430.

Wei, Yanhao Max. 2020. "The Similarity Network of Motion Pictures." *Management Science* 66 (4): 1647-71. https://doi.org/10.1287/mnsc.2018.3261.

Weick, Karl E. 1979. *The Social Psychology of Organizing*. 2nd ed. Reading, MA: Addison-Wesley.

Weick, Karl E. 1989. "Theory Construction as Disciplined Imagination." *Academy of Management Review* 14 (4): 516-31. https://doi.org/10.5465/amr.1989.4308376.

Weick, Karl E. 1993. "The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster." *Administrative Science Quarterly* 38 (4): 628-52. https://doi.org/10.2307/2393339.

Weick, Karl E. 1995a. *Sensemaking in Organizations*. Thousand Oaks, CA: Sage.

Weick, Karl E. 1995b. "What Theory Is Not, Theorizing Is." *Administrative Science Quarterly* 40 (3): 385-90. https://doi.org/10.2307/2393789.

Weick, Karl E., and Karlene H. Roberts. 1993. "Collective Mind in Organizations: Heedful Interrelating on Flight Decks." *Administrative Science Quarterly* 38 (3): 357-81. https://doi.org/10.2307/2393372.

Weick, Karl E., Kathleen M. Sutcliffe, and David Obstfeld. 2005. "Organizing and the Process of Sensemaking." *Organization Science* 16 (4): 409-21. https://doi.org/10.1287/orsc.1050.0133.

Weiner, Bernard. 1985. "An Attributional Theory of Achievement Motivation and Emotion." *Psychological Review* 92 (4): 548-73. https://doi.org/10.1037/0033-295X.92.4.548.

Weiner, Bernard. 2000. "Attributional Thoughts about Consumer Behavior." *Journal of Consumer Research* 27 (3): 382-87. https://doi.org/10.1086/317592.

Weiss, Howard M., and Russell Cropanzano. 1996. "Affective Events Theory: A Theoretical Discussion of the Structure, Causes and Consequences of Affective Experiences at Work." In *Research in Organizational Behavior*, edited by Barry M. Staw and Larry L. Cummings, 18:1-74. Greenwich, CT: JAI Press.

Weiss, Howard M., Jeffrey P. Nicholas, and Catherine S. Daus. 1999. "An Examination of the Joint Effects of Affective Experiences and Job Beliefs on Job Satisfaction and Variations in Affective Experiences over Time." *Organizational Behavior and Human Decision Processes* 78 (1): 1-24. https://doi.org/10.1006/obhd.1999.2824.

Weitzman, Martin L. 1974. "Prices Vs. Quantities." *The Review of Economic Studies* 41 (4): 477-91.

https://doi.org/10.2307/2296698.
Weitzman, Martin L. 1998. "Recombinant Growth." *The Quarterly Journal of Economics* 113 (2): 331-60.
Wernerfelt, Birger. 1984. "A Resource-Based View of the Firm." *Strategic Management Journal* 5 (2): 171-80. https://doi.org/10.1002/smj.4250050207.
Westphal, James D., Ranjay Gulati, and Stephen M. Shortell. 1997. "Customization or Conformity? An Institutional and Network Perspective on the Content and Consequences of TQM Adoption." *Administrative Science Quarterly* 42 (2): 366-94. https://doi.org/10.2307/2393924.
Westphal, James D., and Edward J. Zajac. 1994. "Substance and Symbolism in CEOs' Long-Term Incentive Plans." *Administrative Science Quarterly* 39 (3): 367-90. https://doi.org/10.2307/2393295.
Whetten, David A. 1989. "What Constitutes a Theoretical Contribution?" *Academy of Management Review* 14 (4): 490-95. https://doi.org/10.5465/amr.1989.4308371.
Whetten, David A. 2006. "Albert and Whetten Revisited: Strengthening the Concept of Organizational Identity." *Journal of Management Inquiry* 15 (3): 219-34. https://doi.org/10.1177/1056492606291200.
Whillans, Ashley V., Elizabeth W. Dunn, Paul Smeets, Rene Bekkers, and Michael I. Norton. 2017. "Buying Time Promotes Happiness." *Proceedings of the National Academy of Sciences* 114 (32): 8523-27.
White, Tiffany Barnett, Debra L. Zahay, Helge Thorbjørnsen, and Sharon Shavitt. 2008. "Getting Too Personal: Reactance to Highly Personalized Email Solicitations." *Marketing Letters* 19 (1): 39-50. https://doi.org/10.1007/s11002-007-9027-9.
Whittlesea, Bruce W. A. 1993. "Illusions of Familiarity." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19 (6): 1235-53. https://doi.org/10.1037/0278-7393.19.6.1235.
Whittlesea, Bruce W. A., and Lisa D. Williams. 1998. "Why Do Strangers Feel Familiar, but Friends Don't? A Discrepancy-Attribution Account of Feelings of Familiarity." *Acta Psychologica* 98 (2-3): 141-65.
Whittlesea, Bruce W. A., and Lisa D. Williams. 2001a. "The Discrepancy-Attribution Hypothesis: I. The Heuristic Basis of Feelings of Familiarity." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27 (1): 3-13.
Whittlesea, Bruce W. A., and Lisa D. Williams. 2001b. "The Discrepancy-Attribution Hypothesis: II. Expectation, Uncertainty, Surprise, and Feelings of Familiarity." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27 (1): 14-33.
Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, Dewey Dunnington, and Teun van den Brand. 2025. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. https://ggplot2.tidyverse.org.
Willander, Johan, and Maria Larsson. 2007. "Olfaction and Emotion: The Case of Autobiographical Memory." *Memory & Cognition* 35 (7): 1659-63. https://doi.org/10.3758/bf03193499.
Willer, David, ed. 1999. *Network Exchange Theory*. Westport, CT: Praeger.
Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. New York: Free Press.
Williamson, Oliver E. 1979. "Transaction-Cost Economics: The Governance of Contractual Relations." *The Journal of Law and Economics* 22 (2): 233-61. https://doi.org/10.1086/466942.
Williamson, Oliver E. 1981. "The Economics of Organization: The Transaction Cost Approach." *American Journal of Sociology* 87 (3): 548-77. https://doi.org/10.1086/227496.
Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.

Williamson, Oliver E. 1991. "Comparative Economic Organization: The Analysis of Discrete Structural Alternatives." *Administrative Science Quarterly* 36 (2): 269-96. https://doi.org/10.2307/2393356.

Williamson, Oliver E. 2000. "The New Institutional Economics: Taking Stock, Looking Ahead." *Journal of Economic Literature* 38 (3): 595-613. https://doi.org/10.1257/jel.38.3.595.

Wills, Thomas A. 1981. "Downward Comparison Principles in Social Psychology." *Psychological Bulletin* 90 (2): 245-71. https://doi.org/10.1037/0033-2909.90.2.245.

Wilson, James R., Robert E. Kuehn, and Frank A. Beach. 1963. "Modification in the Sexual Behavior of Male Rats Produced by Changing the Stimulus Female." *Journal of Comparative and Physiological Psychology* 56 (3): 636-44.

Wilson, Timothy D., and Daniel T. Gilbert. 2005. "Affective Forecasting: Knowing What to Want." *Current Directions in Psychological Science* 14 (3): 131-34.

Winer, Russell S. 1986. "A Reference Price Model of Brand Choice for Frequently Purchased Products." *Journal of Consumer Research* 13 (2): 250-56. https://doi.org/10.1086/209064.

Winkielman, Piotr, Jamin Halberstadt, Tedra Fazendeiro, and Steve Catty. 2006. "Prototypes Are Attractive Because They Are Easy on the Mind." *Psychological Science* 17 (9): 799-806. https://doi.org/10.1111/j.1467-9280.2006.01785.x.

Winter, Sidney G. 2003. "Understanding Dynamic Capabilities." *Strategic Management Journal* 24 (10): 991-95. https://doi.org/10.1002/smj.318.

Wiseman, Robert M., and Luis R. Gomez-Mejia. 1998. "A Behavioral Agency Model of Managerial Risk Taking." *Academy of Management Review* 23 (1): 133-53. https://doi.org/10.2307/259103.

Wofford, J. C., and Laurie Z. Liska. 1993. "Path-Goal Theories of Leadership: A Meta-Analysis." *Journal of Management* 19 (4): 857-76. https://doi.org/10.1177/014920639301900407.

Wolfers, Justin, and Eric Zitzewitz. 2004. "Prediction Markets." *Journal of Economic Perspectives* 18 (2): 107-26. https://doi.org/10.1257/0895330041371321.

Womack, James P., Daniel T. Jones, and Daniel Roos. 1990. *The Machine That Changed the World*. New York: Rawson Associates.

Woodward, Joan. 1965. *Industrial Organization: Theory and Practice*. London: Oxford University Press.

Worchel, Stephen, Jerry Lee, and Akanbi Adewole. 1975. "Effects of Supply and Demand on Ratings of Object Value." *Journal of Personality and Social Psychology* 32 (5): 906-14. https://doi.org/10.1037/0022-3514.32.5.906.

Wright, T. P. 1936. "Factors Affecting the Cost of Airplanes." *Journal of the Aeronautical Sciences* 3 (4): 122-28. https://doi.org/10.2514/8.155.

Wrzesniewski, Amy, and Jane E. Dutton. 2001. "Crafting a Job: Revisioning Employees as Active Crafters of Their Work." *Academy of Management Review* 26 (2): 179-201. https://doi.org/10.2307/259118.

Xie, Yihui. 2025a. *Bookdown: Authoring Books and Technical Documents with r Markdown*. https://github.com/rstudio/bookdown.

Xie, Yihui. 2025b. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Xie, Yihui, Joe Cheng, Xianying Tan, and Garrick Aden-Buie. 2025. *DT: A Wrapper of the JavaScript Library DataTables*. https://github.com/rstudio/DT.

Yang, Sybil, and Michael Lynn. 2014. "More Evidence Challenging the Robustness and Usefulness of the Attraction Effect." *Journal of Marketing Research* 51 (4): 508-13.

Yechiam, Eldad, and Guy Hochman. 2013. "Losses as Modulators of Attention: Review and Analysis of the Unique Effects of Losses over Gains." *Psychological Bulletin* 139 (2): 497-518. https:

//doi.org/10.1037/a0029383.
Yoo, Boonghee, and Naveen Donthu. 2001. "Developing and Validating a Multidimensional Consumer-Based Brand Equity Scale." *Journal of Business Research* 52 (1): 1-14. https://doi.org/10.1016/S0148-2963(99)00098-3.
Youn, Hyejin, Deborah Strumsky, Luis M. A. Bettencourt, and José Lobo. 2015. "Invention as a Combinatorial Process: Evidence from US Patents." *Journal of The Royal Society Interface* 12 (106): 20150272. https://doi.org/10.1098/rsif.2015.0272.
Zahavi, Amotz. 1975. "Mate Selection: A Selection for a Handicap." *Journal of Theoretical Biology* 53 (1): 205-14. https://doi.org/10.1016/0022-5193(75)90111-3.
Zaheer, Srilata. 1995. "Overcoming the Liability of Foreignness." *Academy of Management Journal* 38 (2): 341-63. https://doi.org/10.2307/256683.
Zaheer, Srilata, and Elaine Mosakowski. 1997. "The Dynamics of the Liability of Foreignness: A Global Study of Survival in Financial Services." *Strategic Management Journal* 18 (6): 439-63.
Zahra, Shaker A., and Gerard George. 2002. "Absorptive Capacity: A Review, Reconceptualization, and Extension." *Academy of Management Review* 27 (2): 185-203.
Zajonc, Robert B. 1965. "Social Facilitation." *Science* 149 (3681): 269-74. https://doi.org/10.1126/science.149.3681.269.
Zajonc, Robert B. 1968. "Attitudinal Effects of Mere Exposure." *Journal of Personality and Social Psychology* 9 (2, Pt. 2): 1-27. https://doi.org/10.1037/h0025848.
Zajonc, Robert B., and Stephen M. Sales. 1966. "Social Facilitation of Dominant and Subordinate Responses." *Journal of Experimental Social Psychology* 2 (2): 160-68. https://doi.org/10.1016/0022-1031(66)90077-1.
Zanna, Mark P., and Joel Cooper. 1974. "Dissonance and the Pill: An Attribution Approach to Studying the Arousal Properties of Dissonance." *Journal of Personality and Social Psychology* 29 (5): 703-9. https://doi.org/10.1037/h0036651.
Zeithaml, Valarie A. 1988. "Consumer Perceptions of Price, Quality, and Value: A Means-End Model and Synthesis of Evidence." *Journal of Marketing* 52 (3): 2-22. https://doi.org/10.1177/002224298805200302.
Zeithaml, Valarie A., Leonard L. Berry, and A. Parasuraman. 1996. "The Behavioral Consequences of Service Quality." *Journal of Marketing* 60 (2): 31-46. https://doi.org/10.1177/002224299606000203.
Zhao, Eric Yanfei, Greg Fisher, Michael Lounsbury, and Danny Miller. 2017. "Optimal Distinctiveness: Broadening the Interface Between Institutional Theory and Strategic Management." *Strategic Management Journal* 38 (1): 93-113. https://doi.org/10.1002/smj.2589.
Zhao, Hao, Sandy J. Wayne, Brian C. Glibkowski, and Jesus Bravo. 2007. "The Impact of Psychological Contract Breach on Work-Related Outcomes: A Meta-Analysis." *Personnel Psychology* 60 (3): 647-80. https://doi.org/10.1111/j.1744-6570.2007.00087.x.
Zhao, Xinshu, Jr. Lynch John G., and Qimei Chen. 2010. "Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis." *Journal of Consumer Research* 37 (2): 197-206. https://doi.org/10.1086/651257.
Zhu, Feng, and Qihong Liu. 2018. "Competing with Complementors: An Empirical Look at Amazon.com." *Strategic Management Journal* 39 (10): 2618-42. https://doi.org/10.1002/smj.2932.
Zhu, Hao. 2024. *kableExtra: Construct Complex Table with Kable and Pipe Syntax*. http://haozhu233.github.io/kableExtra/.
Zigurs, Ilze, and Bonnie K. Buckland. 1998. "A Theory of Task/Technology Fit and Group Support Systems Effectiveness." *MIS Quarterly* 22 (3): 313-34. https://doi.org/10.2307/249668.
Zollo, Maurizio, and Sidney G. Winter. 2002. "Deliberate Learning and the Evolution of Dynamic Capabilities." *Organization Science* 13 (3): 339-51. https://doi.org/10.1287/orsc.13.3.339.27

80.