

LEARNING MADE EASY



4th Edition

Economics

for
dummies[®]
A Wiley Brand

WITH
**ONLINE
PRACTICE**



Find FREE quizzes
for every chapter online

Discover the principles of
political economy and taxation

Identify the foundations
of the affluent society

Sean Masaki Flynn, PhD

Assistant professor of economics
Scripps College



This book comes with access to more content online.

Quiz yourself, track your progress,
and improve your grade!

Register your book or ebook at
www.dummies.com/go/getaccess.

Select your product, and then follow the prompts
to validate your purchase.

You'll receive an email with your PIN and instructions.



Economics

4th Edition

by Sean Masaki Flynn, PhD

**for
dummies®**
A Wiley Brand

Economics For Dummies®, 4th Edition

Published by: **John Wiley & Sons, Inc.**, 111 River Street, Hoboken, NJ 07030-5774, www.wiley.com

Copyright © 2024 by John Wiley & Sons, Inc., Hoboken, New Jersey

Published simultaneously in Canada

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the Publisher. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, or online at <http://www.wiley.com/go/permissions>.

Trademarks: Wiley, For Dummies, the Dummies Man logo, Dummies.com, Making Everything Easier, and related trade dress are trademarks or registered trademarks of John Wiley & Sons, Inc., and may not be used without written permission. All other trademarks are the property of their respective owners. John Wiley & Sons, Inc., is not associated with any product or vendor mentioned in this book.

LIMIT OF LIABILITY/DISCLAIMER OF WARRANTY: WHILE THE PUBLISHER AND AUTHORS HAVE USED THEIR BEST EFFORTS IN PREPARING THIS WORK, THEY MAKE NO REPRESENTATIONS OR WARRANTIES WITH RESPECT TO THE ACCURACY OR COMPLETENESS OF THE CONTENTS OF THIS WORK AND SPECIFICALLY DISCLAIM ALL WARRANTIES, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. NO WARRANTY MAY BE CREATED OR EXTENDED BY SALES REPRESENTATIVES, WRITTEN SALES MATERIALS OR PROMOTIONAL STATEMENTS FOR THIS WORK. THE FACT THAT AN ORGANIZATION, WEBSITE, OR PRODUCT IS REFERRED TO IN THIS WORK AS A CITATION AND/OR POTENTIAL SOURCE OF FURTHER INFORMATION DOES NOT MEAN THAT THE PUBLISHER AND AUTHORS ENDORSE THE INFORMATION OR SERVICES THE ORGANIZATION, WEBSITE, OR PRODUCT MAY PROVIDE OR RECOMMENDATIONS IT MAY MAKE. THIS WORK IS SOLD WITH THE UNDERSTANDING THAT THE PUBLISHER IS NOT ENGAGED IN RENDERING PROFESSIONAL SERVICES. THE ADVICE AND STRATEGIES CONTAINED HEREIN MAY NOT BE SUITABLE FOR YOUR SITUATION. YOU SHOULD CONSULT WITH A SPECIALIST WHERE APPROPRIATE. FURTHER, READERS SHOULD BE AWARE THAT WEBSITES LISTED IN THIS WORK MAY HAVE CHANGED OR DISAPPEARED BETWEEN WHEN THIS WORK WAS WRITTEN AND WHEN IT IS READ. NEITHER THE PUBLISHER NOR AUTHORS SHALL BE LIABLE FOR ANY LOSS OF PROFIT OR ANY OTHER COMMERCIAL DAMAGES, INCLUDING BUT NOT LIMITED TO SPECIAL, INCIDENTAL, CONSEQUENTIAL, OR OTHER DAMAGES.

For general information on our other products and services, please contact our Customer Care Department within the U.S. at 877-762-2974, outside the U.S. at 317-572-3993, or fax 317-572-4002. For technical support, please visit <https://hub.wiley.com/community/support/dummies>.

Wiley publishes in a variety of print and electronic formats and by print-on-demand. Some material included with standard print versions of this book may not be included in e-books or in print-on-demand. If this book refers to media such as a CD or DVD that is not included in the version you purchased, you may download this material at <http://booksupport.wiley.com>. For more information about Wiley products, visit www.wiley.com.

Library of Congress Control Number: 2023943421

ISBN 978-1-394-16133-1 (pbk); ISBN 978-1-394-16135-5 (ebk); ISBN 978-1-394-16134-8 (epdf)

Contents at a Glance

Introduction	1
Part 1: Economics: The Science of How People Deal with Scarcity	5
CHAPTER 1: What Economics Is and Why You Should Care	7
CHAPTER 2: Cookies or Ice Cream? Exploring Consumer Choices	23
CHAPTER 3: Producing Stuff to Maximize Happiness	37
Part 2: Microeconomics: The Science of Consumer and Firm Behavior	59
CHAPTER 4: Supply and Demand Made Easy	61
CHAPTER 5: Introducing <i>Homo Economicus</i> , the Utility-Maximizing Consumer	87
CHAPTER 6: The Core of Capitalism: The Profit-Maximizing Firm	105
CHAPTER 7: Why Economists Love Free Markets and Competition	129
CHAPTER 8: Monopolies: Bad Behavior When Competition Is Lacking	155
CHAPTER 9: Oligopoly and Monopolistic Competition: Middle Grounds	177
Part 3: Applying the Theories of Microeconomics	195
CHAPTER 10: Property Rights and Wrongs	197
CHAPTER 11: Asymmetric Information and Public Goods	209
CHAPTER 12: Internet Economics: Networks and Platforms	225
CHAPTER 13: Behavioral Economics: Investigating Irrationality	245
Part 4: Macroeconomics: The Science of Economic Growth and Stability	263
CHAPTER 14: How Economists Measure the Macroeconomy	265
CHAPTER 15: Inflation Frustration: Why More Money Isn't Always Good	285
CHAPTER 16: Understanding Why Recessions Happen	303
CHAPTER 17: Fighting Recessions with Monetary and Fiscal Policy	329
CHAPTER 18: Grasping Origins and Effects of Financial Crises	359

Part 5: The Part of Tens	369
CHAPTER 19: Ten Seductive Economic Fallacies	371
CHAPTER 20: Ten Economic Ideas to Hold Dear	379
CHAPTER 21: Ten (Or So) Super-Famous Economists.....	383
CHAPTER 22: Ten (Or So) Recent Nobel Laureates in Economics.....	391
Glossary	397
Index	405

Table of Contents

INTRODUCTION	1
About This Book	1
Foolish Assumptions	3
Icons Used in This Book	3
Beyond the Book	4
Where to Go from Here	4
 PART 1: ECONOMICS: THE SCIENCE OF HOW PEOPLE DEAL WITH SCARCITY	 5
CHAPTER 1: What Economics Is and Why You Should Care	7
Considering a Little Economic History	8
Pondering just how nasty, brutish, and short life used to be	8
Identifying the institutions that raise living standards	9
Looking toward the future	10
Framing Economics as the Science of Scarcity	11
Sending Microeconomics and Macroeconomics to Separate Corners	11
Getting up close and personal: Microeconomics	12
Zooming out: Macroeconomics and the big picture	15
Understanding How Economists Use Models and Graphs	17
Abstracting from reality is a good thing	17
Introducing your first model: The demand curve	18
Drawing your own demand curve	20
 CHAPTER 2: Cookies or Ice Cream? Exploring Consumer Choices	 23
Describing Human Behavior with a Choice Model	24
Pursuing Personal Happiness	25
Using utility to measure happiness	25
Taking “selfless” actions into account	25
Self-interest can promote the common good	26
You Can’t Have Everything: Examining Limitations	27
Resource constraints	27
Technology constraints	28
Time constraints	28
Opportunity cost: The unavoidable cost	29
Making Your Choice: Deciding What and How Much You Want	30

Exploring Violations and Limitations of the Economist's Choice Model	32
Understanding uninformed decision-making	32
Making sense of irrationality	33
CHAPTER 3: Producing Stuff to Maximize Happiness	37
Figuring Out What's Possible to Produce	38
Classifying resources	39
Clarifying human capital	39
Diminishing returns	40
Allocating resources	41
Graphing your production possibilities	42
Reaching new frontiers with better technology	45
Deciding What to Produce	46
Comparing market results and government interventions	47
Opting for a mixed economy	53
Promoting Technology and Innovation	56
 PART 2: MICROECONOMICS: THE SCIENCE OF CONSUMER AND FIRM BEHAVIOR	 59
CHAPTER 4: Supply and Demand Made Easy	61
Deconstructing Demand	62
Prices and other stuff: Looking at what affects quantity demanded	63
Graphing the demand curve	64
Opportunity costs: Setting the slope of the demand curve	66
Elasticity: Looking at extreme demand cases	67
Sorting Out Supply	69
Graphing the supply curve	69
Using elasticity to understand extreme supply cases	73
Bringing Supply and Demand Together	75
Market equilibrium: Seeking a balance	75
Demonstrating the stability of the market equilibrium	77
Adjusting to new market equilibriums when supply or demand changes	79
Price Controls: Keeping Prices Away from Market Equilibrium	81
Setting upper limits with price ceilings	82
Propping up prices with price floors	83
 CHAPTER 5: Introducing <i>Homo Economicus</i>, the Utility-Maximizing Consumer	 87
Choosing by Ranking	88
Getting Less from More: Diminishing Marginal Utility	89

Choosing Among Many Options When Facing a Limited Budget	92
Trying to buy as much (marginal) utility as you can	92
Purchasing the best combination of two goods to maximize total utility	94
Aiming for equal marginal utility per dollar.	97
Deriving Demand Curves from Diminishing Marginal Utility	99
Seeing how price changes affect quantities demanded	99
Graphing the price and quantity changes to form a demand curve.	101
 CHAPTER 6: The Core of Capitalism: The Profit-Maximizing Firm	
A Firm's Goal: Maximizing Profits	106
Facing Competition	107
Listing the requirements for perfect competition	107
Taking prices but setting quantities	108
Distinguishing between accounting profits and economic profits	110
Analyzing a Firm's Cost Structure	111
Focusing on costs per unit of output	112
Examining average variable costs.	113
Watching average fixed costs fall	114
Tracking the movement of average total costs.	115
Focusing on marginal costs	116
Noticing where marginal cost equals average cost	117
Comparing Marginal Revenues with Marginal Costs	119
Finding where marginal revenue equals marginal cost.	119
Visualizing profits	122
Visualizing losses.	124
Pulling the Plug: When Producing Nothing Is Your Best Bet.	125
Distinguishing between the short run and the long run in microeconomics.	126
The short-run shutdown condition: Variable costs exceed total revenues.	126
The long-run shutdown condition: Total costs exceed total revenues	128
 CHAPTER 7: Why Economists Love Free Markets and Competition	
Ensuring That Benefits Exceed Costs: Competitive Free Markets . .	130
Examining the traits of a properly functioning market	131
Analyzing the efficiency of free markets	132
Measuring everyone's gains with total surplus.	135

When Free Markets Lose Their Freedom: Dealing with Deadweight Losses	141
Coming up short: The deadweight loss from a price ceiling	141
Death and taxes: Finding the deadweight loss of a tax	142
Hallmarks of Perfect Competition: Zero Profits and Lowest Possible Costs	146
Understanding the causes and consequences of perfect competition	146
Peering into the process of perfect competition	147
Graphing how profits guide firm entry and exit	149
CHAPTER 8: Monopolies: Bad Behavior When Competition Is Lacking	155
Examining Profit-Maximizing Monopolies	156
Zeroing in on the problems monopolies cause	156
Identifying the source of the problem: Decreasing marginal revenues	157
Choosing an output level to maximize profits	162
Comparing Monopolies with Competitive Firms	166
Looking at output and price levels	166
Deadweight losses: Quantifying the harm caused by monopolies	168
Losing efficiency	169
Considering Good Monopolies	169
Encouraging innovation and investment with patents	170
Reducing annoyingly redundant competitors	170
Keeping costs low with natural monopolies	171
Regulating Monopolies	172
Subsidizing a monopoly to increase output	172
Imposing minimum output requirements	172
Regulating monopoly pricing	173
Breaking up a monopoly into several competing firms	176
CHAPTER 9: Oligopoly and Monopolistic Competition: Middle Grounds	177
Oligopolies: Looking at the Allure of Joining Forces	178
Sharing power over prices	179
Cartel behavior: Trying to imitate monopolists	179
Considering the criteria for coordinating a cartel	180
Understanding Incentives to Cheat on a Cartel Agreement	181
Fleshing out the prisoner's dilemma	181
Enforcing the agreement: Resolving the dilemma with credible threats	184
Seeing that OPEC is trapped in a prisoner's dilemma	185
Using an enforcer to help OPEC members stick to quotas	187

Regulating Oligopolies	188
Breaking up dominant firms	188
Attempting to apply antitrust laws	189
Studying a Hybrid: Monopolistic Competition	189
Benefiting from product differentiation.	189
Facing profit limits	190
PART 3: APPLYING THE THEORIES OF MICROECONOMICS	195
CHAPTER 10: Property Rights and Wrongs	197
Allowing Markets to Reach Socially Optimal Outcomes	198
Examining Externalities: The Costs and Benefits Others Feel from Your Actions	199
Noting the effects of negative externalities.	200
Accepting positive amounts of negative externalities	202
Dealing with negative externalities.	203
Calculating the consequences of positive externalities.	204
Subsidizing things that provide positive externalities	206
Tragedy of the Commons: Overexploiting Commonly Owned Resources	206
Overgrazing on a commonly owned field	206
Extinctions and poor property rights	207
Avoiding the tragedy.	208
CHAPTER 11: Asymmetric Information and Public Goods	209
Facing Up to Asymmetric Information	210
Realizing that asymmetric information limits trade.	210
Souring on the lemons problem: The used car market.	211
Issuing insurance when you can't tell individuals apart	215
Providing Public Goods.	219
Taxing to provide public goods	220
Enlisting philanthropy to provide public goods	220
Providing a public good by selling a related private good.	222
Ranking new technology as a public good.	222
CHAPTER 12: Internet Economics: Networks and Platforms	225
Examining Economies of Scale among Internet Businesses.	226
Pricing, profits, and elasticity.	228
Pricing and competition at zero MC	231
Noticing Network Effects	232
Understanding how network effects enhance market power	233
Visualizing the exponential scaling of network effects	234

Deconstructing Digital Platforms	235
Balancing user engagement against the need to charge fees	236
Parsing platform profit maximization	237
Regulating the Digital Economy.	240
Engaging with the Microsoft and Google cases	241
Distinguishing between winner-take-all and contestable industries	242
Dealing with concentration	243
 CHAPTER 13: Behavioral Economics: Investigating Irrationality	245
Explaining the Need for Behavioral Economics	245
Complementing Neo-Classical Economics with Behavioral Economics.	246
Examining Our Amazing, Efficient, and Error-Prone Brains	248
Deciphering heuristics	248
Deconstructing brain modularity	249
Cogitating on cognitive biases.	250
Surveying Prospect Theory	251
Shrinking packages and loss aversion	252
Framing effects and advertising	253
Anchoring and credit card bills	254
Examining the endowment effect.	255
Stipulating status quo bias.	255
Countering Myopia and Time Inconsistency.	256
Focusing on myopia	256
Tattling on time inconsistency.	257
Beating self-control problems with precommitments.	257
Gauging Fairness and Self-Interest.	258
Defining fairness	259
Examining the experimental evidence for fairness	259
Digesting the experimental evidence on fairness	261
 PART 4: MACROECONOMICS: THE SCIENCE OF ECONOMIC GROWTH AND STABILITY	263
 CHAPTER 14: How Economists Measure the Macroeconomy	265
Getting a Grip on the GDP and Its Parts.	266
Leaving some things out of GDP	267
Tallying up what counts in GDP.	267
Accounting for streams of incomes, and assets	268
Following the funds around and around.	270
Counting stuff when it's made, not when it's sold	271
Watching GDP rise with the good, the bad, and the ugly	272

Diving In to the GDP Equation	273
"C" is for consumption (that's good enough for me!)	274
"I" is for investment in capital stock	275
The big "G" (government, that is).	277
Measuring foreign trade with "NX"	278
Making Sense of International Trade and Its Effect on the Economy.	279
"Trade deficit" ain't fightin' words	279
Considering assets — not just cash	280
Wielding a comparative advantage.	282
CHAPTER 15: Inflation Frustration: Why More Money Isn't Always Good	285
Buying an Inflation: When Too Much Money Is a Bad Thing.	286
Balancing money supply and demand.	287
Giving in to the inflation temptation.	289
Tallying up the effects of inflation.	293
Measuring Inflation	295
Creating your very own market basket	296
Calculating the inflation rate	297
Setting up a price index	297
Determining the real standard of living with the price index	299
Identifying price index problems	299
Pricing the Future: Nominal and Real Interest Rates	301
Using the Fisher equation	301
Realizing that predictions aren't perfect	302
CHAPTER 16: Understanding Why Recessions Happen	303
Introducing the Business Cycle	304
Striving for Full-Employment Output	305
Returning to Y^* : The Natural Result of Price Adjustments	306
Responding to Economic Shocks: Short-Run and Long-Run Effects	307
Defining some critical terms	307
Contemplating the Tao of P: Looking at price adjustments in the long run.	309
A shock to the system: Adjusting to a shift in aggregate demand	310
Dealing with fixed prices in the short run	311
Putting together the long and short of it.	313
Heading toward Recession: Getting Stuck with Sticky Prices	315
Cutting wages or cutting workers	315
Adding up the costs of wages and profits	316
Returning to Y^* with and without government intervention.	317

Achieving Equilibrium with Sticky Prices: The Keynesian Model . . .	318
Adjusting inventories instead of prices	319
Boosting GDP in the Keynesian model.	327
CHAPTER 17: Fighting Recessions with Monetary and Fiscal Policy	329
Stimulating Demand to End Recessions	330
Aiming for full-employment output	330
Back to work: Shifting the AD curve right	332
Generating Inflation: The Risk of Too Much Stimulation.	332
Trying to increase output beyond Y^*	333
Tracing the movement of real wages	334
Failing to stimulate: What happens when a stimulus is expected	336
Figuring Out Fiscal Policy	340
Increasing government spending to help end recessions.	340
Dealing with deficits	342
Dissecting Monetary Policy	344
Identifying the benefits of fiat money over the gold standard. . .	345
Realizing you can have too much money.	346
Getting the basics about bonds.	348
Seeing the link between bond prices and interest rates	349
Changing the money supply to change interest rates	350
Lowering interest rates to stimulate the economy	351
Adding In the Administered Rates	352
Understanding how rational expectations can limit monetary policy	353
Examining quantitative easing and the Great Recession	356
CHAPTER 18: Grasping Origins and Effects of Financial Crises	359
Understanding How Debt-Driven Bubbles Develop.	360
Embracing borrowing in a booming economy	361
Offering larger loans as collateral values rise	361
Relaxing lending standards	362
Borrowing more in hopes of profit.	362
Watching the process gain momentum.	363
Seeing the Bubble Burst.	363
Deleveraging: Trying to ditch debt as prices fall	363
Comprehending bank collapses caused by bursting bubbles . .	364
Leading into a recession.	365

After the Crisis: Looking at Recovery	366
Enduring a broken banking system	366
Struggling with structural mismatches.	367
Noting the limits of government policy	368
PART 5: THE PART OF TENS	369
CHAPTER 19: Ten Seductive Economic Fallacies	371
The Lump-of-Labor Fallacy.	371
The Overpopulation Fallacy	372
The Sequence Fallacy	373
The Protectionism Fallacy	373
The Fallacy of Composition	374
The All-In Fallacy	374
The Market-Instability Fallacy	375
The Cheap-Foreign-Wages Fallacy	375
The Tax-Incentives Fallacy	376
The Unintended-Consequences Fallacy.	376
CHAPTER 20: Ten Economic Ideas to Hold Dear	379
Self-Interest Can Improve Society.	379
Free Markets Require Regulation	380
Economic Growth Relies on Innovation	380
Freedom and Democracy Make Us Richer.	380
Education Raises Living Standards	380
Intellectual Property Boosts Innovation.	381
Weak Property Rights Cause All Environmental Problems	381
International Trade Is a Good Thing.	381
Government Can Provide Public Goods.	382
Preventing Inflation Is Easy	382
CHAPTER 21: Ten (Or So) Super-Famous Economists	383
Adam Smith	383
David Ricardo.	384
Karl Marx	384
Alfred Marshall.	385
John Maynard Keynes.	386
Kenneth Arrow and Gerard Debreu	386
Milton Friedman	386
Paul Samuelson	387
Robert Solow	387
Gary Becker	388
Robert Lucas	388

CHAPTER 22:	Ten (Or So) Recent Nobel Laureates in Economics	391
	Amartya Sen	391
	Daniel Kahneman	392
	Elinor Ostrom	392
	Dan McFadden	393
	Vernon Smith	393
	George Akerlof	393
	Esther Duflo	394
	Paul Krugman	394
	David Card	395
	Richard Thaler	395
	Robert Shiller	396
	Ben Bernanke	396
GLOSSARY		397
INDEX		405

Introduction

Economics is about humanity's struggle to achieve happiness in a world full of constraints. There's never enough time or money to do everything people want, and things like curing cancer are still impossible because the necessary technologies haven't been developed yet. But people are clever. They tinker and invent, ponder and innovate. They look at what they have and what they can do with it and take steps to make sure that if they can't have everything, they'll at least have as much as possible.

Having to choose is a fundamental part of everyday life. The science that studies *how* people choose — economics — is indispensable if you really want to understand human beings both as individuals and as members of larger organizations. Sadly, though, economics has typically been explained quite badly. As a result, people tend to either dismiss it as impenetrable gobbledygook or stand falsely in awe of it. After all, if it's hard to understand, it must be important, right?

I wrote this book so you can quickly and easily understand economics for what it is — a serious science that studies a serious subject and has developed some seriously good ways of explaining human behavior in the (very serious) real world. You'll understand much more about people, the government, international relations, business, and environmental issues after reading these pages. Economics touches on nearly everything, so the returns to reading this book are enormous.

About This Book

The Scottish historian Thomas Carlyle called economics the “dismal science.” I'm going to do my best to make sure that you don't come to agree with him.

I've organized this book to get as much economics into you as quickly and effortlessly as possible. I've also done my best to keep it lively and fun.

In this book, you find the most important economic theories, hypotheses, and discoveries without a zillion obscure details, outdated examples, or complicated mathematical “proofs.” Among the topics covered are

- » How the government fights recessions and unemployment
- » How and why international trade benefits both individuals and nations
- » Why faulty or nonexistent property rights are responsible for most environmental problems, including global warming, air pollution, and species extinctions
- » How profit guides businesses to produce both existing goods and services as well as new products
- » How prices serve as signals in market economies
- » Why competitive firms are almost always better for society than monopolies
- » How the Federal Reserve controls the money supply, interest rates, and inflation at the same time
- » Why government policies such as price controls and subsidies often cause much more harm than good
- » How the simple supply-and-demand model can explain the prices of everything from comic books to open-heart surgeries

You can read the chapters in any order and immediately jump to what you need to know without having to read a bunch of stuff that you couldn’t care less about.

Economists like competition, so you shouldn’t be surprised that there are a lot of competing views when it comes to economic institutions and economic policy. Indeed, it’s only through vigorous debate and careful review of the evidence that the profession improves its understanding of how the world works. This book contains core ideas and concepts that economists agree are valid and important — I try to steer clear of fads or ideas that foster a lot of disagreement. (If you want to be subjected to my opinions and pet theories, you’ll have to buy me a drink.)

Note: Economics is full of two things you may not find very appealing: jargon and algebra. To minimize confusion, whenever I introduce a new term, I put it in *italics* and follow it closely with an easy-to-understand definition. Also, whenever I bring algebra into the discussion, I use those handy *italics* again to let you know that I’m referring to a mathematical variable. For instance, *I* is the abbreviation for investment, so you may see a sentence like this: I think *I* is too big.

I try to keep equations to a minimum, but sometimes they help to clarify things. In such instances, I sometimes use several equations, one after another. To avoid

confusion about which equation I'm referring to at any given time, I give each equation a number, which I put in parentheses. For example,

$$E = mc^2 \quad (1)$$

$$MTV = ESPN + CNN^2 \quad (2)$$

Foolish Assumptions

I wrote this book assuming some things about you:

- » You're sharp, thoughtful, and interested in how the world works.
- » You're either a student trying to flesh out what you're learning in class or you're a citizen of the world who realizes that a good grounding in economics will help you understand everything from business and politics to social issues like poverty and discrimination.
- » You want to know some economics, but you're also busy leading a full life. Consequently, although you want the crucial facts, you don't want to have to read through a bunch of minutiae to find them.
- » You're not totally intimidated by numbers, facts, and figures. Indeed, you welcome them because you like to have things proven to you instead of taking them on faith.
- » You like learning *why* as well as *what*. You want to know why things happen and how they work instead of just memorizing factoids.

Icons Used in This Book

To make this book easier to read and simpler to use, I include a few icons to help you find and fathom critical ideas and information.



REMEMBER

This icon alerts you that I'm explaining a fundamental economic concept or fact that you would do well to stash away in your memory for later use.



TECHNICAL
STUFF

This icon tells you that the ideas and information that it accompanies are a bit more technical or mathematical than other sections of the book. This information can be interesting and informative, but I've designed the book so that you don't need to understand it to get the big picture about what's happening. Feel free to skip this stuff if you want.



TIP

This icon points out time savers. I place this icon next to suggestions for ways to do or think about things that can save you some effort.



WARNING

This icon indicates troublesome areas. I've used it to identify potential pitfalls.

Beyond the Book

To view this book's Cheat Sheet, go to www.dummies.com and search for "Economics For Dummies Cheat Sheet" for a handy reference guide that answers common questions about economics.

To gain access to the online practice, all you have to do is register. Just follow these simple steps:

1. **Register your book or ebook at Dummies.com to get your PIN.** Go to www.dummies.com/go/getaccess.
2. **Select your product from the drop-down list on that page.**
3. **Follow the prompts to validate your product.** Then check your email for a confirmation message with your PIN and instructions for logging in.

If you don't receive this email within two hours, please check your spam folder before contacting us through our Technical Support website at <http://support.wiley.com> or by phone at 877-762-2974.

Now you're ready to go! You can return to the practice material as often as you want. Just log on with the username and password you created during your initial login. No need to enter the access code a second time.

Your registration is good for one year from the day you activate your PIN.

Where to Go from Here

This book is set up so that you can understand what's happening even if you skip around. The book is also divided into independent parts so that you can, for instance, read all about microeconomics without having to read anything about macroeconomics. The table of contents and index can help you find specific topics easily. But if you don't know where to begin, just do the old-fashioned thing and start at the beginning.

1

Economics: The Science of How People Deal with Scarcity

IN THIS PART . . .

Find out what economics is, what economists do, and why these things are important.

Decipher how people decide what brings them the most happiness.

Understand how goods and services are produced, how resources are allocated, and the roles of government and the market.

IN THIS CHAPTER

- » Taking a quick peek at economic history
- » Observing how people cope with scarcity
- » Separating macroeconomics from microeconomics
- » Getting a grip on economic graphs and models

Chapter **1**

What Economics Is and Why You Should Care

Economics studies how people and societies make decisions that allow them to get the most from their limited resources. And because every country, every business, and every person has to deal with constraints, economics is everywhere. At this very moment, for example, you could be doing something other than reading this book. You could be grueling through your hundredth pushup, bingeing the hottest series on Netflix, or texting with that friend you reconnected with at your last reunion. You should only be reading this book if doing so is the best possible use of your limited time. In the same way, you should hope that the resources used to create this book have been put to their best use and that every dollar your government spends is being managed in the best possible way.

Economics gets to the heart of these issues by analyzing the behavior of individuals, firms, governments, and other social and political institutions to see how well they convert humanity's limited resources into the goods and services that would best satisfy needs and wants.

Considering a Little Economic History

To better understand today's economic situation and what sorts of changes might promote the most significant improvements in economic efficiency, you have to look back on economic history to see how humanity got to where it is now. Please stick with me. I'll make this discussion as efficient and painless as possible.

Pondering just how nasty, brutish, and short life used to be

For most of human history, people didn't manage to squeeze much out of their limited resources. Living standards were quite low, and people lived poor, short, and painful lives. (Forget about being able to microwave the Lean Cuisine you had delivered to your doorstep or have a video chat with your doctor about the peculiarity of your Tuesday-only food allergy.) Consider the following facts, which didn't change until just a few centuries ago:

- » Life expectancy at birth was about 25 years.
- » More than 30 percent of newborns never made it to their fifth birthdays.
- » Most people experienced horrible diseases and starvation.
- » The standard of living was low and stayed low, generation after generation. Except for the nobles, everybody lived at or near subsistence, century after century.

In the past 250 years or so, everything changed. For the first time in history, people figured out how to use electricity, computers, television and radio, antibiotics, aviation, and a host of other technologies, including artificial intelligence and genetic engineering. Each has allowed people to accomplish much more with the limited amounts of air, water, soil, and sea they were given on Earth. The result has been an explosion in living standards, with life expectancy at birth now greater than 70 years worldwide and many people able to afford much better housing, clothing, and food than imaginable a few hundred years ago.

Of course, not everything is perfect. Grinding poverty is still a fact of life in a significant fraction of the world. And even the wealthiest nations have to cope with pressing economic problems like unemployment and how to transition workers from dying industries to growing ones. But the fact remains that the modern world is much richer than it was in centuries past. Even better, most nations now enjoy sustained economic growth so that their living standards rise almost every year.

Identifying the institutions that raise living standards

The obvious reason for higher living standards is that humans have recently discovered lots of valuable technologies. But if you dig a little deeper, you have to wonder why a technologically innovative society didn't happen earlier.

The ancient Greeks, for example, invented a simple steam engine as well as a sophisticated mechanical computer that could calculate dates centuries into the future. They even developed a mechanical way of storing what today would be called a *computer algorithm*. But they never got around to having an industrial revolution or figuring out how they could raise the average person's standard of living.

Ancient Greece wasn't alone. Even though there have always been brilliant people in every society in history, it wasn't until the late 18th century, in England, that the Industrial Revolution started. Only then did living standards begin to rise substantially, year after year.



REMEMBER

What factors combined in the late 18th century to so radically accelerate economic growth? The short answer is that the following institutions were in place:

- » **Democracy:** Because the commoners outnumbered the nobles, the advent of democracy meant that, for the first time, governments reflected the interests of society at large. A significant result was the creation of government policy that favored city-dwelling merchants and manufacturers over landed nobles living out in the countryside.
- » **The limited liability corporation:** Under this business structure, investors could lose only the amount of their investments rather than being personally liable for all of a business's debts and liabilities. That reduction in risk led to way more investing — enough, in fact, to pay for factories and railroads and steamships.
- » **Patent rights to protect inventors:** Before patents, inventors usually saw their ideas stolen before they could make any money. By giving inventors the exclusive right to market and sell their inventions, patents provided a financial incentive to produce lots of inventions. Indeed, after patents came into being, the world saw its first full-time inventors — people who made a living by inventing things. (Hello, Thomas Edison, Alexander Graham Bell, and Grace Hopper!)
- » **Widespread literacy and education:** After the Industrial Revolution began, many nations began to make primary education mandatory. They understood that without widespread literacy, there would be few inventors and little

progress. In addition, an educated workforce made spreading and implementing new technologies much easier. Soon, middle school and high school were mandated, too, paving the way for rapid and sustained economic growth.

Institutions and policies like these have given people a world of growth and opportunity. We enjoy a material abundance so unprecedented in world history that the most significant public health problem in many countries today is obesity. (Maybe if we had to forage for our own Lean Cuisine. . .)

Looking toward the future

Moving forward, the challenge is getting even more of what people want out of the world's limited pool of resources. We must face this challenge because problems like infant mortality, child labor, malnutrition, endemic disease, illiteracy, and unemployment are alleviated by higher living standards.

Economists believe that many of the poverty-related problems found in less economically developed areas might be alleviated if those areas adopt some of the institutions that have helped already-developed areas reach high levels of material prosperity.

On the other hand, economists point out that developing areas should learn from the mistakes of the past. There's no need to blindly copy the "smokestack industrialization" and heavy pollution that characterized the economic rise of England, Japan, and the United States in the 19th and 20th centuries. Modern development can be clean and green.



TIP

To summarize, three interrelated and excellent reasons should motivate you to read this book and get a firm grasp of economics:

- » **You can discover how modern economies function.** Doing so can help you understand how they've raised living standards and where they could improve.
- » **By getting a thorough handle on fundamental economic principles, you can judge the economic policy proposals that politicians and activists promote.** After reading this book, you'll be much better able to sort the good from the bad.
- » **You'll be able to spout off some interesting factoids at parties.** Hey, whatever it takes to get a date, right?

Framing Economics as the Science of Scarcity

Chapter 2 delves into scarcity and the tradeoffs that it forces people to make.

The fundamental and unavoidable constraint that generates a need for the science of economics is *scarcity*: There isn't nearly enough time or stuff to satisfy human desires.

To compensate, people have to make hard choices about what to produce and what to consume. If they do so wisely, they can at least have the best of what's possible, even though what they end up with will likely fall far short of everything they dreamed of.

The processes that people follow when attempting to deal with scarcity turn out to be intimately connected with a phenomenon known as *diminishing returns*, which describes the sad fact that each additional amount of a resource that's thrown at a production process brings forth successively smaller amounts of output.

Like scarcity, diminishing returns is unavoidable. In Chapter 3, I explain how people deal with diminishing returns to get the most out of humanity's limited pool of resources.

Sending Microeconomics and Macroeconomics to Separate Corners

The central organizing principle I use in this book is to divide economics into its two broad pieces: macroeconomics and microeconomics:

- » **Microeconomics** focuses on individual people and individual firms. It explains how individuals behave when deciding where to spend their money or how to invest their savings. And it describes how profit-maximizing firms behave when competing against each other.
- » **Macroeconomics** looks at the economy as an organic whole by focusing on statistics that reflect economy-wide trends, such as interest rates, inflation, and unemployment. It also encompasses the study of economic growth and the methods governments use to try to moderate the harm caused by recessions.

Some basic principles, such as scarcity and diminishing returns, underly both microeconomics and macroeconomics. Consequently, I spend the rest of Part 1 explaining these fundamentals before diving into microeconomics in Part 2 and macroeconomics in Part 4.

It's a good idea, though, for us to pause a little and use the next few pages to give you a slightly more detailed overview of microeconomics and macroeconomics. That way, you'll be able to jump around and read the book out of order without losing context. Don't worry. It's not like turning to the last page of a murder mystery to find out who done it.

Getting up close and personal: Microeconomics

Microeconomics gets down to the nitty-gritty, studying the most fundamental economic agents: individuals and firms. Topics include supply and demand, competition, property rights, market failures, and how the government might want to regulate giant internet companies like Google and Microsoft.

Balancing supply and demand

In a modern economy, individuals and firms produce and consume everything that's made. Producers determine supply, consumers determine demand, and their interaction in markets determines what's made and how much it costs. (See Chapter 4 for details.)

Individuals make economic decisions about how to get the most happiness out of their limited incomes. They do this first by assessing how much *utility*, or satisfaction, each possible course of action would give them. They then weigh costs and benefits to select the course of action that will yield the greatest amount of utility possible given their limited incomes. These decisions generate the demand curves that affect prices and output levels in markets. I cover these decisions and demand curves in Chapter 5.

In a similar way, the profit-maximizing decisions of firms generate the supply curves that affect markets. Every firm will decide what to produce and how much to produce by comparing costs and revenues. A unit of output will only be produced if doing so will increase its maker's profit. In particular, a firm will only produce a unit if the increase in revenue from selling it exceeds the unit's cost of production. This behavior underpins the upward slope of supply curves and affect prices and output levels in markets, as I discuss in Chapter 6.

Considering why competition is so great

You may not feel warm and fuzzy about profit-maximizing firms, but economists love them — just as long as they're stuck in competitive industries. That is true because firms that are forced to compete end up satisfying two wonderful conditions:

- » They're *allocatively efficient*, producing the goods and services that consumers most greatly desire.
- » They're *productively efficient*, producing those most-desired products at the lowest possible cost.



REMEMBER

The allocative and productive efficiency of competitive firms is the basis of Scottish economist and philosopher Adam Smith's *invisible hand* — the famous idea that, when constrained by competition, each firm's greed ends up causing it to act in a socially optimal way, as if guided to do the right thing by an invisible hand. I discuss this idea, and much more about the benefits of competition, in Chapter 7.

Examining problems caused by lack of competition

Unfortunately, not every firm is constrained by competition. And when that happens, firms don't end up acting in socially optimal ways. The most extreme case is *monopoly*, which is characterized by only one firm in an industry — meaning that the lone firm has absolutely no competition. Monopolies misbehave, restricting output to drive up prices and inflate profits. These actions hurt consumers and may go on indefinitely unless the government intervenes.

A less extreme case of lack of competition is *oligopoly*, which is a situation in which there are only a few firms in an industry. In such situations, firms often make deals not to compete against each other so that they can keep prices high and make big profits. However, these firms often have a hard time keeping their agreements with each other. This fact means that oligopoly firms often end up competing against each other despite their best efforts not to. Consequently, government regulation isn't always needed. You can read more about monopolies in Chapter 8 and oligopolies in Chapter 9.

Reforming property rights



REMEMBER

You can rely upon markets and competition to produce socially beneficial results only if society sets up a sound system of property rights. A *property right* gives a person the exclusive authority to determine how a productive resource can be used. Thus, for example, a person who has the property right (ownership) over a piece of land can determine whether it will be used for farming, as the next

Disneyland, or as a nature preserve. All pollution issues, as well as all cases of species loss, are the direct result of poorly designed property rights generating perverse incentives that lead people to do bad things. Economists take this problem seriously and have done their best to reform property rights to alleviate pollution and eliminate species loss. I discuss these issues in detail in Chapter 10.

Dealing with other common market failures

Monopolies, oligopolies, and poorly designed property rights lead to what economists like to call *market failures* — situations in which markets don't deliver socially optimal outcomes. Market failures are also caused by asymmetric information and public goods:

- » **Asymmetric information:** Asymmetric information occurs when a potential buyer or a potential seller knows more about the quality of the good that they are negotiating over than does the other party. Because of the uneven playing field and the suspicion it creates, many potentially beneficial economic transactions are never completed.
- » **Public goods:** Public goods are goods or services that are impossible to provide to just one person; if you provide them to one person, you have to provide them to everybody. (It's an "I scream, you scream, we all scream for ice cream" kind of deal.) The problem is that most people try to get the benefit without paying for it.

I discuss both of these situations, and ways to deal with them, in Chapter 11.

Investigating networks, platforms, and the economics of the internet

Large oligopoly firms like Google, Meta, and Microsoft run the networks and platforms that power the internet. But unlike the brick-and-mortar oligopolies of earlier eras, they often give away their main products for free. We study why they do that in the first part of Chapter 12 before investigating how governments are approaching the regulation of these powerful companies in the latter part of Chapter 12.

Understanding behavioral economics

People aren't always rational. That matters because much of economics was developed by asking what a rational person would do in one situation or another. *Behavioral economics* fills in the gaps by looking at decision-making when people aren't acting perfectly rationally. Four billion years of evolution have left us with brains prone to errors, including being overconfident, being overly focused on the

present, being easily confused by irrelevant information, and being unable to see the bigger picture when making financial decisions. I spend Chapter 13 rationally explaining all this irrational behavior. It's crazy fun.

Zooming out: Macroeconomics and the big picture

Macroeconomics treats the economy as a unified whole. Studying macroeconomics is insightful because certain factors, such as interest rates and tax policy, have economy-wide effects and because when the economy goes into a recession or a boom, every person and every business is affected. Here are some of the main topic areas within macroeconomics.

Measuring the economy

Hard-working government accountants calculate *gross domestic product* (GDP), the value of all goods and services produced in a nation's economy in a given period of time, usually a quarter or a year. Totaling this number is vital because if you can't measure how the economy is doing, you can't tell whether government policies intended to improve the economy are helping or hurting. Chapter 14 explains GDP in more depth.

Inflation measures how prices in the economy increase over time. This topic, inflation, is the focus of Chapter 15, and it's crucial because high rates of inflation usually accompany huge economic problems, including deep recessions and countries defaulting on their debts.

It's also important to study inflation because poor government policy is the sole culprit behind high rates of inflation — meaning that governments are responsible when big inflations happen.

Looking at international trade

International trade occurs when consumers, firms, or governments purchase products or resources from other countries. Imported goods often compete with locally produced ones. That's why international trade is the subject of endless political controversy and attempts to erect import duties or numerical quotas to keep foreign goods out and make life easier for domestic producers.

Those disputes are intensified by concerns about whether foreign working conditions are humane, whether foreign producers are unfairly subsidized by their governments, and whether foreign governments are manipulating currency exchange rates to give their own firms a cost advantage. Chapter 14 explains how economists analyze these and other globalization issues.

Understanding and fighting recessions



REMEMBER

A *recession* occurs when the total amount of goods and services produced in an economy declines. Recessions are painful for two reasons:

- » Less output means less consumption.
- » Many workers lose their jobs.

Recessions linger because institutional factors in the economy make it hard for prices to fall. If prices *could* decline quickly and easily, recessions would quickly resolve themselves. But because prices can't quickly and easily fall, economists have had to develop antirecessionary policies to get economies out of recessions as soon as possible.

The person most responsible for developing those antirecessionary policies was the English economist John Maynard Keynes. Chapter 16 introduces you to his model of the economy and how it explicitly takes account of the fact that prices can't quickly and easily fall to get you out of a recession.



REMEMBER

Chapter 17 applies Keynes's model of the macroeconomy to discuss the two broad sets of government policies to shorten and soften recessions:

- » **Monetary policy:** Monetary policy uses changes in the money supply to change interest rates and stimulate economic activity. For instance, if the government causes interest rates to fall, consumers borrow more money to buy things like houses and cars, thereby stimulating economic activity and getting the economy moving faster.
- » **Fiscal policy:** Fiscal policy refers to using increased government spending or lower tax rates to fight recessions. For instance, economic activity increases if the government buys more goods and services. In a similar fashion, if the government cuts tax rates, consumers end up with higher after-tax incomes, which, when spent, increase economic activity.

When Keynes's antirecessionary ideas began to be implemented after World War II, they seemed to work extremely well. But things didn't go so well in the 1970s, and it soon became apparent that although monetary and fiscal policy were powerful antirecessionary tools, they had their limits.

For this reason, Chapter 17 also covers how and why monetary policy and fiscal policy are constrained in their effectiveness. The key concept is called *rational expectations*. It explains how rational people very often change their behavior in response to policy changes in ways that limit the effectiveness of those changes. It's a concept you need to understand if you're going to generate informed opinions about current macroeconomic policy debates.

Financial crises are recessions triggered by the failure of important financial institutions to keep their financial promises. Such failures often happen after consumers or businesses take on too much debt and are unable to repay loans to banks. Sometimes they occur when a government takes on too much debt and can't repay its bondholders. Chapter 18 discusses the causes and consequences of financial crises.

Understanding How Economists Use Models and Graphs

Economists like to be logical and precise, which is why they use a lot of equations and math. But they also like to present their ideas in easy-to-understand and highly intuitive ways, which is why they use so many graphs.

To avoid a graph-induced panic as you flip through the pages of this book, let's spend the next few pages helping you get acquainted with the sorts of graphs that economists like to use. Take a deep breath; I promise this won't hurt.

Abstracting from reality is a good thing

The graphs that economists use are almost always visual representations of economic models. An *economic model* is a mathematical simplification of reality that allows you to focus on what's truly important by ignoring lots of irrelevant details.



TIP

For instance, the economist's model of consumer demand focuses on how prices affect the amounts of goods and services that people want to buy. Other things, such as changing styles and tastes, also affect consumer demand. But because a product's own price is the key factor that affects the demand for a given product, that price is emphasized both in the mathematical model of consumer demand as well as in the graphs and figures that economists draw to illustrate consumer demand.

Let's use orange juice as an example. The price of orange juice is the primary factor that affects how much people are going to buy. (I don't care which dietary trend is in vogue — if orange juice costs \$50 a gallon, you'll probably find another diet.) Therefore, it's helpful to abstract from those other things so that we can concentrate on how the price of orange juice affects the quantity of orange juice that people want to buy. (*Concentrate on orange juice. . .get it?!*)

Introducing your first model:

The demand curve

Suppose that some economists go out and survey consumers, asking them how many gallons of orange juice they'd buy each month at three hypothetical prices: \$10 per gallon, \$5 per gallon, and \$1 per gallon. The results are summarized in the following table.

Gallons of Orange Juice That Consumers Want to Buy

Price	Gallons
\$10	1
\$5	6
\$1	10

Economists refer to the quantity that people would be willing to purchase at a particular price as the *quantity demanded* at that price. If you look at the data in the preceding table, you'll find that the price of orange juice and the quantity demanded of orange juice have an *inverse relationship* with each other — when one goes up, the other goes down.



REMEMBER

Because this inverse relationship between price and quantity demanded holds for nearly all goods and services, economists refer to it as the *law of demand*. But, quite frankly, the law of demand becomes much more immediate and interesting if you can see it rather than just *think* about it.

Creating a demand curve by plotting out the data

The best way to see the quantity demanded at various prices is to plot it out on a graph. In the standard demand graph, the horizontal axis represents quantity, and the vertical axis represents price.

In Figure 1-1, I've graphed the orange juice data in the preceding table and marked three points, labeling them A, B, and C. The horizontal axis of Figure 1-1 measures the number of gallons of orange juice that people demand each month at various prices per gallon. The vertical axis measures the prices.

Point A is the visual representation of the data in the top row of the preceding orange juice table. It tells you that at a price of \$10 per gallon, people want to purchase only 1 gallon per month of orange juice. Similarly, Point B tells you that they demand 6 gallons per month at a price of \$5, and Point C tells you that they demand 10 gallons per month at a price of \$1 per gallon.

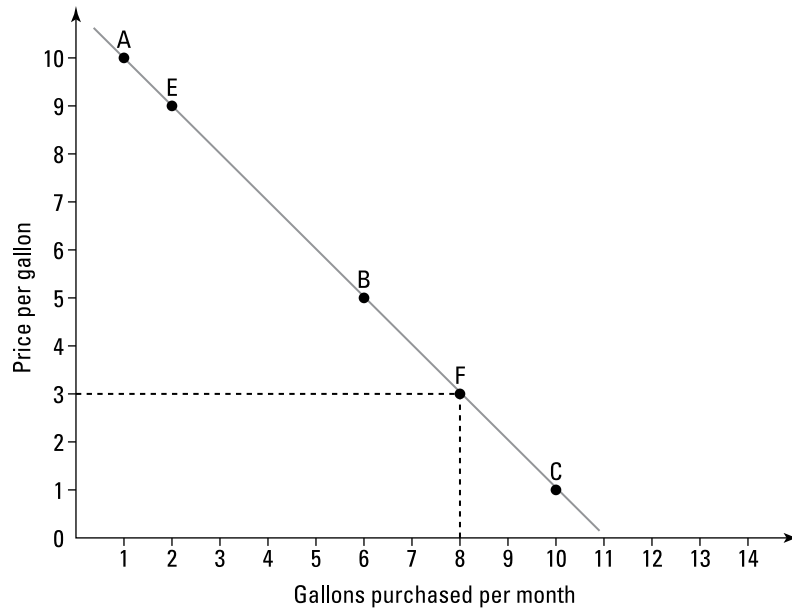


FIGURE 1-1:
Graphing the
demand for
orange juice.

© John Wiley & Sons, Inc.

Notice that I've connected Points A, B, and C with a line. I've done this to make up for the fact that the economists who conducted the survey asked about what people would do at only three prices. If they'd had a big enough budget to ask consumers about every possible price (\$8.46 per gallon, \$2.23 per gallon, and so on), there'd be an infinite number of dots on the graph. But because they didn't do that, I drew a straight line passing through points A, B, and C to estimate the quantities demanded at all the prices that weren't surveyed.

That straight line is a *demand curve*. I know it doesn't curve at all, but for simplicity, economists use the term "demand curve" to refer to all plotted relationships between price and quantity demanded, regardless of whether they're straight lines or curvy lines. (This is consistent with the fact that economists are both eggheads *and* squares.) Straight or curvy, you can visualize the fact that price and quantity demanded have an *inverse relationship*: When price goes up, quantity demanded goes down. The inverse relationship implies that demand curves slope downward.

Generalizing a bit, you can also see that the *slope* of a demand curve gives quick intuition about the sensitivity of the inverse relationship between price and quantity demanded. If a demand curve is steep, you know it would take a large change in price to cause a small change in quantity demanded. By contrast, a flat demand curve tells you that a small change in price would result in a large change in quantity demanded.

Extending that reasoning even further, you can see that demand curves with changing slopes (that is, demand curves that aren't perfectly straight lines) tell you that the relationship between price and quantity demanded varies. On the steeper parts of such curves, a change in price causes a relatively small change in quantity demanded. On the flatter part of such curves, a change in price causes a relatively large change in quantity demanded.

Using the demand curve to make predictions

Graphing out a demand curve makes it much easier to make quick predictions. For instance, you can use the straight line I've drawn in Figure 1-1 to estimate that at a price of \$9 per gallon, people would want to buy about 2 gallons of orange juice per month. I've labeled this as Point E on the graph.

Suppose that you can see only the data in the preceding orange juice table and can't look at Figure 1-1. Can you quickly estimate how many gallons per month people are likely to demand if the price of orange juice is \$3 per gallon? Looking at the table's second and third rows, you have to conclude that people will demand between 6 and 10 gallons per month. But figuring out exactly how many gallons will be demanded would take some time and require some calculations.

By contrast, if you look at Figure 1-1, it's easy to figure out how many gallons per month people would demand at \$3 per gallon. You start at \$3 on the vertical axis, move to the right until you hit the demand curve at Point F, and then drop down vertically until you reach the horizontal axis, where you discover that you're at 8 gallons per month. (To make sure you understand, I've drawn a dotted line that follows this path.) As you can see, using a figure rather than a table makes coming up with model-based predictions much simpler.

Drawing your own demand curve

Try a simple graphing exercise for yourself. Imagine that the government came out with a research report showing that people who drink orange juice have lower blood pressures, fewer strokes, and better sex lives than people who don't drink orange juice. What do you think will happen to the demand for orange juice? It should increase.

To verify this, our intrepid team of survey economists goes out again, asking people how much orange juice they would like to buy each month at \$10 per gallon, \$5 per gallon, and \$1 per gallon. The new responses are here.

Gallons of OJ That Consumers Want to Buy After Reading New Government Report

Price	Gallons
\$10	4
\$5	9
\$1	13

Your assignment, should you choose to accept it, is to plot these three points in Figure 1-1. After you've done that, connect them with a straight line. (Yes, you can write in this book if it's yours, but the library frowns on your doing that!)

You've just created a new demand curve that reflects people's new preferences for orange juice in light of the government survey. Their increased demand is reflected in the fact that they now demand a larger quantity of juice at any given price than they did before. For instance, whereas before they wanted only 1 gallon per month at a price of \$10, they now would be willing to buy 4 gallons per month at that price.

There's still an inverse relationship between price and quantity demanded, meaning that even though the health benefits of orange juice make people demand more of it, people are still sensitive to higher orange juice prices. Higher prices still mean lower quantities demanded, and your new demand curve still slopes downward.

Use your new demand curve to figure out how many gallons per month people will want to buy at a price of \$7 and at a price of \$2. Figuring these things out from the data in the preceding table would be challenging, but figuring them out using your new demand curve should be easy. Once you have a handle on things, go pour yourself some liquid vitamin C. You've earned it.

IN THIS CHAPTER

- » Deciding what will bring the most happiness
- » Cataloguing the constraints that limit choice
- » Modeling choice behavior like an economist
- » Evaluating the limitations of the economics choice model

Chapter 2

Cookies or Ice Cream? Exploring Consumer Choices

Economics is about how groups and individuals make choices and why they choose the things they do. Economists have spent a great deal of time analyzing how groups make choices, but because group choice behavior often turns out to be broadly similar to individual choice behavior, my focus in this chapter is on individuals.

To keep things simple, my explanation of individual choice behavior focuses on *consumer behavior* because most of the choices people make on a day-to-day basis involve which goods and services to consume. Human beings are constantly forced to choose because their wants almost always exceed their means. Limited resources, or *scarcity*, is at the heart not only of economics but also of ecology and biology. Darwinian evolution is about animals and plants competing over limited resources to produce the largest number of offspring. Economics is about human beings choosing among limited options to maximize happiness.

Describing Human Behavior with a Choice Model

Human beings may be complicated creatures with sometimes mystifying behavior, but most people are fairly predictable and consistent. You can gain a lot of knowledge by studying choice behavior because if you can understand the choices people made in the past, you stand a good chance of understanding the choices they'll make in the future.

Understanding (and even predicting) future choice behavior is critical because significant shifts in the economic environment are typically the result of millions of individual decisions that add up to a major trend. For instance, the circumstances under which millions of individuals choose to pursue work or school have major effects on the unemployment rate. And the choices these individuals make about how much of their paychecks to save or spend affect whether interest rates will be high or low and whether gross domestic product (GDP) and overall economic output will increase or decrease. (I discuss GDP in Chapter 14.)



REMEMBER

To predict how self-interested individuals make their choices, economists have created a model of human behavior that assumes that people are rational and can calculate subtle trade-offs between possible choices. This model is a three-stage process:

- 1. Evaluate how happy each option can make you.**
- 2. Look at the constraints and trade-offs limiting your options.**
- 3. Choose the option that will maximize your overall happiness.**

Although not a complete description of human choice behavior, this model generally makes accurate predictions. However, many people question this explanation of human behavior. Here are three common objections:

- » Are people really so self-interested? Aren't they often motivated by what's best for others?
- » Are people really aware of all their options? How are they supposed to rationally choose among new things that they've never tried before?
- » Are people really free to make decisions? Aren't they constrained by legal, moral, and social standards?

I spend the following few sections of this chapter expanding on the three-stage, economic-choice model and addressing common objections.

Pursuing Personal Happiness

Economists like to think of human beings as free agents with free will. To economists, people are usually rational and, thus, normally capable of making sensible decisions. But that begs the question of what motivates people and, in turn, what sorts of things people will choose to do.

In a nutshell, economists assume that the primary motivation driving most people is a desire to be happy. This assumption implies that people make choices based on whether or not those choices will make them as happy as they can be given their circumstances.

Using utility to measure happiness

The first stage of the economic choice model examines how the pursuit of happiness affects consumer behavior.

If people make choices based on maximizing happiness, they need a way of comparing how much happiness each option brings with it.

Along these lines, economists assume that people get a sense of satisfaction or pleasure from the things life offers. Sunsets are nice. Eating ice cream is excellent. Friendship can be joyful. And I happen to like driving fast.



REMEMBER

Economists suppose that you can compare all possible things that you may experience with a common measure of happiness or satisfaction called *utility*. Things you like a lot have high utility. Things you like only a little have low utility. And things you hate (like toxic waste or foods that cause you to have allergic reactions) have negative utility. Utility is a common denominator that allows people to compare even radically different options in a systematic manner.

The concept of utility is broad. For a hedonist, utility may be the physical gratification of experiencing sensual pleasures. But for a morally conscientious person, utility may be the sense of moral satisfaction when doing the right thing in a particular situation. The critical idea for economists is that people can sort out and compare the utilities of various activities.

Taking “selfless” actions into account

Economists take it as a given that people make their choices in life to maximize their personal happiness. This viewpoint immediately raises objections because people are often willing to endure great personal suffering to help others.

However, an economist can view *altruism*, helping others at one's own expense, as a personal preference. The mother who doesn't eat so she can give the little food she has to her infant may be pursuing a goal (helping her child) that maximizes the mother's happiness. The same can be said about people who donate to charities. Most people consider such generosity selfless, but it's also consistent with assuming that people do things to make themselves happy. If people give because doing so makes them feel good, their selfless action is motivated by self-interest.



REMEMBER

Because economists view human motivation as intrinsically self-interested, economics is often accused of being immoral; however, economics is concerned with how people achieve their goals rather than with the morality of those goals. For instance, some people like honey, but others don't. Economists make no distinction between these two groups regarding the rightness or wrongness of their preferences. Rather, what interests economists is how each group behaves, given its preferences. Consequently, economics is *amoral* rather than immoral.

Economists are people, too, and they're concerned with things like social justice, climate change, and poverty. They just tend to interpret the desire to pursue morality and equity as an individual goal that maximizes individual happiness rather than as a group goal that should be pursued to achieve some sort of collective good.

Self-interest can promote the common good

Adam Smith, one of the fathers of modern economics, believed that if society were set up correctly, people chasing after their own happiness would also provide for other people's happiness. As he famously pointed out in *An Inquiry into the Nature and Causes of the Wealth of Nations*, published in 1776, "It is not from the benevolence of the butcher, the brewer, or the baker, that we can expect our dinner, but from their regard to their own interest."



REMEMBER

Put differently, the butcher, the brewer, and the baker make stuff for you not because they like you but because they want your money. Yet because they want your money, they end up producing for you everything that you need to have a nice meal. When you trade your money for their goods, everyone is happier. Not having to prepare that food is worth more to you than keeping your money. And getting your money is worth more to them than the toil involved in preparing that food.

Adam Smith expanded on this notion by saying that a person pursuing their own selfish interests may be "led by an invisible hand to promote an end which was no part of his intention."

Because economists recognize that benevolent outcomes can result from self-interested behavior, they're (1) less concerned with intentions than with results and (2) less focused on what makes people happy than on how people pursue happiness and whether those pursuits lead to good or bad outcomes.

You Can't Have Everything: Examining Limitations

The second stage of the economic choice model examines how limits and trade-offs affect consumer behavior.

Life has limits. Time, for instance, is always in limited supply, as are natural resources. The second stage of the economic choice model looks at the constraints that force you to choose among your happy options.

For example, oil can be used to manufacture pharmaceuticals that can save many lives. But it can also be used to make gasoline, which can be used to drive ambulances, which also save lives. Both pharmaceuticals and gasoline are good uses for oil, so society has to come up with some way of deciding how much oil gets to each of these two good uses, knowing all the while that each gallon of oil that goes to one can't be used for the other.

This section outlines the various constraints and the unavoidable cost — *opportunity cost* — of getting what you want. For more on how markets use supply and demand to allocate resources in the face of constraints, please see Chapter 4.

Resource constraints

The most obvious constraints on human happiness are the physical limitations of nature. Not only are the supplies of oil, water, and fish limited, but so are the radio frequencies on which to send signals and the hours of sunshine that can be used to generate electricity by means of photovoltaic solar panels. There's simply not enough of most natural resources for everyone to have as much as they want.

The limited supply of natural resources is allocated in different ways. In some cases, as with certain endangered species, laws guarantee that nobody can have any of a particular resource. With the electromagnetic spectrum, national governments portion out the spectrum to broadcasters or mobile phone operators. But for the most part, private property and prices control the allocation of natural resources.

Under such a system, the use of a resource goes to the highest bidder. Although this system can discriminate against the poor because they don't have much to bid with, it does ensure that the resource is used by those who value it highly — in other words, by those who have chosen to purchase the resource to maximize their happiness.

Technology constraints

You have a much higher standard of living than your ancestors did. You have a cushier life because of improvements in the technology of converting raw resources into things people like to use. Yet technology improves less quickly than people would like, and people's choices are limited at any given moment by how advanced technology is right then. Therefore, it's natural to think of technology as a constraint that restricts choices.



REMEMBER

As technology improves, people are able to produce more from the limited supply of resources on the planet. Or, put slightly differently, individuals have more and better choices as technology improves. In the past 200 years, people have figured out how to immunize children against deadly diseases, how to use electricity to provide light and mechanical power, how to build a rocket capable of putting people on the moon, and how to increase farm yields dramatically to feed more people. In just the past 30 years, the internet and cheap mobile phones have revolutionized everything from entertainment to how governments communicate with their citizens.

Time constraints

Time is a precious resource. Worse yet, it's a resource in fixed supply. Therefore, the best that technology can do for people is to allow them to produce more in the limited time they have or to grant them a few more years of life through better medical technology.

But even with a longer life span, you can be in only one place at a time. Consequently, you only have a finite amount of time to work with. This means you must choose how to allocate your limited amount of time between leisure and labor, between taking time to do things you like and selling your time to employers so that you can earn wages to pay for things you want. This trade-off implies that time is a precious commodity requiring people to make serious choices.

Opportunity cost: The unavoidable cost

The economic idea of *opportunity cost* is closely related to time constraints. You can do only one thing at a time, which means that, inevitably, you're always giving up other things.



REMEMBER

The *opportunity cost* of any activity is the value of the best alternative thing you could've done instead. For instance, this morning, I could've chatted on the phone with a friend, watched TV, or worked hard writing this chapter. I chose to chat with my friend because that made me happiest. (Don't tell my editor!) Of the two things I didn't choose, I consider working on the chapter better than watching TV. So the opportunity cost of chatting on the phone was not getting to spend the time working on this chapter.



TIP

Opportunity cost depends only on the value of the best alternative option because you can always reduce a complicated choice with many options to a simple choice between Option X and the best alternative option out of all the other options you can choose from. It doesn't matter whether you have 3 alternative options or 3,000.

Simplifying a decision down to only two options makes choosing easy. You should go with Option X (rather than the best alternative option) only if the pleasure you will receive from Option X exceeds the opportunity cost of not getting to enjoy the best alternative option. And you should select the best alternative option only if the opportunity cost of forgoing it exceeds the pleasure you would get from consuming Option X.

Suppose that you can choose only one item from a selection of desserts that includes butter pecan ice cream, glazed donuts, chocolate chip cookies, and peach cobbler. Select one of these at random — say, butter pecan ice cream. Then, out of all the other desserts, identify the one that you like best out of that group. In my case, it'd be chocolate chip cookies.

My decision about which dessert to eat now comes down to comparing how I feel about butter pecan ice cream and chocolate chip cookies. Selecting the ice cream means enduring the opportunity cost of not eating the cookies. I'll do that only if the pleasure from eating the ice cream exceeds the opportunity cost of forgoing the chocolate chip cookies. And I'll opt for the chocolate chip cookies only if the opportunity cost of forgoing them exceeds the pleasure I'd get from eating the ice cream.

Making Your Choice: Deciding What and How Much You Want

At its most basic, the third stage of the economic choice model is nothing more than cost-benefit analysis. In the third stage, you simply choose the option for which the benefits outweigh the costs by the largest margin.

The cost-benefit model of how people make decisions is potent because it seems to correctly describe how most decisions are made. But this version of cost-benefit analysis can tell you only *whether* people choose a given option. In other words, it's only good at describing all-or-nothing decisions like whether or not to eat ice cream. A much more powerful version of cost-benefit analysis uses the concept of *marginal utility* to tell you not just whether I'm going to eat ice cream but also *how much* of it I'll eat.

To see how marginal utility works, recognize that the amount of utility a given item brings usually depends on how much of that item a person has already had. For instance, if you've been really hungry, the first slice of pizza brings you a lot of utility. The second slice is also pleasant but not quite as good as the first because you're no longer starving. The third, in turn, brings less utility than the second. And if you keep forcing yourself to eat, you may find that the 12th or 13th slice of pizza makes you sick and brings you negative utility.



REMEMBER

Economists refer to this phenomenon as *diminishing marginal utility*. Each additional, or *marginal*, unit that is consumed brings less utility than the previous unit, so that the extra utility, or *marginal utility*, brought by each successive unit diminishes as you consume more and more units. Here, each successive slice of pizza brings with it less additional (or marginal) utility than the previous slice.

To see how diminishing marginal utility guides how people make decisions about how much of something to consume, consider having \$20 to spend on either \$4 pizza slices or \$4 baskets of fries.

If I buy four slices of pizza, I'll have \$4 left over to spend on a basket of fries. But should I spend that last \$4 on a fifth slice of pizza or a first basket of fries? That depends on whether the marginal utility that I'd get from a fifth slice of pizza would exceed the marginal utility that I'd get from a first basket of fries. If spending the last \$4 on a first basket of fries would bring me more marginal utility, then I'll spend that money on a first basket of fries. But if a fifth slice of pizza would bring more marginal utility than a first basket of fries, I'll be a pizza-only guy with my limited \$20 budget.

Generalizing that strategy, if my goal is to maximize the *total utility* I can get from my \$20 budget, I'll keep adjusting the quantities of each food (slices of pizza vs. baskets of fries) until I find the combination that maximizes how much total utility I can purchase using my \$20.

Because different people have different preferences, the quantities of each good that will maximize each person's total utility are usually different. Someone who detests pizza will spend all their money on fries. A person who can't stand fries will spend all their money on pizza. And for people who choose to have some of both, the optimal quantities of each depend on their feelings about the two goods and how fast their respective marginal utilities decrease. Check out Chapter 5 for more detail on diminishing marginal utility and how it causes demand curves to slope downward.

Allowing for diminishing marginal utility makes this model of choice behavior compelling. It tells you not only what people will choose but how much of each thing they'll choose. It's not perfect, however. For example, it assumes that people have a clear sense of the utility of various things, a good idea of how fast marginal utilities diminish, and no trouble making comparisons. I discuss these substantial criticisms in the next section.

MARGINAL UTILITY IS FOR THE BIRDS!

Economists are confident that cost-benefit analysis and diminishing marginal utility are good descriptions of decision-making because there's plenty of evidence that other species also behave in ways consistent with these concepts.

For instance, scientists can train birds to peck at one button to earn food and another button to earn time on a treadmill. If scientists increase the cost of one of the options by increasing the number of clicks required to get it, the birds respond rationally by not clicking so much on the button for that option. But even more interesting is that they also switch to clicking more on the button for the other option.

The birds seem to understand that they have only a limited number of clicks they can make before they become exhausted, and they allocate these clicks between the two options to maximize their total utility. Consequently, when the relative costs and benefits of the options change, they change their behavior quite rationally in response.

Most species also seem to be affected by diminishing marginal utility and become indifferent to marginal (that is, additional) units of something that they've recently enjoyed a lot of. So although economists' models of human behavior may seem to ignore some relevant factors, they do take into account some fundamental and universal behaviors.

Exploring Violations and Limitations of the Economist's Choice Model

For simplicity, economists often assume that people are fully informed and completely rational when they make decisions. You may think that gives people too much credit, but models based on those assumptions work surprisingly well.

However, in the real world, people aren't always informed about the decisions they need to make, and they aren't always as reasonable as economists assume. In this section, I note some of the limitations of the choice model and explain why they may not matter all that much in the long run.

Understanding uninformed decision-making

When economists apply the choice model, they assume a situation in which a person knows all the possible options, how much utility each will bring, and the opportunity cost associated with each option. But how do you evaluate whether it would be better to sit on top of Mount Everest for five minutes or hang glide over the Amazon for ten minutes? Because you've never done either, you aren't well informed about the constraints and costs of the choices and probably don't know what the utilities of the two options would be if you ever did get to experience them.

Politicians touting novel programs often ask voters to make similarly uninformed choices. They make their proposals sound as good as possible, but in many cases, nobody really knows what they may be getting into.

Things are similarly murky for choices involving luck or uncertainty. People buying lottery tickets in state lotteries have no idea about the eventual possible gain because the size of the prize depends on how many tickets are sold before the drawing is made. The people who choose to play lotteries also tend to have highly exaggerated "guesstimates" about their chances of winning.

Economists account for this reality by assuming that when faced with uninformed decisions, people make their best guesses about not only uncertain outcomes but also about how much they may like or dislike things with which they have no previous experience. This may seem like a fudge, but people in the real world *do* making decisions in such situations — meaning that they must be fudging a bit as well.

Whether people make good choices when they're poorly informed is hard to say. People would prefer to be better informed before choosing. And some people do shy away from less certain options. But overall, the economist's model of choice behavior seems quite capable of dealing with situations of incomplete information and uncertainty about random outcomes.

Making sense of irrationality

Even when people are fully informed about their options, they often make logical errors in evaluating costs and benefits. Don't be alarmed if you've made these errors yourself. After people have these choice errors explained to them, they typically stop making the errors and start behaving in a manner consistent with rationally weighing marginal benefits against marginal costs. I go through three of the most common mistakes in the following subsections.

Sunk costs are sunk!



REMEMBER

Economists refer to costs that have already been incurred and that shouldn't affect your current and future decision-making as *sunk costs*. Rationally speaking, you should consider only the future, potential marginal costs and benefits of your current options.

Suppose you just spent \$25 to get into an all-you-can-eat sushi restaurant. How much should you eat? More specifically, when deciding how much to eat, should you care about how much you paid to get into the restaurant? To an economist, the answer to the first question is, "Eat exactly the amount of food that makes you most happy." And the answer to the second question is, "How much it cost you to get in doesn't matter because whether you eat 1 piece of sushi or 80, the cost was the same."

Put differently, because the cost of getting into the restaurant is now in the past, it should be completely unrelated to your current decision of how much to eat. To see why, suppose that you were suddenly offered \$5,000 to leave the sushi restaurant and eat next door at a competitor's restaurant. Would you refuse just because you felt obligated to stay and eat sushi "to get your money's worth" out of the \$25 you spent getting into the sushi restaurant? Not if you're sane.

Unfortunately, many people let sunk costs affect their decision-making until an economist points out that sunk costs are irrelevant — or, as economists never tire of saying, "Sunk costs are sunk!" (Conversely, noneconomists quickly tire of hearing this phrase.)

Mistaking a big percentage for a big dollar amount

Suppose you decide to save 10 percent on a mobile phone by making a one-hour round trip to a store in another town. You plan to buy the phone for \$360 instead of buying it at your local store for \$400. Next, ask yourself whether you'd also be willing to drive one hour to buy a home theater system for \$1,960 in the next town rather than for \$2,000 at your local store. You do the math, and because you'd save only 2 percent, you decide to buy the system for \$2,000 at the local store.

You may think you're being smart, but you've just behaved in a colossally inconsistent and irrational way. In the first case, you were willing to drive one hour to save \$40. In the second, you weren't. Costs and benefits are absolute, but people make the mistake of thinking of costs and benefits as percentages or proportions. Instead, you should compare the total dollar amount of marginal costs against the total dollar amount of marginal benefits because the potential gain is the absolute dollar amount you save, not the percentage you save.

Confusing marginal and average

Suppose your local government has recently built three bridges at a total cost of \$30 million. That's an average cost of \$10 million per bridge. A local economist does a study and estimates that the total benefits of the three bridges to the local economy add up to \$36 million, or an average of \$12 million per bridge.

A politician then starts trying to build a fourth bridge, arguing that because bridges on average cost \$10 million to build but on average bring \$12 million in benefits, it would be foolish not to build another bridge. Should you believe the politician? After all, if each bridge brings society a net gain of \$2 million, you'd want to keep building bridges forever.



REMEMBER

What really matters in this and other choice situations are *marginal* costs and *marginal* benefits, not *average* costs and *average* benefits. (See the earlier section “Making Your Choice: Deciding What and How Much You Want” to review marginal utility.) Who cares what costs and benefits all the previous bridges brought with them? You have to compare the costs of that extra, marginal bridge with the benefits of that additional, marginal bridge. You should build the bridge only if the marginal benefits exceed the marginal costs. If the marginal costs exceed the marginal benefits, you should walk away.

For example, suppose that an independent watchdog group hires an engineer to estimate the cost of building one more bridge and an economist to estimate the benefits of erecting one more bridge. The engineer finds that because the three shortest river crossings have already been taken by the first three bridges, the fourth bridge will have to be much longer. The extra length will raise the construction cost to \$15 million.

At the same time, the economist does a survey and finds that a fourth bridge isn't all that necessary. At best, it will bring only \$8 million in benefits. Consequently, this fourth bridge shouldn't be built because its marginal cost of \$15 million exceeds its marginal benefit of \$8 million. By telling voters only about the earlier bridges' *average* costs and *average* benefits, the politician supporting the project is grossly misleading them. So watch out anytime somebody tries to sell you a bridge.

- » Determining your production possibilities
- » Allocating resources in the face of diminishing returns
- » Choosing outputs to maximize happiness
- » Understanding the roles of government and markets in production and distribution

Chapter 3

Producing Stuff to Maximize Happiness

Although it's true that human beings face scarcity and can't have everything they want (as I discuss in Chapter 2), it's also true that they have a lot of options. Productive technology is now so advanced that people can convert the planet's limited supply of resources into a fantastic variety of goods and services, including delivery drones, AI chatbots, genetically engineered cancer treatments, fully immersive 3D video games — and totally awesome *For Dummies* books like this one.

Thanks to advanced technologies, people are spoiled for choices. The enormous variety of goods and services currently available means that people must choose wisely if they want to convert the planet's limited resources into the goods and services that will provide the greatest possible happiness when consumed.

This chapter explains how economists analyze the process by which societies choose precisely what to produce to maximize human happiness. For every society, the process can be divided into two simple steps:

1. **Figure out all the possible combinations of goods and services that it could produce given its limited resources and currently available technology.**

2. **Choose one of those output combinations — presumably, the combination that maximizes happiness.**

Economists view success in each of the two steps in terms of two particular types of efficiency:

- » **Productive efficiency:** Producing any given good or service using the fewest possible resources.
- » **Allocative efficiency:** Producing the kinds of goods and services that will make people the happiest, and producing each of them in the correct amounts.

This chapter shows you how a society achieves both productive and allocative efficiency — that is, how a society can produce the things that people most want at the lowest possible cost. I give you the lowdown on diminishing returns, production possibilities frontier graphs (yay, graphs!), and the interplay between markets and governments.

Figuring Out What's Possible to Produce

In determining what's possible to produce in an economy, economists list two primary factors that affect both the maximum amounts and the types of output that will be produced:

- » **Limited resources:** The first factor is obvious. If resources were unlimited, goods and services would be as well.
- » **Diminishing returns:** The more you make of something, the more expensive it becomes to produce. Even with mass production, the cost of additional units will begin to rise after some output level. Eventually, the increasing cost exceeds the benefit of producing additional units. That limits how much of the product in question you will want to produce, even if it's your favorite thing. In such situations, resources should be reallocated to producing units of other products for which the benefits still outweigh the costs.

A key result of diminishing returns is that societies are usually better off when they devote their limited resources to producing moderate amounts of many goods rather than a massive amount of just one thing.

This section gives you the lowdown on how limited resources and diminishing returns determine production possibilities. It also shows you how to represent these possibilities graphically.

Classifying resources

You can't get output without inputs of resources. Economists traditionally divide inputs, or factors of production, into four classes:

- » **Land:** Land isn't just real estate but all naturally occurring resources that can produce things people want to consume. Land includes the weather, plant and animal life, geothermal energy, and the electromagnetic spectrum.
- » **Labor:** Labor is the work people must do to produce things. A tree doesn't become a house without human intervention.
- » **Capital:** Capital comes in two main varieties.

Physical capital refers to human-made machines, tools, and structures that aren't directly consumed but are instead used to produce goods and services that people do directly consume. For instance, a car that you drive for pleasure is a consumption good, but an identical vehicle that you use to haul around bricks for your construction business is physical capital. Physical capital includes factories, roads, sewers, electrical grids, the internet, and so on.

Intellectual capital encompasses society's accumulated stock of knowledge about how to make things and do things. Examples include the recipe for cement, the computer algorithm that generates Google search results, and understanding how to run a food-processing factory in a clean and hygienic manner.

- » **Entrepreneurial ability:** The human resource, distinct from labor, that combines the other three factors of production (land, labor, and capital) to produce new products or make innovations in the production of existing products. The difference between labor and entrepreneurial ability is that labor simply works at a known task. In contrast, entrepreneurial ability is the skill of improving how we make an existing product or the wherewithal to invent a completely new product. Without entrepreneurial ability, we'd be stuck making the same things the same way, forever.

Clarifying human capital

With respect to the factor of production known as labor, economists often speak of *human capital*, by which they mean the knowledge, skills, experience, and attributes individuals possess that can be used in the production of goods or services. For instance, I have a lot of human capital where teaching economics is concerned but extremely low human capital where painting and singing are concerned.

Say that you put a person to work at a job for which she has high human capital. She'll produce much better or much more output than a person with low human capital, even though they both supply the same amount of labor in terms of hours

worked. An important consequence is that skilled workers (high human capital) get paid more than unskilled workers (low human capital). Therefore, a good way for societies to become more prosperous is to improve the skills of their workers through education and training. If societies can raise workers' human capital levels, they can produce more with the same land, labor, and capital inputs. Their workers will also be paid more and enjoy higher standards of living.

However, building up human capital is costly, and at any given instant, you should think of the level of human capital in a society as being fixed. Combined with limitations on the amount of land, labor, capital, and entrepreneurial ability, the limit on human capital means that the society will be able to produce only a limited amount of output. And along these same lines, the decisions about where to best allocate these limited resources become crucial because the resources must be used to produce the goods and services that will bring the greatest amount of happiness. (For more on limited resources and production possibilities, see the upcoming section "Allocating resources.")

Diminishing returns



REMEMBER

Diminishing returns is probably the most important economic factor in determining precisely what to produce out of all the things that could be produced, given the limited supply of resources. For virtually everything people make, the additional output you get from each unit of input decreases as you use more and more of the input.

Diminishing returns is sometimes referred to as the low-hanging fruit principle. Imagine being sent into an apple orchard at harvest time to pick apples. During the first hour, you pick a lot of apples because you go for the low-hanging ones that are the easiest to reach. In the second hour, however, you can't pick as many because you have to start reaching awkwardly for fruit that's higher up. During the third hour, you pick even fewer apples; you now have to jump off the ground every time you try to pick an apple because the only ones left are even farther away. Table 3-1 demonstrates how your productivity — your output for a given amount of input — diminishes with each additional hour you work.

TABLE 3-1

Diminishing Returns to Apple Picking

Hour Worked	Apples Picked	Labor Cost per Apple
First	300	6 cents
Second	200	9 cents
Third	120	15 cents

Another way to see the effect of diminishing returns is to note the increasing costs for producing output. If you pay workers \$18 per hour to pick apples, your cost to have 300 apples picked in the first hour is 6 cents per apple (Table 3-1). The second hour yields only 200 apples, costing you 9 cents per apple (because you still have to pay the worker \$18 for that hour's work). Only 120 get picked in the third hour, so the labor cost rises to 15 cents per apple. Eventually, the effects of diminishing returns drive prices so high that you stop devoting further labor resources to picking additional apples.

Virtually all production processes show diminishing returns, and not just for labor. Additional amounts of any particular input usually result in smaller and smaller increments of output, holding all other inputs constant.

Allocating resources

Because diminishing returns guarantees that a production process will eventually become too costly, a society normally allocates its limited resources to many different production processes.

Imagine that you can allocate workers to either picking apples or picking oranges. You can sell both apples and oranges for \$1 each, but the production of both fruits involves diminishing returns so that additional workers acting as fruit pickers yield successively smaller increments of output, no matter which fruit they're picking.

Allocating all your workers to picking oranges, for example, is unproductive because the output you'll get from the last worker picking oranges will be much less than the output you'd get from the first worker picking oranges. The intelligent thing to do is to take a worker away from picking oranges and reassign him to picking apples. As the last worker picking oranges, he didn't produce much. But as the first worker picking apples, he'll pick a lot of them. Because you pay him the same wage regardless of which fruit he's picking, you use your labor more intelligently by having him pick apples, as one apple sells for as much money as one orange.

You may also want to reassign a second worker, and perhaps a third or a fourth. But you don't want to reassign all the workers because diminishing returns applies just as much to picking apples as to picking oranges. Each additional worker assigned to picking apples produces less than the previous worker picking apples. At some point, moving additional workers from picking oranges to picking apples no longer benefits you. When that happens, you will have reached what economists refer to as an optimal allocation of your labor resource. As soon as you've found this sweet spot, you have no further incentive to move workers from picking one fruit to picking the other because no additional moving of workers will increase total fruit picking. At this point, you've maximized your fruit-picking potential.

Graphing your production possibilities

Economists have a handy graph called the production possibilities frontier (PPF) that lets you visualize the effect of diminishing returns and view the trade-offs you make when you reallocate inputs from producing one thing to producing another. The production possibilities frontier, which is sometimes referred to as the production possibilities curve, also shows how limited resources constrain your ability to produce output. Figure 3-1 shows a PPF graph that corresponds to the data in Table 3-2.

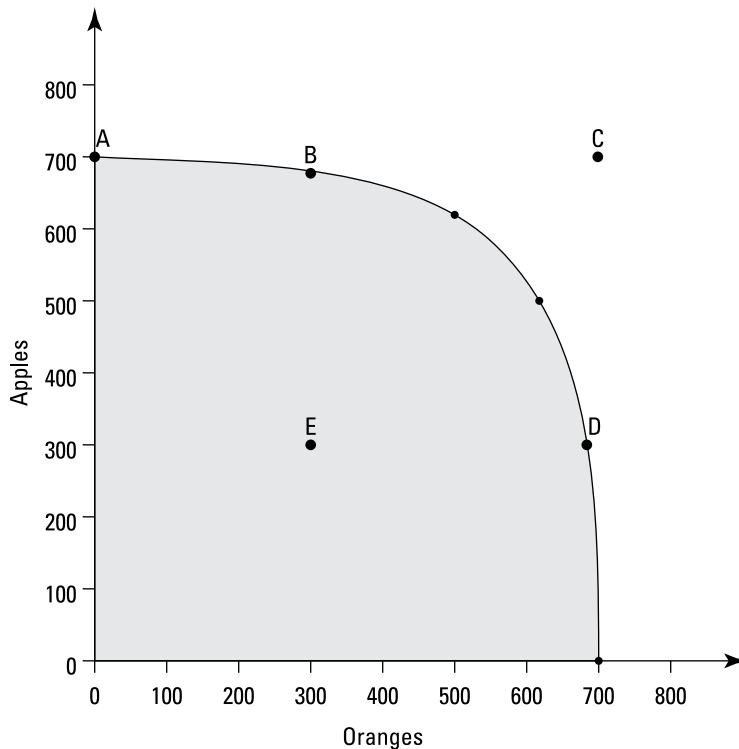


FIGURE 3-1:
The production possibilities frontier (PPF) for the data in Table 3-2.

© John Wiley & Sons, Inc.

Table 3-2 shows how the total output of apples and the total output of oranges changes as you make different allocations of five available workers to picking apples or oranges. For instance, if you put all five people to work picking only apples for one whole day, you get 700 apples and zero oranges. If you move one worker to oranges (so four workers are picking apples and one worker is picking oranges), you get 680 apples picked and 300 oranges picked. Because of diminishing returns, taking one worker away from apples reduces apple output by only 20. But moving that worker to oranges increases orange production by 300 because that worker is the first one picking oranges and can get the low-hanging fruit.

TABLE 3-2

Outputs of Apples and Oranges as the Labor Allocation Changes

	Combo 1	Combo 2	Combo 3	Combo 4	Combo 5	Combo 6
Workers picking oranges	0	1	2	3	4	5
Workers picking apples	5	4	3	2	1	0
Output of oranges	0	300	500	620	680	700
Output of apples	700	680	620	500	300	0

Graphing the combinations

You can graph your production possibilities by plotting the various combinations of two output goods that can be produced as you vary the amount of a resource allocated between them. Figure 3-1 plots the six output combinations that result from varying the allocation of workers in Table 3-2, thereby graphing all your production possibilities. Point A corresponds to putting all five of your workers to work picking apples. Point B corresponds to the output you get from four workers picking apples and one worker picking oranges, and so on.

Note that each of the six points is attainable in the sense that you can produce the corresponding quantities of each fruit through some allocation of the five workers' labor. On the other hand, a point like C isn't attainable. You can't allocate your five workers in any way to produce that many apples and oranges. You could produce such an output combination if you had more workers, but you're limited to five workers.

Imagine that instead of allocating labor by worker, you allocate it by time. The five workers each work for one day, so you have 5 worker-days of labor to allocate. You can now allocate, for instance, 3.2 worker-days to apple picking and 1.8 worker-days to orange picking. This arrangement allows you to fill in the graph and draw a line connecting the six points corresponding to the output combinations you get when allocating labor by worker, one worker at a time.



REMEMBER

The line passing through the six points is called the production possibilities frontier, or PPF, because it divides the area of the graph into two parts. The combinations of output that are possible to produce given your limited supply of labor are under the line in the shaded area, and those that are not possible to produce are above it in the unshaded area. In this way, the PPF graph captures the effect of scarce resources on production. Some output combinations just aren't producible, given the limited supply of labor.

The PPF is a simplification of the real world, derived by allocating one input between just two outputs. The real world is more complicated, of course, with

several resources allocated among the different outputs. But the principles of limited resources and diminishing returns that show up so clearly on the PPF graph also apply to the much greater variety of inputs and outputs in the real world.

Interpreting the shape of the graph

The bowed-out curvature of the PPF graph illustrates the effects of diminishing returns. In Figure 3-1, the changing slope as you move along the frontier shows that the trade-off between apple production and orange production depends on where you start. If you're at Point A, where you're allocating all your resources to the production of apples, then you can, by reallocating resources, produce a lot more oranges at the cost of giving up only a few apples. But if you start at Point D, where you're already producing a lot of oranges, you have to give up a lot of apples to get just a few more oranges.

In economic jargon, the changing slope of the PPF in the face of diminishing returns is due to the fact that the opportunity cost of production varies, depending on your current allocation of resources. (Check out Chapter 2 for more on opportunity cost, which equals the value of the best option you didn't choose.) Suppose you're already producing a lot of apples. In that case, the opportunity cost of devoting even more labor to apple production is high because you're giving up a lot of potential orange production. On the other hand, the opportunity cost of devoting that labor to orange production is low because you have to give up producing only a few apples. Clearly, you should devote the labor to picking the fruit with lower opportunity cost because, in this example, both fruits bring the same benefit: \$1 per fruit sold.

Gauging efficiency



TIP

The PPF is handy because any points that lie on the PPF itself (on the frontier) clearly show the output combinations you get when you're productively efficient, or wasting none of your resources.

Starting from any point on the PPF, you can't increase the production of apples without reducing the production of oranges, and vice versa. For instance, if you start at Point B, the only way to increase apple production is to slide up along the frontier, which implies reducing orange production. You have to make this trade-off because you don't have any wasted labor lying around with which you could get more apples without reducing the amount of labor already devoted to orange picking.

All the points below the line are productively inefficient. Consider Point E in Figure 3-1, which corresponds to producing 300 apples and 300 oranges. You produce at a point like E only if you're productively inefficient. In fact, you can see

from Table 3-2 that you can produce these numbers by sending only one worker to pick apples and another worker to pick oranges. You're using just two of your five workers; the labor of the other three workers is either being wasted or not used.

In the real world, you end up at points like *E* because of inefficient production technology or poor management. For one reason or another, the available resources aren't being used to produce as much output as they could. Any manager who has five workers to allocate but produces only output combination *E* should be fired!



REMEMBER

Economies that want to be efficient should always produce at some point on their production possibility frontiers. Otherwise, they're wasting precious resources.

Reaching new frontiers with better technology

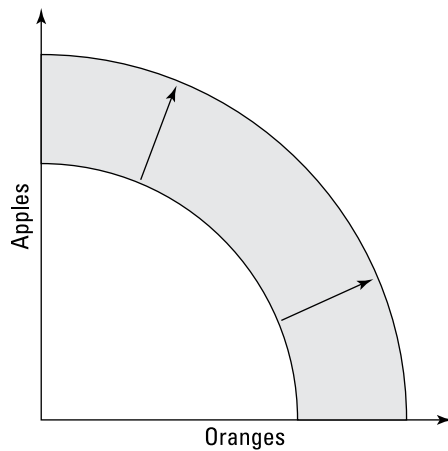
One simplification that's assumed when you draw the PPF is that, aside from the particular input you're allocating, you implicitly hold constant all other productive inputs, including technology. But humanity's technological sophistication is constantly increasing, allowing people to produce more from any given set of resources.

Economists represent this increase in productivity by shifting the PPF outward. In Figure 3-2, the shaded area represents new combinations of output that can be produced using the same amount of resources as before, thanks to better technology. The PPF is still curved because better technologies don't eliminate diminishing returns. Even with a better technology, if you start increasing the amount of a particular input, you get successively smaller additional increases in output.

In Figure 3-2, the technology shift is "balanced" in the sense that it increases your ability to produce more of both goods. An example of a balanced technological change would be improvements in fertilizers or pesticides that increase crop yields for both apples and oranges.

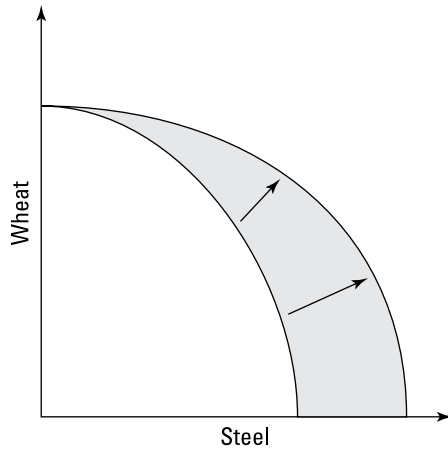
Most technological innovations are biased. For instance, suppose you're considering a PPF where the two output goods are wheat and steel. An improvement in steel-making technology obviously allows you to make more steel from your limited resources but doesn't affect your ability to produce wheat. In this situation, the PPF wouldn't shift out evenly. This is illustrated in Figure 3-3, where the PPF shifts out at the end where all your particular input (say, labor) is devoted to steel but remains fixed at the end where all your particular input is dedicated to wheat production.

FIGURE 3-2:
A technologically
balanced outward
shift of the PPF.



© John Wiley & Sons, Inc.

FIGURE 3-3:
A technologically
biased outward
shift of the PPF.



© John Wiley & Sons, Inc.

Deciding What to Produce

After a society locates the frontier of efficient output combinations (see the preceding sections), the next step is choosing the point along the frontier that produces the mix of goods and services that makes people happiest. Choosing only from among frontier combinations guarantees productive efficiency. Selecting the single frontier combination that maximizes happiness ensures allocative efficiency.

Because determining where the frontier lies is mostly a matter of engineering and applying current technology to available resources, it creates little

controversy — but deciding which particular combination of outputs a society as a whole should choose is much more complicated. People have preferences both as individuals and as groups about which products make them happiest. An individual selecting a point along his own personal PPF encounters no conflict; he just determines what combination of outputs makes him happiest, and then he produces and consumes it. The decision-making process becomes vastly more complicated when you consider an entire society's PPF, in which case you're sure to have a vigorous disagreement about which combination of outputs to produce with the society's limited resources.

For instance, your neighbor may not mind all the pollution because she likes driving her gas-powered SUV day and night. If she were living in her own world, the pollution wouldn't matter, but because you live near her, you're affected by the pollution and object. Perhaps you'll seek government intervention to limit your neighbor's actions. Similarly, the government argues over what it should produce with its limited resources. Some people favor farm subsidies, whereas others favor defense spending or programs to aid the poor.

Because of competing priorities, some sort of decision-making process must be established to determine what will actually be produced and to attempt to please most people most of the time. In most modern economies, this process results from private and public decisions acting through a combination of free markets and government action. The process isn't always smooth, but it has delivered the highest standards of living in world history.

In this section, I discuss the pros and cons of free markets and government interventions and why most nations have opted for some mixture of the two rather than relying exclusively on one or the other.

Comparing market results and government interventions

When analyzing how modern economies and societies select a combination of goods and services to produce, realize that current laws and institutions are the result of conflicting pressures about either:

- » Leaving markets to their own devices when turning resources into output.
- » Using the power of government to intervene in markets to secure a different set of outcomes.

Keep the following three factors in mind when considering the political and social conflicts that arise with regard to leaving markets alone or intervening:

- » **Complexity:** Modern economies are hugely complicated, with millions of goods and services being produced from limited supplies of land, labor, capital, and entrepreneurial ability. Markets handle this complexity easily, but government interventions usually don't — meaning that government interventions often risk substantial reductions in productive and allocative efficiency.
- » **Side effects:** Some goods and services, such as cocaine and coal-burning power plants, have negative consequences that often fall on “third parties” who weren't directly involved in either the production or the consumption of those products. These negative consequences bring forth substantial pressure for government intervention in the economy because markets, if left alone, will produce a lot of these goods and services and, consequently, a lot of negative consequences for third parties.
- » **Inequality:** Some people end up consuming a considerable proportion of the goods and services produced, whereas others end up with very little. Such an unequal distribution also generates a great deal of pressure for government intervention in the economy to equalize living standards.

These factors are both a consequence and a cause of the fact that modern economies are largely a mix of market production and government intervention. For the most part, what to produce, how much to produce, and who gets it are decided by voluntary transactions made by individuals and businesses. But sometimes the government uses its powers to achieve outcomes that wouldn't happen if individuals and businesses were left to their own devices.

In both cases, a gigantic apparatus of law and tradition helps society produce a combination of output that is, hopefully, both productively efficient (so that every unit of output gets made from the smallest amount of inputs possible) and allocatively efficient (so that the correct mix of goods and services is produced). Next, I outline the benefits and drawbacks that markets and governments bring to the economic table.

The magic of markets: Directing resources automatically

Market economies are collections of billions of small, face-to-face transactions between buyers and sellers. Economists use the term *market production* to capture what happens when one individual offers to make or sell something to another individual at a price agreeable to both.



REMEMBER

In markets, the allocation of resources is facilitated by the fact that each resource has a price, and whoever is willing to pay the price gets the resource. In fact, market economies are often called *price systems* because prices serve as the signals that direct resources. Holding supply constant, products in high demand have high prices, and products in low demand have low prices. Because businesses like to make money, they follow the price signals and produce more of what has a high price and less of what has a low price. In this way, markets tend to take limited resources and use them to produce what people most want — or at least, what people are most willing to pay for.

For instance, the guy who sells you a TV at a local retailer has no idea about the total demand for TVs in the world, how many tons of steel or plastic are needed to produce them, or how many other things weren't produced because the steel and plastic needed to make the TVs were used for TVs rather than other things. All he knows is that you're willing to pay him for a TV. And if he's making a profit selling TVs, he orders more TVs from the factory. The factory, in turn, increases production, taking resources away from the production of other things.

Markets have the benefit of figuring out, automatically, the things people want. To grasp why this is so amazing, consider that you live in a world of more than 8 billion people. It would be impossible for any single person to gather enough information to figure out what each of them most wants to buy. It would take several lifetimes to speak with each of them, even just to find out what they want for dinner, let alone all the other things they would most like to purchase on a typical day. But because production and distribution in modern economies aren't centrally controlled (by a small number of government officials working together in a single place), nobody needs to know the big picture.

INDIA'S LICENSE RAJ

Raj is a Hindi word that means rule or reign. The License Raj was a system of centralized economic planning and regulation in India that lasted from independence in 1947 until 1991. Under this system, businesses needed to obtain licenses and permits from the government to start or expand their operations, and these licenses were often difficult to obtain due to corruption and bureaucratic red tape.

The License Raj created an environment in which government officials had extensive power over businesses, which often led to cronyism and corruption in addition to all the difficulties associated with central planners not having enough information to make good decisions about the allocation of resources. In addition, the regulations and restrictions made it difficult for businesses to grow and be competitive, which resulted in a lack of innovation and efficiency.

(continued)

(continued)

One of the key consequences of the License Raj was the stunting of economic growth in India. The heavy regulation and bureaucracy discouraged investment and entrepreneurship, and the country's economic growth rate was consistently lower than that of other countries during the same period. The ending of the License Raj in 1991 ushered in a new era of economic liberalization and privatization in India, which helped to more than double India's growth rate while unleashing a massive wave of high-tech innovation.

In a competitive market, many sellers compete against each other to attract customers. Here are some of the advantages of competition:

- » **Using resources efficiently and keeping prices low:** A competitive market tends to guarantee productive efficiency because the best way for sellers to keep prices low is to make sure they're using all their resources efficiently, with nothing going to waste. Because competition is ongoing, the pressure to be productively efficient is constant.
- » **Underselling their rivals and stealing their customers:** Because every firm has this incentive, prices tend to be driven so low that the businesses can just barely make a profit.
- » **Encouraging improvements in technology:** Market production with a lot of competition tends to ensure not only that economies produce along the frontier but also that they have frontiers that are constantly being pushed outward. That constant outward shift results from constant improvements in productive efficiency generated by *entrepreneurs*, who supply the entrepreneurial ability to generate innovation and invention.

The misdeeds of markets: Ignoring morals and fairness

Markets aren't perfect. In particular, they suffer from two significant problems:

- » Markets produce whatever people are willing to pay for, even if these things aren't good for the people or the environment. As long as profits are to be made, you can be confident that a supply will arise to satisfy any demand. The fact that illegal drugs are widely and cheaply available — despite vigorous government programs to stop their production and distribution — is probably the best example of the robustness of markets. But although it's nice that markets are so hell-bent on giving people what they're willing to pay for, illegal drugs are an excellent example of the fact that markets will deliver things

without caring about their negative consequences, such as drug addiction, broken families, and increased crime.

Along the same lines, producers sometimes utilize production methods that consumers don't like. Child labor and sweatshop labor are primary examples. Often the government must intervene to change these practices when the price system fails to incentivize producers to reform such objectionable practices.

- » Markets don't in any way guarantee fairness or equity. Markets cater to those with money to spend. If someone is poor, they can't incentivize producers to provide them with even basic necessities, such as medicine and food.

A related problem with markets is income and wealth inequality. Because market systems reward those who can best provide the goods and services that people want to buy, some sellers become rich because they're better at delivering what people want. This invariably leads to significant inequalities in wealth that many people find offensive.

The case for government intervention

Many societies use their governments to intervene and address the problems that markets create or can't fix. Government interventions in the economy usually take one of three forms:

- » **Penalties or bans on producing or consuming goods or services considered dangerous or immoral:** For example, governments may ban drugs or impose "sin taxes" on alcohol and tobacco, which, though legal, are considered products whose use should be discouraged. However, these bans often work only partially because the market still has a significant financial incentive to provide such products.
- » **Subsidies to encourage the production of desirable goods and services:** For instance, most governments heavily subsidize children's education and medical care. They do so because of the fear that, without subsidies, insufficient education and inadequate medical care will be provided. Education and medical care also offer extensive spill-over benefits to the broader public in the form of higher productivity, less crime, and fewer disease outbreaks.
- » **Taxes on the well-off to provide goods and services to the less fortunate and to reduce inequalities in income and wealth:** These taxes are put toward things like good parks, clean air, and art, as well as goods and services for the poor. Governments tax individuals and businesses to raise the money to provide such things.

In terms of the PPF graph, each of these government interventions causes the economy to produce and allocate an output combination, or “output bundle,” that’s different from the one society would have ended up with if the markets had made all the production and allocation decisions.

Politicians and political activists tend to promote only those interventions for which a persuasive case can be made for improvements in either productive efficiency, allocative efficiency, or fairness. Whether any particular government-influenced output bundle will in fact end up being better or worse than what the market would have produced will depend on the specific situation, however.

The case against government intervention

Government intervention is a powerful force for redirecting economic activity, but it doesn’t necessarily improve the economy. There are at least three good reasons to worry that government intervention will make things worse:

- » **Special interests:** Government programs are often the result of special interest lobbying that persuades the legislature to put the interests of a small group ahead of the interests of the general public.
- » **Inefficiency:** Even when pursuing the common good, government programs often deliver poor service because they have no competition to create incentives to efficiently produce government goods and services.
- » **Inflexibility:** Government interventions usually lack the flexibility of the price system, which is able to constantly redirect resources to accommodate people’s changing willingness to pay for one good rather than another. Government policies take years to pass, and laws are usually written in a precise manner that doesn’t allow for changing circumstances and rapid innovation — things that the price system handles easily.

THE SUPER-SWEET SUGAR SUBSIDY

The U.S. federal government subsidizes domestic sugar growers through price supports, import quotas, and loan programs. This subsidy program has been widely criticized for driving up the price of sugar for consumers and for causing a number of negative economic, environmental, and health effects. For example, the sugar subsidy has led to the destruction of natural habitats in the Florida Everglades and has been linked to higher rates of obesity, diabetes, and other health problems due to the fact that the artificially high price of sugar generated by the sugar subsidy has caused many food companies to substitute high-fructose corn syrup for sugar in processed foods.

So how does a government intervention like the sugar subsidy survive if it appears to clearly *reduce* productive and allocative efficiency? Because U.S. sugar producers lobby the U.S. government hard each year to renew the subsidy. For each producer, the subsidy is worth a lot of money. And because only a few cents out of the average voter's tax bill go toward the sugar subsidy, no one protests it. Consequently, the sugar subsidy survives not because it does society any good but because lobbying pays off in a democracy. Many other government programs are similarly deficient of widespread social benefits.

Although markets sometimes fail to deliver everything society wants, government intervention isn't a cure-all. Markets are good at providing the majority of things people want and can usually do so at the lowest possible cost. Consequently, government intervention should be well thought out lest it make things worse rather than better.

Opting for a mixed economy

In the real world, few societies opt for an extreme type of economy, such as one that's totally market-based or one that features constant and pervasive government intervention. Instead, most societies opt for some mixture of markets, government intervention, and what economists call traditional production. In their purest forms, these three types of economy can be defined as follows:

- » **Market:** A market economy is one in which almost all economic activity happens in markets with little or no interference by the government. Because of the lack of government intervention, this system is also often referred to as *laissez faire*, which is French for "let it be," or, "leave it alone."
- » **Command:** A command economy is one in which the government directs all economic activity.
- » **Traditional:** A traditional economy is one in which production and distribution are handled along the lines of longstanding cultural traditions. For instance, until the caste system was abolished in India, nearly every good and service could be produced only by someone born into the appropriate *caste*, or social strata. Similarly, in medieval Europe, you couldn't typically be part of the government or attain high military rank unless you were born a noble.

Because nearly every modern economy is a mixture of these three pure forms, most modern economies fall into the inclusive category called mixed economies. Except for a few isolated traditional societies, however, the traditional economy part of the mixture has tended to decline in significance because most production

has shifted to markets and because traditional economic restrictions concerning personal characteristics like age and gender have become less influential (and more illegal).



REMEMBER

Today's mixed economies are a blend of command and market economies. The ones you find in most countries typically feature governments that mostly allow markets to determine what's produced, but that also fuse in limited interventions in an attempt to improve what the market would do if left to its own devices.

The precise nature of the mixture depends on the country, with the United States and the United Kingdom emphasizing markets and France and Germany, for instance, featuring more emphasis on government intervention. On the other hand, a few totalitarian states like North Korea persist in running pure command economies as part of their all-encompassing authoritarian regimes.

Noting the failure of command and the absence of *laissez faire*

Command economies have historically been dismal failures. (See the sidebar "India's License Raj" for a relatively mild example.) Even well-intentioned governments can't gather enough information about production and distribution to allocate resources well. In fact, they do a much worse job than price systems do.

Consequently, the opposite extreme — no government intervention — can seem attractive. French economists first suggested such *laissez-faire* systems several hundred years ago in response to the propensity of governments of that era to intervene heavily in economic activity. However, no pure *laissez-faire* economy has ever existed or probably could ever exist. The simple fact is that properly functioning market economies that use price mechanisms to allocate resources require tremendous government support.



REMEMBER

Among other things, market economies need governments to:

- » Enforce property rights, so people don't steal
- » Provide legal systems to write and enforce contracts so people can make purchases and sales of property and resources as well as goods and services
- » Enforce standardized systems of weights and measures so people know they aren't being cheated
- » Provide a stable money supply that's safe from counterfeiters
- » Enforce patents and copyrights to encourage innovation and creativity

Notice that all these things must be in place for markets to function. Consequently, a more moderate, more modern version of *laissez-faire* says that governments should provide the institutional framework necessary for market economies to operate. It should then get out of the way and let people make and sell whatever is demanded under that institutional framework.

Deciding on the amount of government intervention

Because command economies don't work very well and *laissez-faire* economies can't really exist, most societies have opted for one form or another of mixed economy in which governments and markets share economic responsibilities. The precise nature of that mix varies from country to country. All such mixtures feature some instances of direct governmental command and control of economic activity interacting with markets that use a price system to allocate resources.

But most people want governments to do more than just set up the institutions necessary for markets to function. They want governments to stop the production and sale of things like drugs or to subsidize the production of things that the market economy may not provide much of, like housing for the poor. They often also want to tax well-off citizens to pay for government programs.

Many government programs are so commonplace that you don't even consider them government interventions. For instance, free public schools, safety features on cars, warning labels on medicine bottles, taxes on alcohol and tobacco, and mandatory contributions to retirement systems are all government interventions in the economy.

The government interventions needed to implement such programs are, in many cases, inefficient. But many people would argue that there's quite a bit more to life than efficiency and that the inefficiencies caused by many government interventions are well worth the benefits they produce. For such people, the government interventions in question increase overall happiness even though they are, strictly speaking, inefficient.



REMEMBER

At the end of the day, all government interventions — both good and bad — are the result of a political process. In democracies, the amount of government intervention is, broadly speaking, a reflection of the will of the people. Nations with more trust in markets, like the United States and the United Kingdom, tend to feature mixed economies with less government intervention than nations that are more suspicious of corporations and impersonal market forces, like France and Germany.

Promoting Technology and Innovation

Technology is, in many ways, like any other good that a market can provide. If there's a profit incentive to invent a new technology, businesspeople will figure out a way to create it, just as they figure out ways to deliver all the other things people are willing to pay for. However, firms won't find it profitable to innovate unless the government cultivates the right set of laws, institutions, and property rights.

Here are several ways in which governments have promoted technology and innovation.



REMEMBER

- » **Funding research:** Governments supply a good deal of direct support through research grants and university subsidies.
- » **Protecting intellectual property rights:** A crucial thing to understand about innovation is the indirect role that governments play not by subsidizing new technology but by guarding it. In particular, the patents granted by governments provide a substantial economic incentive for individuals and businesses to innovate. A patent guarantees inventors of new products or business methods the exclusive right to profit from their innovations, usually for about 20 years in most countries.

It's not a coincidence that economic growth in the United States and Western Europe took off 200 years ago, right after patents became widely enforced. For the first time in world history, there was a secure financial incentive to use your brain to innovate. Before that time, innovating was extremely risky because, after all your hard work, others would simply copy your invention and sell it without your permission.

Copyrights for literary, musical, and cinematic works serve a similar purpose. Artists produce more art when they know they can make a living creating it. Along these lines, the easy duplication and distribution of digital media on the internet is troubling because it has weakened artists' ability to charge for the art they work so hard to produce.

- » **Encouraging education:** You shouldn't be surprised that every wealthy country in the world has a policy of universal primary and secondary education as well as strong universities. Innovative new technologies require well-educated researchers from sound educational systems.

Advanced economies also require competent, well-educated workers to implement the new technologies. Consequently, education must be available to everyone if an economy is to utilize the constant flow of innovative new processes and tools that researchers develop.

POLITICAL SYSTEMS VERSUS ECONOMIC SYSTEMS

The three categories economists use to distinguish economic systems — market, command, and traditional — are only meant to convey information about a nation's economic system. By contrast, terms like *communism*, *capitalism*, *democracy*, and *socialism* tend to focus either on a nation's political system or on how its political system is interwoven with its economic system.

Consider the word *democracy*. It refers to a political system in which the citizens control the nation's government. Whether they direct the government to institute an economic system that is market, command, or traditional isn't specified by the fact that the country is run democratically. The fact that a country is a democracy also says nothing about whether the government favors private ownership or government control of businesses.

In contrast, communist and socialist systems are supposed to feature democratic governments owning and operating all the businesses in an economy on behalf of the people. That is, both systems are, by definition, supposed to encompass both a nation's political system as well as its economic system.

Please note, however, that actual communist and socialist governments have tended to fail in achieving their democratic political aspirations. Instead, communist and socialist nations like the old Soviet Union, North Korea, Cuba, and, recently, Venezuela, tend to feature nondemocratic totalitarian governments that use their almost total control over commerce as yet another tool for suppressing their citizens.

Finally, understand that the term *capitalism* has many conflicting definitions. For some, it refers to economies featuring private property and free markets. To others, it's defined by industrialization and the use of large amounts of capital equipment in production. Yet another group uses it to describe a political-economic system in which the wealthy owners of large businesses use their power and influence over commerce and the government to exploit the poor. Because of these many conflicting (and controversial!) definitions, economists tend to avoid the term *capitalism* in favor of a more precise descriptive term like *market-based economy*.



Microeconomics: The Science of Consumer and Firm Behavior

IN THIS PART . . .

Discover the relationships among prices, quantity supplied, and quantity demanded and how changes in supply or demand affect market equilibrium.

Understand how people maximize their happiness by navigating marginal utility, weighing alternatives, and choosing just the right amounts given their limited budgets.

Know how and why firms set output levels to maximize profits and what they do when they lose money.

Check out the social benefits of markets, how they can maximize total surplus, and what taxes and price controls do in a market setting.

Find out what monopolies are and how they behave.

Examine what collusion is and why some collusive oligopolies work while others don't.

IN THIS CHAPTER

- » Understanding the relationship between price and the quantity supplied or demanded
- » Focusing on market equilibrium
- » Understanding how shifts in demand or supply affect market equilibrium
- » Identifying policies that prevent market equilibrium

Chapter 4

Supply and Demand Made Easy

In the modern economy, most economic activity occurs in markets, where buyers and sellers come together to trade money for a good or service. A market doesn't have to be an actual place, and many markets nowadays exist only in cyberspace. But no matter what sort of institutional arrangement markets have, they all tend to behave in the same way.



REMEMBER

A straightforward model called *supply and demand* does an excellent job of describing how markets work, regardless of which good or service is being bought and sold. This model separates buyers from sellers and then summarizes each group's behavior with its own curve on a graph. The demand curve captures the buyers' behavior, and the supply curve captures the sellers' behavior. By putting these two curves on the same graph, economists can show how buyers and sellers interact to determine how much of any particular item will be sold and for how much it'll be sold.

The supply-and-demand model is the economics profession's most famous contribution to human understanding. It's helpful in so many areas, shedding light on exactly how markets set prices and allocate resources, as well as giving

accurate predictions about how government policies will affect the behavior of markets. For instance, this model can tell you why gas prices go up during the summer and wheat prices go down after a good harvest. It also explains why chocolates put on store shelves in anticipation of Valentine's Day are marked down significantly a week or two afterward. (But, take it from me, it's best to give your sweetie the chocolates on February 14, not February 23 when they've been marked down to half price. I've learned a few things over the years!) The supply-and-demand model can also predict — correctly — that agricultural price supports will cause an overproduction of food and that rent control will lead to a housing shortage.

I begin this chapter by showing you how to draw and manipulate supply and demand curves. The next step is to watch the curves interact to see how markets function both when left to their own devices and when subject to government regulation or intervention. After reading this chapter, you should have new insights into virtually everything you read or hear about commerce, business, and politics.

Deconstructing Demand

People want to buy things, and economists refer to that desire as *demand*. When they say *demand*, economists aren't referring to pie-in-the-sky dreams or to mere wishful thinking along the lines of, "I want a billion-jillion scoops of ice cream!" Rather, when they say *demand*, economists mean how much of something people are both *willing and able* to pay for. So although I may want a billion-jillion scoops of ice cream (butter pecan, please!), that's not my demand in the sense that economists use it. Instead, my demand is three scoops because that's how much I'm *willing and able* to buy at the price my local ice cream shop charges for it.

To be precise, what I've just described is my *quantity demanded*, which refers to how much I demand at a specific price, holding constant everything else in the world, including my income, my preferences, the prices of other goods and services, and the nagging voice of my doctor who doesn't want me to eat three scoops of ice cream in one sitting. By contrast, when economists use the word *demand*, they mean the whole range of quantities a person will demand at various prices while holding constant all other possible influencing factors.

If that sounds confusing, don't worry. I'll make it very clear in this section how "quantity demanded" and "demand" differ from each other. We'll also look at how a product's own price determines a demand curve's slope and how changes

in any other factor — including individual preferences and the prices of competing products — can shift the position of the demand curve.

Prices and other stuff: Looking at what affects quantity demanded

Economists divide everything that can possibly affect the quantity demanded into two groups: a product's (own) price and *everything* else. Here's what "everything else" includes:

- » **Tastes and preferences:** Chief among "everything else" are tastes and preferences. For instance, no matter how low the price gets, I won't buy even a single container of Cherry Dream ice cream because I think that flavor tastes awful. At the same time, however, lots of people love Cherry Dream ice cream so much that even if the price increased by 50 percent, they'd still be willing to buy quite a bit of it. No matter how much a container of a particular flavor costs, the people who love it will always have a higher quantity demanded than I will. Because this is true for every possible price, you can say that they have a *higher demand* than I do.
- » **Other prices:** People's demand for Product X is affected not only by its price but also by the prices of everything else they can possibly buy. If, for example, the price of movie tickets rises, people are likely to not only go to the movies less often but also to spend the money they save at the theater to use on other entertainment activities, such as concerts. Thus, a change in the price of movie tickets will affect their quantity demanded of concert tickets — even when the price of concert tickets hasn't changed.
- » **Income:** As you get richer, you increase your quantity demanded of specific goods that you've always liked and can now afford to buy more of. These are called *normal goods*. On the other hand, you decrease your quantity demanded of those things that you had been buying only because you were too poor when you bought them to get what you really wanted. These are called *inferior goods*. For example, new cars are normal goods, while clunkers are inferior goods. Similarly, freshly made organic salads are normal goods, and three-day-old, discounted bread is an inferior good.

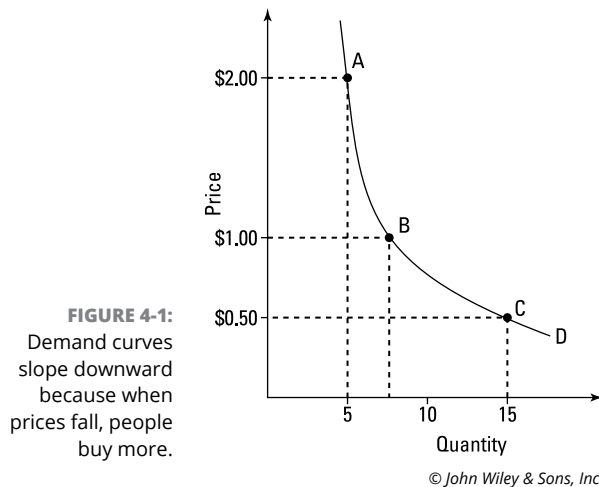
Given the complexity of variables such as preferences and income, why do economists insist on dividing everything that could influence your quantity demanded into only two groups: a have a very different effect than do the other variables. Any change in a product's (own) price will move you along a given demand curve, whereas a change in any other variable will shift the location of the curve and how it's shaped.

Graphing the demand curve

A product's (own) price has an *inverse relationship* with the product's quantity demanded. In other words, the higher the price, the fewer units people demand (while holding constant all the other things that can affect the quantity demanded).

Geometrically, that inverse relationship between a product's (own) price and its quantity demanded implies that the product's demand curve must slope downward.

I've drawn a demand curve in Figure 4-1. To break things down, let's suppose that this demand curve represents the demand for cabbages. On the vertical axis is the price of cabbages, measured in dollars. The horizontal axis is the number, or quantity, of cabbages that are demanded at any given price.



As you can see, the demand curve slopes downward. As noted, this downward slope reflects the fact that there's an inverse relationship between the price of cabbages and the number of cabbages people want to buy. For instance, consider Point A on the demand curve. At that point, people demand five cabbages at a price of \$2 per cabbage. However, as you can see by looking at Point B, if the price drops to \$1 per cabbage, people demand eight cabbages. And if the price drops to only \$0.50 per cabbage, they demand 15 cabbages.

Here's how changes in price and other factors are reflected in the demand curve:

- » Increases or decreases in (own) price simply move you along the current demand curve.

- » Changes in any other factor that affect people's willingness and ability to buy (such as changes in their incomes or changes in the price of a related product, like corned beef) shift the demand curve either left or right.

For instance, suppose that a government health study says that cabbages make people attractive to members of the opposite sex. Naturally, that's going to increase the demand for cabbages. Geometrically, the effect is to shift the demand curve to the right. I've illustrated this effect in Figure 4-2, where the demand curve before the study is announced is labeled D , and the demand curve after the study is announced is labeled D' .

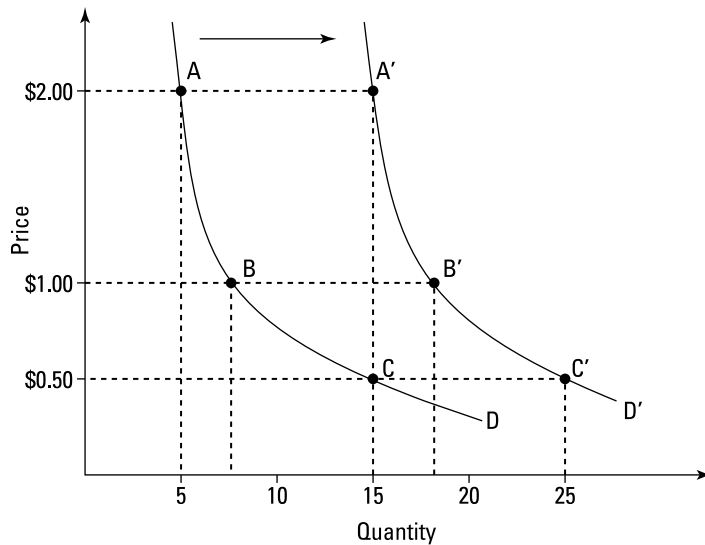


FIGURE 4-2:
An increase in demand causes the demand curve to shift right from D to D' .

© John Wiley & Sons, Inc.

Whenever a demand curve moves, economists say that there has been a *shift in demand*. In this case, demand has increased. In contrast, if the curve shifts to the left, you would say that demand decreased. Implicit in this way of describing the movements is the fact that the quantities demanded either increase or decrease *while holding prices constant*.



WARNING

You have to distinguish between changes in quantities demanded that occur because the price changes (these are movements along a given curve) and changes in quantities demanded that occur because something besides the price changes (these are shifts of the entire curve). Anything besides the price that affects the quantity demanded shifts the demand curve left or right.

To see the difference, compare Point A and Point A'. Both points share the same price of \$2 per cabbage, but thanks to the recently released government study, people now demand 15 cabbages at that price (at Point A') rather than 5 cabbages at that price (at Point A). Because the price is the same for the two points, you know that the change in the quantity demanded was caused by something other than (own) price. Similarly, you can look at what happens to the quantity demanded while holding the price of cabbages constant at \$1: The quantity demanded at that price increases from 8 before the study to 18 after, moving from Point B to Point B'.

Opportunity costs: Setting the slope of the demand curve

A product's own price is a measure of sacrifice, of the quantity of other goods and services that must be given up to obtain one additional unit of the product in question. The slopes of demand curves indicate how people react to changes in prices — and, consequently, to changes in how much of other goods and services must be sacrificed to obtain one unit of the product in question.

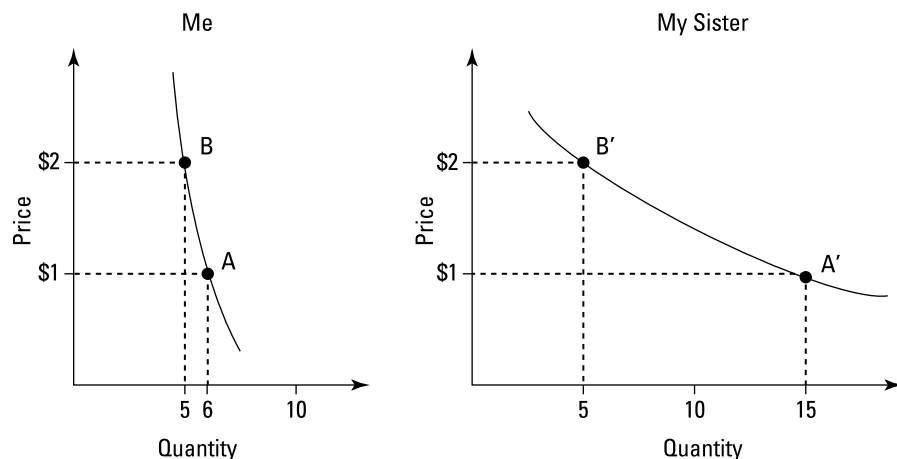
For instance, imagine that the price of a good you currently buy falls from \$10 to \$9. How do you respond? Well, that depends on how you feel about the good in question relative to other goods you could spend your money on:

- » You may buy a lot more of the good in question because extra units bring you a lot of happiness, and you're consequently grateful to be able to purchase them for \$9 instead of \$10.
- » You may barely increase your buying because although it's nice that you can now buy the good for \$9 instead of \$10, extra units just don't make you all that much happier. In such a situation, the best thing about the price cut is that it frees up money to buy more of something else.

In terms of demand curves, these different reactions lead to different slopes. The person who buys a lot more when the price falls has a flat demand curve, and the person whose purchases barely budge when the price falls has a steep demand curve.

Consider Figure 4-3, where I've drawn two separate demand curves on two separate graphs. The one on the left is my demand for pecans. The one on the right is my sister's demand for pecans.

FIGURE 4-3:
Two demand
curves for
pecans.



© John Wiley & Sons, Inc.

Notice that my demand curve is steep, whereas my sister's demand curve is pretty flat. The difference is entirely the result of differences in how we react to price changes. You can see this by comparing my quantity demanded at Point A with my quantity demanded at Point B. Even though the price doubles from \$1 per bag of pecans to \$2 per bag of pecans, my quantity demanded falls only from six bags to five bags. By contrast, when the price doubles from \$1 per bag to \$2 per bag, my sister's quantity demanded falls considerably, from 15 bags to only 5 bags.

This means that my sister is much less attached to pecans than I am. When I see the price double, it barely reduces my quantity demanded — I'm willing to give up a lot of other things that I could spent money on so I can keep buying almost as many pecans as before.

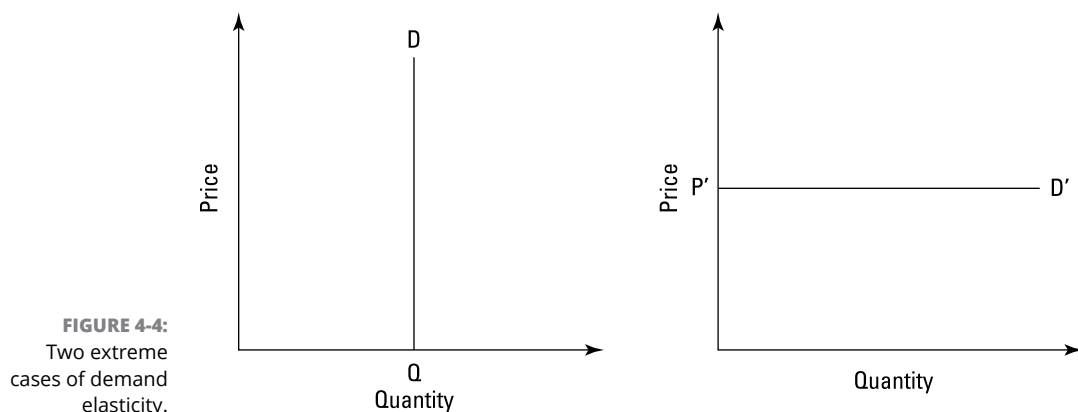
My sister, on the other hand, reacts very differently. Although she initially buys more pecans than I do when the price is only a dollar, when the price doubles, she cuts her pecan-buying by ten bags. When the price doubles, she decides that she'd be better off cutting back sharply on pecan purchases in order to be able to spend her money on other things. In plain English, she's not nearly as attached to pecans as I am. (Pecan pie, pecan peanut butter cookies, butter pecan ice cream . . . mmmmmmm!)

Elasticity: Looking at extreme demand cases

Economists have borrowed the word *elasticity* to describe how changes in one variable affect another variable. If they say *demand elasticity*, they're referring to how much the quantity demanded changes when the price changes. If the same change

in price causes my quantity demanded to fall much less than my sister's quantity demanded (as in Figure 4-3), then my demand curve has a lot less demand elasticity than my sister's.

Extreme cases of demand elasticity are illustrated in Figure 4-4 using two demand curves, the first being perfectly vertical and the second being perfectly horizontal.



Perfectly inelastic: Paying any price

The vertical demand curve, D in Figure 4-4, is said to be *perfectly inelastic* because exactly Q units are demanded no matter what the price is. You may be wondering what sort of a good would have such a demand curve, and the answer is life-saving drugs. If you need precisely Q units to keep living, you'll pay any price up to your maximum ability to pay. Ransoms in kidnappings are also probably like this because people will pay any price they can to get their family members back. Also, those battling drug addictions probably feel this way about their drug of choice; they're so desperate to get high that they don't care about the price.

Perfectly elastic: Buying all or nothing

The horizontal demand curve, D' in Figure 4-4, is said to be *perfectly elastic*. To understand this name, try to imagine a gradually sloping demand curve that's almost horizontal. On such a shallowly sloped demand curve, even a small change in price causes a big change in the quantity demanded. Indeed, the flatter a demand curve becomes, the greater the change in the quantity demanded for any given price change. For instance, look at Figure 4-3 and compare how a \$1 change in the price of pecans causes a much bigger change in my sister's quantity demanded on her flatter demand curve than it does on my steeper demand curve.

You can think of a perfectly horizontal demand curve as being the most extreme case of this phenomenon so that even the tiniest change in price brings forth an infinite change in quantity demanded. That is, if prices are above P' in the right-hand graph in Figure 4-4, you buy nothing, and if prices are at P' or just a penny less, you buy a whole lot. (*Infinite* is a whole lot.)

A concrete example of such a situation would be if you worked for a large restaurant chain and had to buy tons of ketchup. Your choices are Brand X and Brand Y, but because they taste the same, the only thing that matters is the price. Consequently, if the price of Brand X is even the slightest bit lower than that of Brand Y, you'll buy tons of Brand X and none of Brand Y. If the price of X is even slightly higher than that of Brand Y, you'll buy tons of Y and none of X.

Just know that demand curves that are perfectly elastic or perfectly inelastic are rare. Nearly all demand curves slope downward, meaning that moderate changes in prices bring forth moderate changes in quantities demanded. In Chapter 5, I explain why this is so by looking at how consumers make trade-offs between different goods to maximize the happiness they can get from spending their limited budgets.

Sorting Out Supply

When an economist thinks about the *supply* of goods and services, the key consideration is cost. Supplying things is costly, which means that you have to pay people to supply the things you want. Even more interesting, though, is the fact that the more you want them to supply, the higher their costs of supplying each additional unit. In other words, the first units tend to be relatively inexpensive to produce, whereas later units become successively more costly to produce. (In Chapter 6, I explain why this holds true.) If you want producers to make more and more, you have to pay them more and more.

In this section, I explain how rising costs cause supply curves to slope upward and how entire supply curves can shift when input costs change.

Graphing the supply curve

A *supply curve* shows the minimum prices at which someone is willing to sell various amounts of a good or service. Because production costs rise as people make more of something, suppliers insist on more money for larger quantities. That's why supply curves slope upward.

Imagine a farmer named Babbage who likes to grow cabbage. In Figure 4-5, I graph Mr. Babbage's supply of cabbages and label it S . (I was tempted to label it B^S , for *Babbage's Supply*, but I didn't want you telling your friends that my book was full of B^S !)

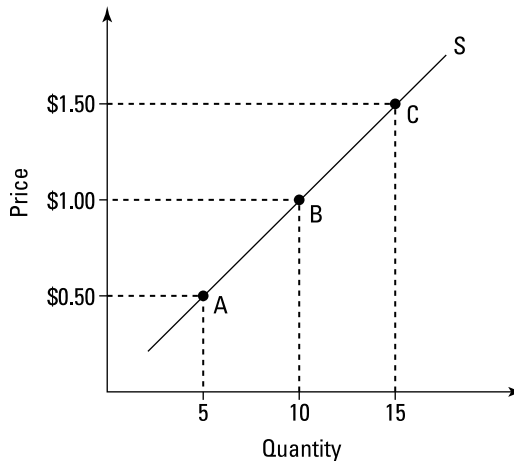


FIGURE 4-5: Supply curves have an upward slope because of increasing production costs.

© John Wiley & Sons, Inc.

The horizontal axis gives the number of cabbages supplied, and the vertical axis gives the price per cabbage that you have to pay to get Mr. Babbage to supply you with any given number of cabbages. Thus, Point A says that you must pay Mr. Babbage 50 cents per cabbage if you want him to supply you with five cabbages.

Because Mr. Babbage's production costs rise as he tries to grow more and more cabbages, you have to pay him \$1 per cabbage if you want him to grow you 10 cabbages, as shown by Point B. And you have to pay \$1.50 per cabbage if you want 15 cabbages, as demonstrated by Point C.



REMEMBER

Keep in mind that the points on the supply curve don't represent the prices that Mr. Babbage *wants* to receive for any given amount of cabbages — obviously, he wants to receive a gazillion dollars for each one. Instead, each dollar amount on a supply curve represents the minimum you could pay him and still get him to produce the desired amount. At Point A, you can get him to produce five cabbages if you pay him 50 cents per cabbage; if you offer him 49 cents, he won't do it. Why not? Because he has costs, which he can cover at 50 cents per cabbage but not at 49 cents per cabbage.

Separating sales price and production cost

Economists split all the things that can affect the quantity supplied into two groups: a product's own price and everything else. The things that go into *everything else* all relate to production costs — the costs of supplying the good in question.



TIP

When you see a particular supply curve, imagine that it derives from a particular production technology the supplier uses. Because each possible technology creates its own unique relationship between output levels and costs, some technologies give rise to steeply sloped supply curves, whereas others generate fairly flat supply curves. (See Chapter 6 for the skinny on firms' supply curves.)

Regardless of exactly how the curve is sloped or where it's positioned, the fact that costs increase as output increases means that you need to offer a higher and higher price to the supplier if you want to obtain more units. And that's basically why prices move you upward along supply curves.

A world-wide example of that phenomenon happened at the start of the Covid-19 pandemic. Rapidly increased production of medical supplies resulted in higher per-unit costs as medical-device makers struggled to hire more workers, run factories 24-hours per day, and obtain enough raw materials and inputs. The result was a strongly upward sloping supply curve for medical supplies during the crisis.

Price changes: Moving along the supply curve

Varying the price of an item moves you along a given supply curve because the supply curve represents the minimum payment you need to give the supplier to supply the amount of output you want.



REMEMBER

Suppliers look at whatever price is being offered and make as many units as are profitable at that price but no more. Because costs rise with each additional unit produced, the only way to get suppliers to produce more is to offer them higher prices. Therefore, raising or lowering prices moves you along the supply curve as the suppliers' quantities supplied respond to changing prices.

To see how this works, consider what happens if you offer to pay Mr. Babbage \$1 per cabbage and then let him choose how many cabbages he wants to produce. Given his supply curve in Figure 4-5, he'll want to produce precisely ten cabbages and no more. That's because for cabbage number one through cabbage number nine, the cost of production is less than what you're paying him. For example, consider Point A from Figure 4-5. At Point A, his production costs are 50 cents per cabbage. That means he'll make a nice profit if you pay him \$1 per cabbage. Similarly, because his cost per cabbage for producing six cabbages is also less than

\$1 per cabbage, he'll also want to make number six. The same is true of cabbages seven, eight, and nine.

Mr. Babbage is indifferent at ten cabbages because his cost per cabbage is \$1, and you're offering him \$1. In such cases, economists assume that he'll produce the tenth just to keep the buyer happy. But notice that Mr. Babbage would not produce at Point C if you offered him \$1 per cabbage. That's because his cost of production is \$1.50 per cabbage, and he would lose money.

Cost changes: Shifting the supply curve

Because a supplier's production costs determine where his supply curve is located and how it's sloped, changes in production costs cause changes in the supply curve. Things that make production more costly will shift the supply curve up, and things that lower costs will shift the supply curve down.

Suppose Mr. Babbage's production costs increase because the government imposes a new organic farming law requiring him to grow cabbages without using pesticides. In response, he has to hire lots of extra workers to kill pests with tweezers instead of simply spraying cheap chemicals.

Figure 4-6 shows the change in the supply curve. Because Mr. Babbage's production costs have increased, the minimum you have to pay him to produce any given level of output also goes up. Consequently, you can think of his supply curve as shifting upward vertically from S_0 to S_1 .

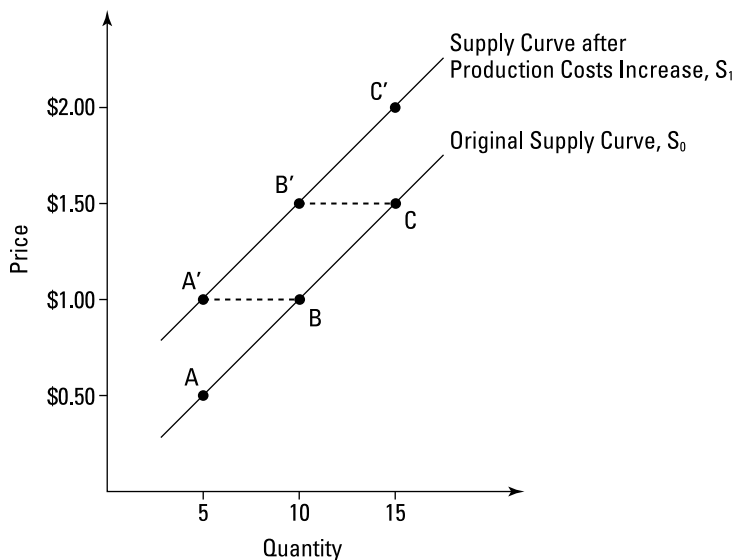


FIGURE 4-6:
Increased
production
costs shift the
supply curve.

© John Wiley & Sons, Inc.

I've drawn the shift to show that Mr. Babbage's cost of production is 50 cents higher for each cabbage, no matter how many cabbages are produced. Compare Points A and A'. Before the new environmental regulation, Mr. Babbage was willing to grow five cabbages if you paid him 50 cents per cabbage. After the policy change, you have to pay him \$1.00 per cabbage if you want him to grow you five cabbages.

Similarly, Points B and B' show that before the regulation, he would grow you ten cabbages if you offered him \$1 per cabbage. Now you have to offer him \$1.50 per cabbage if you want him to grow ten.



TIP

It's perfectly kosher to think of supply curves as moving left and right when cost structures shift. That is, you can say that the supply curve shifts left when costs increase, and you can quickly extrapolate that a decrease in costs would shift the supply curve to the right.

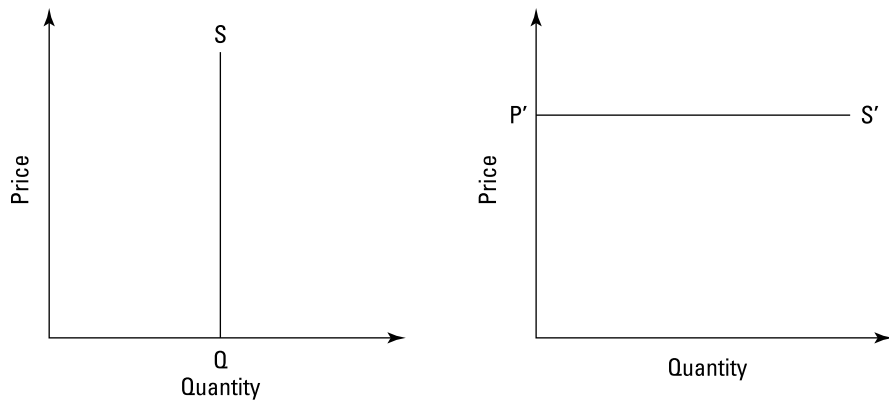
For instance, consider the quantity supplied at a price of \$1.00 both before and after the cost increase. Before the cost increase, Mr. Babbage is willing to supply you with 10 cabbages for \$1.00, putting you at Point B on the original supply curve. But after the cost increase, he's willing to supply you with only 5 cabbages for \$1.00, putting you at Point A' on the shifted supply curve. Similarly, at a price of \$1.50, Mr. Babbage was previously willing to supply you with 15 cabbages (Point C), whereas, after the cost increase, he's willing to supply only 10 cabbages at that price (Point B').

Having two ways to interpret supply curve shifts is rather handy. In some situations, it's easier to think of the shifts as either right or left, whereas in others, it's easier to think of them as up or down.

Using elasticity to understand extreme supply cases

Two extreme supply curves help to illustrate how production costs and prices combine to determine the quantity that will be supplied at any particular price. I've illustrated these two cases in Figure 4-7. The graph on the left shows a vertical supply curve and illustrates what economists call *perfectly inelastic supply*. The graph on the right with a horizontal supply curve illustrates what economists call *perfectly elastic supply*.

FIGURE 4-7:
Extreme supply
curves.



© John Wiley & Sons, Inc.

Being unable to make more: Perfectly inelastic supply

The left graph of Figure 4-7 illustrates a situation in which the price has no effect on the quantity supplied. As you can see in the diagram, no matter how low or how high the price, the quantity Q is supplied. Because the quantity supplied is utterly unresponsive to the price, it is said to be *perfectly inelastic*. Supply situations that look like this are usually referred to as situations of *perfectly inelastic supply*.

I expect you're curious about what has perfectly inelastic supply curves. The answer is unique things that can't be reproduced. Examples include:

- » **The Hope Diamond:** Because it's one of a kind — there'd still be only one Hope Diamond, no matter how much anyone wanted to pay — its supply curve is vertical.
- » **Land:** As comedian Will Rogers said back in the early 20th century, "Buy land. They ain't making more of it."
- » **The electromagnetic spectrum:** There's only one set of radio frequencies, and everyone has to share because there's no way to make more.

An interesting thing about each of these is that production costs are nonexistent. Because of this, offering the owner a price isn't an incentive in the way it is when you pay a producer enough to make something for you. Instead, the price serves only one purpose: to transfer the right of ownership and usage from one person to another.

Historically, the fact that the quantity of supplied land has nothing to do with production costs has been the justification for property taxes. The way

governments see it, they can tax land as harshly as they want because there's no need to worry that the amount of land — and, consequently, the tax base — will ever decrease.

Producing however much you want: Perfectly elastic supply

The right-hand graph in Figure 4-7 illustrates the case where the supply curve is perfectly horizontal. The idea here is that the supplier is producing something that has nonincreasing costs. No matter how many units you want the supplier to produce, it always costs only P' dollars to make a unit. Consequently, whether you want one unit produced or one jillion units produced, you only have to pay the supplier P' dollars per unit.

In the real world, the number of supply curves that are perfectly elastic is probably zero because production costs typically rise with output levels, as I explain in Chapter 6. But a few supply curves do come close. For instance, the supply curve for pencils looks nearly perfectly elastic because pencil companies seem to be able to increase production levels by millions of units with only minimal increases in costs.

Bringing Supply and Demand Together

Now it's time to bring supply and demand curves together so they can interact. First up is the market equilibrium point, where the supply and demand curves cross. In that discussion, I show you how markets determine the amounts, as well as the prices, of goods and services sold. Then I explain how markets find new equilibrium points when demand or supply curves shift. Are you ready to find out how?

Market equilibrium: Seeking a balance

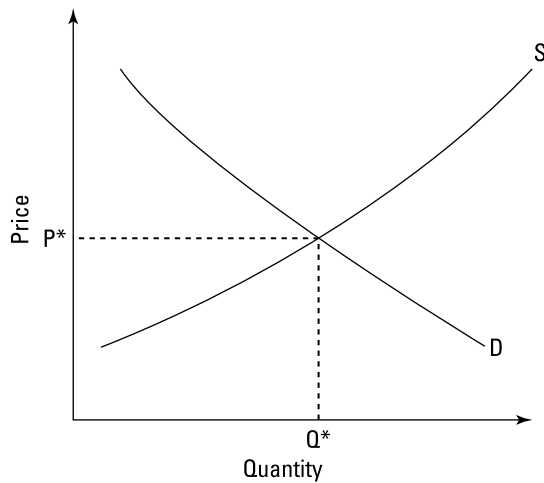
Supply and demand curves are especially useful when you graph them on the same axes. In Figure 4-8, I've drawn a demand curve and a supply curve and labeled them D and S , respectively.



REMEMBER

The point where the supply and demand curves cross marks how much the good or service in question costs and how much of it is sold. Just remember: X marks the spot! This price and quantity are known as the *market price* and the *market quantity*.

FIGURE 4-8:
The market equilibrium price and quantity happen where the demand curve crosses the supply curve.



© John Wiley & Sons, Inc.

In Figure 4-8, I label the market price and market quantity as P^* and Q^* , respectively. What makes this price and this quantity special is that at price P^* , the quantity that buyers demand is equal to the quantity that producers want to supply. It's the ideal scenario.

Start at price P^* and move along the dashed line to the right. You can see that buyers demand Q^* at that price, and sellers supply Q^* at that price. Because quantity demanded equals quantity supplied, both producers and consumers are content. The consumers get the quantity they want to buy at price P^* , and the producers sell the quantity they want to sell at price P^* . With everyone getting what they want, nobody is going to change anything.

Economists use the word *equilibrium* to describe situations like this in which all involved parties are happy with the status quo and thus have no reason to change their current behavior. By contrast, situations in which one or more of the parties have an incentive to change their behavior are referred to as *disequilibrium* situations.



REMEMBER

A wonderful thing about the model of demand and supply is that at any price besides the market price, P^* , there's always some sort of pressure from either buyers or sellers to bring the model back to the equilibrium price and quantity. Consequently, no matter where the market starts, it always ends at equilibrium — market forces will always push the price and quantity back to these values. Hence, the market price and market quantity are also called the *equilibrium price* and the *equilibrium quantity*.

Demonstrating the stability of the market equilibrium

The market equilibrium is called a *stable equilibrium* because no matter where the demand and supply model starts, it always gravitates back to the market equilibrium. This is nice because it means that markets are self-correcting, and if you know where the demand and supply curves are, you know where prices and quantities will end up. Especially gratifying is the fact that the actions of the market participants — buyers and sellers — move the market toward equilibrium without the need for any outside intervention, such as government regulations.

In this section, I show that the market equilibrium is indeed stable. First I focus on the fact that if prices start higher than P^* , they fall down to P^* . After that I show you that if prices start lower than P^* , they rise up to P^* . The fact that prices always move toward P^* indicates that the market equilibrium is stable.

Excess supply: Reducing prices until they reach equilibrium

In Figure 4-9, you can see what happens when you have a price like P^H that starts out higher than the market equilibrium price, P^* . At price P^H , the quantity demanded by buyers, Q^D , is less than the quantity supplied by sellers, Q^S . (I use dashed lines to show where P^H intersects the demand and supply curves.) Economists refer to such a situation as an *excess supply* or a *surplus*. A situation of excess supply can't be an equilibrium because sellers can't sell everything they want to sell at price P^H .

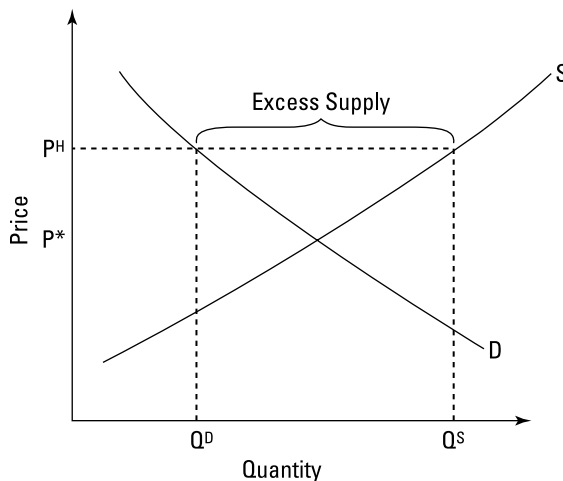


FIGURE 4-9:
Excess supply.

© John Wiley & Sons, Inc.

In fact, of the total amount that sellers want to sell, Q^S , only the amount Q^D is sold, meaning that the remaining amount, $Q^S - Q^D$, remains unsold unless something is done. Well, something *is* done. Sellers see the massive pile of unsold goods and do what any store does when it can't sell something at current prices: They have a sale.

Sellers lower the price and keep lowering it until supply no longer exceeds demand. You can see in Figure 4-9 that sellers keep lowering the price until it falls to P^* because that's the only price at which the quantity demanded by buyers equals the quantity that sellers want to supply.

Excess demand: Raising prices until they reach equilibrium

Figure 4-10 shows a situation in which the initial price, P^L , is lower than the market equilibrium price, P^* . You can see that, in this case, the problem isn't excess supply but *excess demand*. At price P^L , the amount that buyers want to buy, Q^D , exceeds the amount that suppliers want to sell, Q^S .

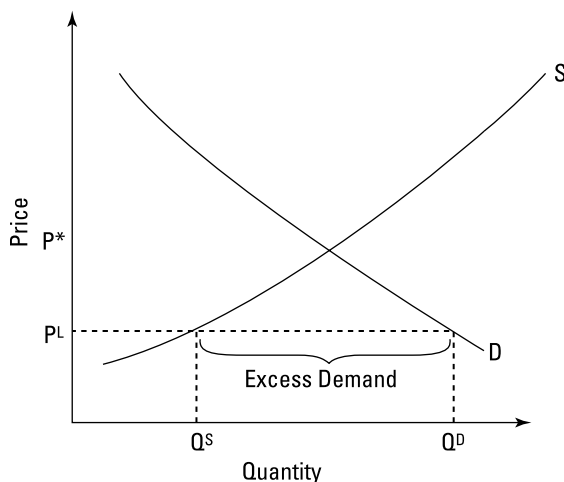


FIGURE 4-10:
Excess demand.

© John Wiley & Sons, Inc.

In other words, there's a *shortage* of $Q^D - Q^S$ units. As a result, buyers start bidding the price up, competing against each other for the insufficient amount of the good.

As long as the price is less than P^* , some degree of shortage exists, and the price continues to be bid up. This means that whenever you start out with a price less than P^* , the price is pushed back up to P^* , returning the market to its equilibrium — the only place where there's neither a shortage nor an excess supply.

Adjusting to new market equilibriums when supply or demand changes

Market forces adjust the price and quantity of a good or service until they correspond to where the demand curve crosses the supply curve. (See the preceding section for details.) After the price and quantity reach that point — the *market equilibrium* — they don't change. They stay right there as long as the demand and supply curves don't move.

In this section, I demonstrate how prices and quantities *do* adjust if the demand and supply curves change. I illustrate by showing you a demand curve shift followed by a supply curve shift.

Reacting to an increase in demand

If demand increases and supply stays the same, the equilibrium price and quantity both increase. Take a close look at Figure 4-11, which shows what happens if the demand curve shifts to the right from D_0 to D_1 while the supply curve S stays the same. Before the shift, the market equilibrium price is P_0^* , and the market equilibrium quantity is Q_0^* . When the demand curve shifts to the right to D_1 , the price momentarily stays the same at P_0^* . But this price can't last because, with the new demand curve, there's now an excess demand. That is, at price P_0^* , the quantity demanded, Q_1^D , exceeds the quantity supplied, Q_0^* .

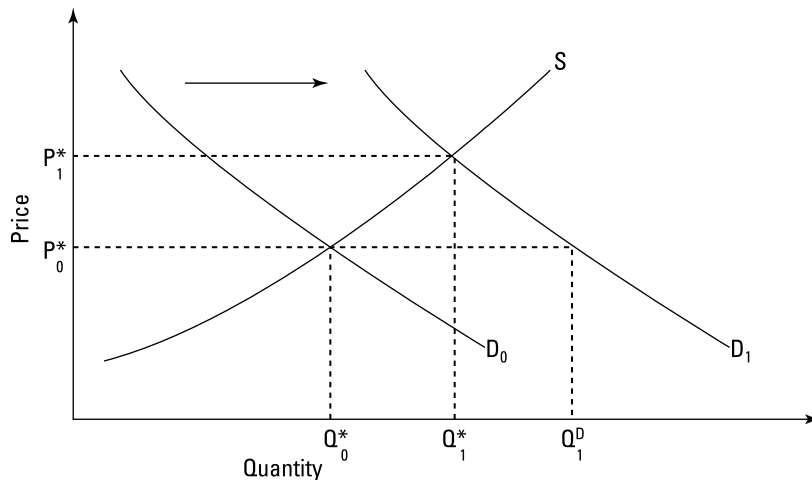


FIGURE 4-11:
A rightward shift
of the demand
curve.

© John Wiley & Sons, Inc.

Any such shortage causes buyers to bid up the price. (See the earlier section “Excess demand: Raising prices until they reach equilibrium.”) The result is that the price rises and continues to rise until it reaches P_1^* , the price where demand curve D_1 crosses supply curve S .

Note that when moving from the first equilibrium to the second, the equilibrium quantity increases from Q_0^* to Q_1^* . This result makes good sense because if demand increases and buyers are willing to pay more for something, you would expect more of it to be supplied. Also, the price goes up from one equilibrium to the other because to get suppliers to supply more in a world of rising costs, you have to pay them more.

However, a much more subtle thing to realize is that the slope of the supply curve *interacts* with the demand curve to determine how big the changes in price and quantity will be. Think of a vertical (perfectly inelastic) supply curve, such as the one in the left-hand graph of Figure 4-7. For such a supply curve, any increase in demand increases *only* the price because the quantity can’t increase. On the other hand, if you’re dealing with a horizontal (perfectly elastic) supply curve, as in the right-hand graph of Figure 4-7, a rightward shift in demand increases *only* the quantity because the price is fixed at P' .

Neither demand nor supply is in complete control in a situation like Figure 4-11. Their interaction jointly determines equilibrium prices and quantities and how they change if the demand curve or the supply curve shifts.

Reacting to a decrease in supply

If supply decreases and demand stays the same, the equilibrium price increases, but the equilibrium quantity decreases. Consider Figure 4-12, where the supply curve shifts from S_0 to S_1 because of increased production costs. (As I discuss in the earlier section “Cost changes: Shifting the supply curve,” this increase in costs can be considered to shift the supply curve either up or to the left. In Figure 4-12, I’ve drawn a vertical arrow to indicate a vertical shift, but I could’ve just as correctly put in a left arrow to show a leftward shift.)

The shift in supply causes the market equilibrium to adjust. The original equilibrium is at price P_0^* and quantity Q_0^* , which is the point where the demand curve D and the initial supply curve S_0 cross. When production costs increase, the supply curve shifts to S_1 .

For a moment, the price remains at P_0^* . But this price can’t continue because the quantity demanded at this price, Q_0^* , exceeds the quantity supplied, Q_1^S . This situation of excess demand causes the price to be bid up until reaching the new equilibrium price of P_1^* , at which price the quantity demanded equals the quantity supplied at Q_1^* .

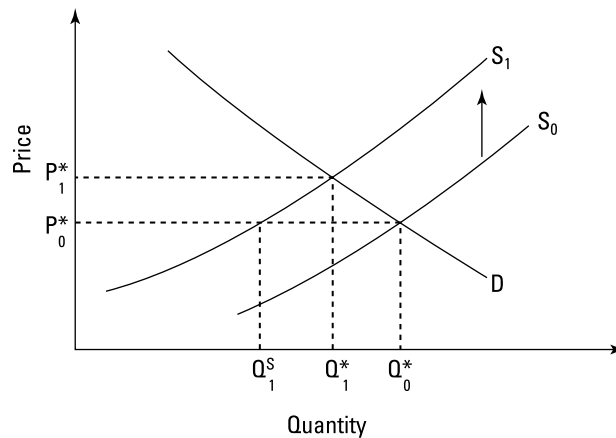


FIGURE 4-12:
A vertical shift of
the supply curve.

© John Wiley & Sons, Inc.

If you compare this increasing-cost situation with the increasing-demand situation of the preceding section, you'll notice that the equilibrium price rises in both scenarios. However, be sure to note that the equilibrium quantities go in opposite directions. An increase in demand causes an increase in equilibrium quantity, but an increase in costs causes a reduction in equilibrium quantity.



REMEMBER

The equilibrium quantity falls when supply decreases because the increase in production costs doesn't just affect the producer. To stay in business, the producer has to pass along the cost increase. But when the increase is passed on, it tends to discourage buyers. The result is that the equilibrium quantity falls because some buyers aren't willing to pay the higher prices. Those who still want to buy are willing to pay the higher prices — a fact that's reflected in the increased market price.

Price Controls: Keeping Prices Away from Market Equilibrium

Left to its own devices, a market always adjusts until the price and quantity are determined by where the demand and supply curves cross. The benefit of the market equilibrium price is that everyone who wants to buy at that price can do so, and everyone who wants to sell at that price can also. In other words, the quantity demanded equals the quantity supplied.

However, the market price isn't always the politically expedient price, and governments sometimes interfere in markets to prevent the market equilibrium from being reached. Such interventions happen because politically influential buyers think the market price is too high or because politically influential sellers believe it's too low.

Unfortunately, government intervention creates a whole new set of problems related to the misallocation of resources. In some cases, government intervention even hurts those whom the intervention is designed to help. In this section, I explain the effects of two kinds of government intervention: price ceilings and price floors. Price ceilings prevent prices from rising to the market equilibrium, whereas price floors keep prices from falling to the market equilibrium. (Obviously, you use only one or the other!)

Setting upper limits with price ceilings

Sometimes the government intervenes in a market to ensure that the price stays below the market equilibrium price, P^* . Such policies are called *price ceilings* because they prevent the price from rising as high as it would if left alone. Prices hit the ceiling and then go no higher.

You may have heard about price ceilings on apartment rent. For instance, in an attempt to provide low-cost housing for those lacking much in the way of finances, city governments may place price ceilings on how much a landlord can charge. The problem is that with prices artificially low, the quantity demanded is greater than the quantity supplied, creating a shortage.

To see how this works, look at Figure 4-13, in which the price ceiling P^C lies below the market equilibrium price of P^* . To make clear that the government has imposed a ceiling price that the market equilibrium price can't rise above, I've drawn a solid horizontal line starting from P^C and extending to the right.

However, at the ceiling price, the quantity demanded, Q^D , far exceeds the quantity supplied, Q^S . This may not seem like such a big problem, but the shortage has to be dealt with somehow. You have to figure out a way to allocate the insufficient supply among all the people who want it. What often happens is that people end up waiting in lines, or *queues*, to get the limited supply.

During my grad school days at Berkeley, there were price ceilings for how much rent a landlord could charge — a policy euphemistically referred to as *rent control*. Because rents were kept far below their market equilibrium value, there were always many more people who wanted to rent apartments than there were apartments available. The result was that any time an apartment became available, you stood in line with — literally — 200 people to fill out a rental application. With so many potential renters vying for one apartment, the property manager could take his pick — hence, rental applications were often five to ten pages long and asked you *everything*. If you weren't picked, you had to get in line for the next apartment that happened to come on the market.

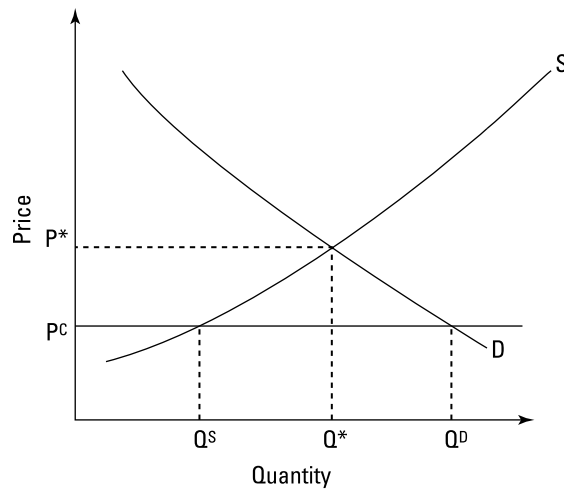


FIGURE 4-13:
A price ceiling.

© John Wiley & Sons, Inc.

It didn't matter if you had enough money to pay higher rent. It also didn't matter if you were much more desperate than other potential renters. Because the price supports imposed by the Berkely City Council had allowed for a situation of excess demand, we had to wait in line and hope and pray that we'd get an apartment.

The main result of the policy was that thousands of people wasted tens of thousands of hours each year waiting in line — and some of them still didn't get apartments! Even worse was the fact that the policy reduced the total number of apartments available in the City of Berkeley. You can see this by the fact that $Q^S < Q^*$ in Figure 4-13. The quantity supplied of apartments falls from Q^* to Q^S because the rent-control price isn't high enough to compensate many property owners for the costs of making their apartments available for rent. Those with the highest costs cease providing apartments when the price of rentals is forced down from P^* to P^C .

Propping up prices with price floors

With a *price floor*, the government keeps the price of a product above its market equilibrium value to raise the incomes of those producing the product. Figure 4-14 shows an example where the floor price, P^F , exceeds the market equilibrium price, P^* . To clarify that prices can't fall below P^F , I've drawn a solid horizontal line at that price.

The problem here is that at price P^F , the quantity supplied, Q^S , is much bigger than the quantity demanded, Q^D . The normal response to such a situation of excess supply would be for the price to fall. The government prevents this and keeps the price propped up by stepping in and buying the excess supply.

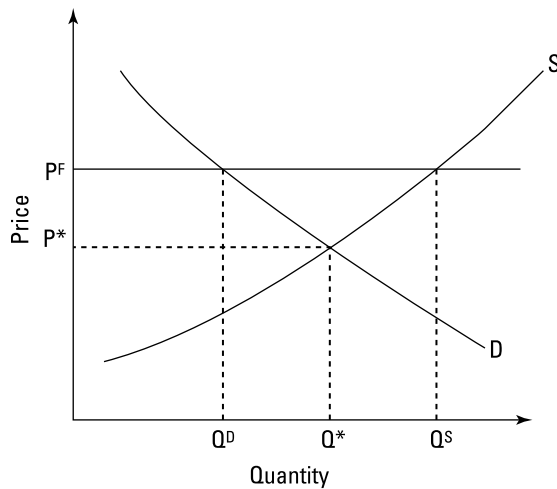


FIGURE 4-14:
A price floor.

© John Wiley & Sons, Inc.

In other words, of the total amount Q^S that's supplied at price P^F , regular consumers demand and purchase Q^D . The government must purchase the remainder, $Q^S - Q^D$. That doesn't sound so bad until you read about price floors in agriculture, which are usually referred to euphemistically as *price supports* (as in, "You poor thing! All you need is a little *support*!").

Price supports generate vast piles of surplus farm output that nobody wants to buy. Consider the price support system for milk that the U.S. Department of Agriculture ran in the 1980s and 1990s. Each year, the Department had to buy up hundreds of thousands of gallons of milk that nobody wanted to buy at the high price the government was maintaining to help dairy farmers. What did the department do with all that milk? It turned it into cheese that it couldn't sell (because it was also supporting the price of cheese) and stored that surplus cheese in massive, refrigerated warehouses — indefinitely! And yes, this was all done at the taxpayers' expense.

Because many people protested this wastefulness, the government switched policies for some crops. Nowadays, the government pays many farmers *not* to farm. That way, the farmers still get paid, but there's no worry about an excess supply that has to be destroyed or given away. (For concerns about food aid, see the nearby sidebar "First-world supports, third-world suffering.") In terms of Figure 4-14, the goal of these policies is to shift the supply curve left until it intersects the demand curve directly over quantity Q^D . If you can do that, there's no excess supply at price P^F .

For both price ceilings and price floors, the message you should take away is that interfering with markets typically causes great mischief. (Yet, deep inside, I'm actually hoping that the government decides to support the salaries of academic economists. I'd love to end up getting paid *not* to teach.)

FIRST-WORLD SUPPORTS, THIRD-WORLD SUFFERING

A perverse result of agricultural price floors (or “supports”) in rich countries like the United States and the nations of the European Union is the significant damage they inflict on developing nations. For instance, the U.S. price of sugar is substantially higher than the world price because the United States restricts imports of cheaper foreign sugar to help U.S. food production giants sell corn-based sweeteners. The result is that thousands of disadvantaged Third World farmers who could otherwise make a living selling sugar to Americans are left without a livelihood.

Even worse is what the United States does with some of the many tons of excess agricultural products that pile up due to its agricultural price supports. Not wanting to sell the excess in the United States and thereby depress U.S. prices, the government often sends the stuff free to developing countries as food aid. That sounds nice and friendly, but when all that free wheat hits Nigeria, for example, it puts Nigerian farmers out of business. That in turn ends up making Nigeria dependent on foreign food because it no longer has enough local farmers to produce sufficient quantities of wheat to feed its own people.

Even worse, many of the newly bankrupted farmers move from the countryside to cities in search of employment, worsening urban-overcrowding problems and themselves becoming dependent on free foreign wheat to survive when most of them can't find work. Thus, the U.S. government's attempt to make its relatively very wealthy farmers a bit richer ends up causing catastrophic harm to the poor farmers of developing nations.

To avoid these problems, economists recommend that rich countries only send food aid to developing countries during times of famine or drought when local farmers can't produce enough to feed their country. Food aid should be withdrawn as soon as a crisis ends so that it doesn't permanently undercut local food prices and the ability of local farmers to make a living.

IN THIS CHAPTER

- » Studying how people maximize their utility (happiness)
- » Taking account of diminishing marginal utility
- » Watching how people weigh alternatives
- » Choosing exactly the right amounts within a limited budget

Chapter 5

Introducing *Homo Economicus*, the Utility-Maximizing Consumer

This chapter gets behind the demand curve (which I introduce in Chapter 4) by showing you how people come to choose the things they do. This decision-making process is critical because human wants drive the economy. Firms don't randomly produce goods and services; they produce the things that people want to buy and are spending money on.

The thing that makes studying this process hard is that people have so many different things they can spend their money on. If economists were asked to research how you'd spend \$100 in a store that sold only blueberry muffins, their job wouldn't be so hard. What's impressive is that economists have come up with a way to explain how you'd spend \$100 in a store that has hundreds or even thousands of items for sale.

Even more impressive is that economists can explain not only which items you'd buy but also how many of each you'd buy. In other words, economic theory can

explain not just *what* you demand but also the *quantities* you demand, which is where demand curves come from.

I start this chapter by discussing *utility*, which is how economists measure happiness or satisfaction. Economists assume that people act in ways that maximize their utility, but human actions are constrained, especially by limited budgets. I explain how people navigate these constraints to get the maximum utility possible, given the limits involved. Finally, I show how these decisions underlie and explain the slope and position of demand curves.

Choosing by Ranking

As you know from Chapter 1, people must make choices because their means for satisfying their wants are limited. There's never enough money or time to do everything. Consequently, people need to choose wisely to get the most happiness out of the limited resources they have.

But for people to choose wisely between the exceedingly different goods and services available in the economy, they must have a way of comparing them. Comparing costs is pretty easy; you just compare prices. But how do you compare the benefits of various goods and services? How do you assess whether it's better to spend \$40 on a giant Swiss chocolate bar or on a new plaid shirt? In what ways are chocolate and shirts even comparable?



REMEMBER

Obviously, people *do* manage to compare and rank the two choices. Economists imagine that people do this by assigning a common measure of happiness to each possible thing they could buy and use. Economists call this common measure of happiness *utility*, and they imagine that if they could somehow get inside your brain and measure utility, they could do so using a unit that they (uncreatively) refer to as a *util*.

Given those assumptions, economists have come up with two alternative ways of thinking about how people make comparisons and choices:

» **Cardinal utility:** Decision-making under this system is based on assigning specific numbers of utils to different things — for instance, 25 utils to the pleasure associated with eating a brownie or 75 utils to the pleasure associated with watching a sunset. Making such specific assignments is called *cardinal utility* (like cardinal numbers: 1, 2, 3 . . .). Some people object to cardinal utility because it's not clear that people make such assessments — after all, how many utils do you think you receive from a sunny day or an infant's smile?

» **Ordinal utility:** A much less objectionable thing to do is to think in terms of *ordinal utility*, a system in which you simply rank things. For instance, instead of saying that the sunset has a utility of 75, which makes it preferred over the brownie with a utility of 25, you can simply say that you prefer sunsets to brownies. This system has a much more intuitive feeling for most people and eliminates the need to try to measure things using the imaginary unit called the util.

It's been proven mathematically that you can describe the same human choice behavior using ordinal utility that you can using cardinal utility, which means that economists don't *have* to use cardinal utility. But I'm going to anyway. Why? Because explaining the crucial concept of diminishing marginal utility using the cardinal utility system is much easier. You can also explain diminishing marginal utility using the ordinal system, but the math is so complex that it's typically taught only to PhD students. So please forgive me if the cardinal utility system seems a bit unrealistic, but it's the best way to convey this fundamental idea.

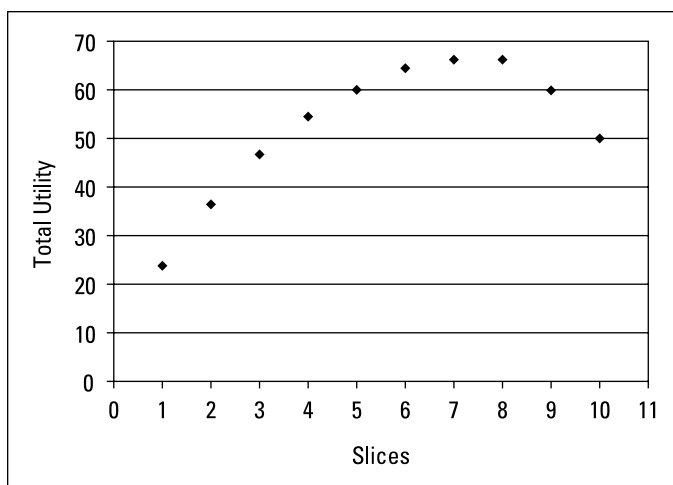
Getting Less from More: Diminishing Marginal Utility

People get bored even with things they like; they get tired of repetition and sameness. Economists have to take account of this when studying how people choose to spend their money.

For instance, if I haven't had pizza in a long time, I get a tremendous amount of utility from eating a slice. The melted cheese, the basil and garlic in the sauce, and the warmth in my mouth all make me very, very happy. But the thrill of pizza is dampened if I eat a second slice; it's still good, but not as good as the first. And if I have a third slice, it's not as good as the second. And if I keep eating and eating and eating, the additional slices of pizza will soon get sickening and produce pain instead of pleasure if I eat them.

This phenomenon isn't limited to pizza; it applies to nearly everything. Unless you're addicted to something, you get tired of it as you have more of it, and each additional unit brings you less happiness than the previous unit. To clarify this phenomenon, look at Figure 5-1, which shows my cumulative, or total, utility as I eat more and more slices of pizza. For instance, my total utility after eating one slice of pizza is 24 utils. After eating two slices, it's 36 utils. And after three slices, it's 46 utils.

FIGURE 5-1:
My total utility as
I eat more and
more slices
of pizza.



© John Wiley & Sons, Inc.

Notice that the amount of total utility changes with each successive pizza slice. These incremental changes in total utility are known as *marginal utility*. As you further consider the numbers, you should notice that the extra, or marginal, utility associated with each slice is decreasing:

- » **First slice:** My total utility increases by 24 utils, from 0 utils to 24 utils. So the marginal utility of my first slice is 24 utils.
- » **Second slice:** My total utility increases from 24 to 36 utils, so the marginal utility for my second slice is 12 utils.
- » **Third slice:** My marginal utility for the third slice is 10 utils because my total utility increases from 36 to 46 utils.



REMEMBER

Economists refer to this phenomenon as *diminishing marginal utility* because the extra utility, or *marginal utility*, that each successive slice brings with it decreases relative to the marginal utility brought by the previous slice. Diminishing marginal utility simply reflects the fact that people get fed up or bored with things. Or, in the case of food and drink, their appetite decreases with each unit they consume.

Look at what happens in Figure 5-1 after slice number eight. My total utility goes down because slice nine makes me a little sick. And if I have slice number ten, I'm even sicker, so total utility falls again. This decrease in total utility implies that marginal utility must be negative for slices nine and ten. The data in Figure 5-1 shows that although my total utility increases for slices one through seven, it stalls at slice number eight and falls for slices nine and ten.

In Figure 5-2, I plot the marginal utility that I get for each slice. I feel diminishing marginal utility as I eat more and more pizza because the marginal utility that comes with each additional slice is always less than that of the previous slice. Specifically, although my marginal utility is 24 utils for the first slice, it falls to 0 utils for slice eight and then becomes negative for slices nine and ten because eating them makes me ill. You can see quite clearly from the downward slope of the points that my marginal utility diminishes as I eat more and more pizza.

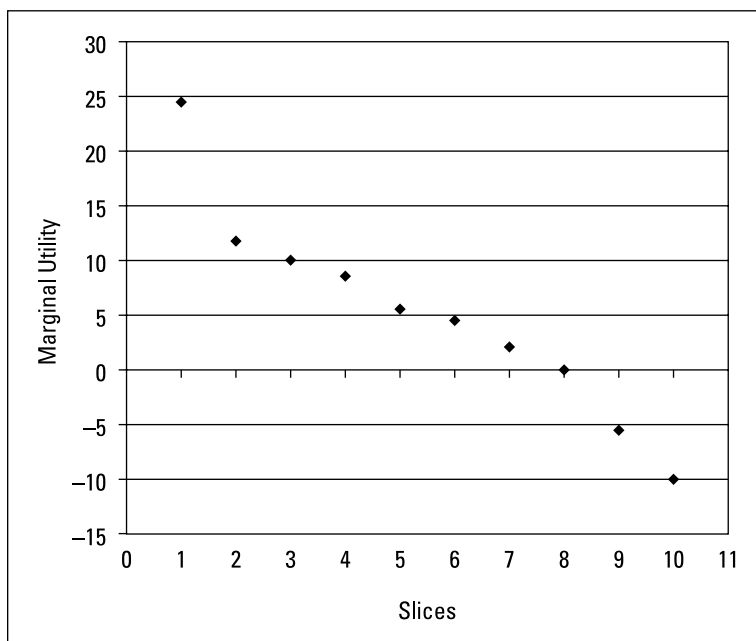


FIGURE 5-2:
The marginal utility I derive from each slice of pizza.

© John Wiley & Sons, Inc.



TIP

You have to be careful not to confuse diminishing marginal utility with *negative* marginal utility. As you see in Figure 5-2, there's diminishing marginal utility for all slices of pizza starting with the second because each successive slice has a smaller marginal utility than the previous one. But the marginal utilities are still positive for all slices up to slice seven, and they become negative only for slices nine and ten.

That fact implies that you enjoy eating every slice up to and including the seventh slice because doing so increases utility (happiness). So don't think that just because marginal utility is diminishing for a particular slice, you wouldn't want to eat it. Marginal utility can be diminishing but still positive. The only slices you'll outright want to avoid are the ninth and tenth.

Choosing Among Many Options When Facing a Limited Budget

The phenomenon of diminishing marginal utility makes studying human choices interesting because preferences can't be determined in the abstract. Instead, they depend on what you've already had.

If I haven't had any ice cream for months and you ask me whether I want chocolate or vanilla, I'll say chocolate. But suppose you ask me whether I want chocolate or vanilla after I've just eaten a gallon of chocolate. In that case, I'm going to say vanilla because I've already more than satisfied my chocolate cravings.

So the answer to the question "Chocolate or vanilla?" isn't as straightforward as it seems. Your preferences exhibit diminishing marginal utility, and even something that you generally like a lot won't bring you much marginal utility (additional happiness) if you've just indulged in it in excess.

This fact leads to a simple rule about how people make decisions when faced with limited budgets. Essentially, people want to maximize their total utility when buying a combination of goods. To do that, the marginal utilities per dollar have to be equal for the final units of each good they purchase. I explain this rule with a concrete example in the next subsection before showing you the mathematical formula that summarizes the simple rule.

Trying to buy as much (marginal) utility as you can

A person interested in maximizing their utility will want to figure out how to best allocate any limited budget so they can purchase the combination of goods and services that will bring them the largest possible amount of utility. But in doing so, this person has to remember that each product is subject to diminishing marginal utility, and each successive unit purchased and consumed will bring with it less additional, or marginal, utility than the previous unit.

Suppose that I have \$10 to spend and, because I'm going to the local student bar, the only two things I can spend it on are pints of beer and slices of pizza. (Do you really crave anything else as a college student anyway?) It becomes clear to me that the intelligent thing to do is to think in terms of buying up as much utility as I can with my limited budget. Both beer and pizza make me happy, but my goal isn't just to be happy; it's to be as happy as possible, given my limited budget. So I want to make sure that every dollar buys me the maximum possible amount of utility.

I don't care where the utility comes from. One util from beer makes me just as happy as one util from pizza; I only care about buying up as many utils as possible. To do that, the critical factor turns out to be the price of utility. Beer and pizza have prices measured in dollars, but what's the price of a util?

Well, it depends. The first three columns in Table 5-1 repeat the data from Figure 5-1 that gave my total and marginal utilities for ten slices of pizza. But the final two columns include new data. (*MU* stands for *marginal utility*.)

TABLE 5-1

Determining the Price of Utility for Pizza

Slice	Total Utility	Marginal Utility	MU per Dollar at \$1 per Slice	MU per Dollar at \$2 per Slice
1	24	24	24	12
2	36	12	12	6
3	46	10	10	5
4	54	8	8	4
5	60	6	6	3
6	64	4	4	2
7	66	2	2	1
8	66	0	0	0
9	60	-6	-6	-3
10	50	-10	-10	-5

What I've done in these last two columns is calculate how much it costs to get some additional happiness (marginal utility) if the way you're getting it is by buying slices of pizza. Consider the fourth column, which assumes that each slice of pizza costs \$1. If you buy one slice, it brings you a marginal utility of 24 utils at a cost of \$1. So the MU per dollar of the first slice is 24.

But now consider spending a second dollar to buy a second slice of pizza. Because that second slice brings a marginal utility of only 12 utils, the MU per dollar spent here is 12. And because diminishing marginal utility continues to decrease the marginal utility of each additional slice of pizza, each additional dollar buys you less additional utility than the previous dollar.

The final column of Table 5-1 shows you that the MU per dollar that you get from pizza depends on how much each slice of pizza costs. If pizza costs \$2 per slice, each dollar spent brings you only half as much marginal utility as when pizza costs \$1 per slice. For instance, because each slice now costs \$2, when you buy the first slice and it brings you 24 utils, you get only 12 utils per dollar spent. Similarly, while the second slice still brings you 12 additional utils of happiness, because it now costs you \$2 to get those utils, your MU per dollar is only 6 utils.

In Table 5-2, I give you the same sort of information as in Table 5-1, but this time it's for my total utility, marginal utility, and MU per dollar when I'm drinking beer that costs \$2 per pint. As you can see from the third column, I exhibit diminishing marginal utility with regard to beer, as my MU for each beer falls from 22 utils for the first pint down to -12 utils for the tenth pint. As a result, my MU per dollar spent in the fourth column falls from 11 per dollar for the first pint down to -6 per dollar for the last pint.

TABLE 5-2

Determining the Price of Utility for Beer

Pint	Total Utility	Marginal Utility	MU per Dollar at \$2 per Pint
1	22	22	11
2	40	18	9
3	56	16	8
4	70	14	7
5	80	10	5
6	86	6	3
7	88	2	1
8	88	0	0
9	82	-6	-3
10	70	-12	-6

Purchasing the best combination of two goods to maximize total utility

When facing a limited budget, the trick is to see how to allocate purchases to get the most utility. Tables 5-1 and 5-2 show how much utility I can get by spending money on either pizza or beer. I have \$10 to spend. As a first attempt, consider the

two most extreme options: blowing all the money on pizza or blowing all the money on beer. (Pizza costs \$1 per slice, and beer costs \$2 per pint.)

If I spend all \$10 on pizza, I can buy 10 slices of pizza, giving me a total utility of 50 utils. On the other hand, if I spend all \$10 on beer, I can buy 5 pints at \$2 each and thereby get 80 total utils. If these were my only two options, I'd clearly prefer to spend all my money on beer because it brings me more utils than does buying only pizza.

However, there's a much better thing to do. I can get even more total utility if I wisely mix up my consumption a bit and spend some of my money on beer and some of it on pizza.



REMEMBER

The way I get the most utility possible out of my \$10 is simple: I take each of the ten dollars in turn and spend it on whichever good brings more utility. I don't think of my task as buying slices of pizza or pints of beer. Rather, my job is buying utility. For every dollar spent, I want to buy as much utility as possible, and I don't care whether that utility comes from beer or pizza.

The only thing complicating this process of spending each dollar on whichever good will bring the most utility is the fact that I have diminishing marginal utility for both beer and pizza. Thus, the amount of utility I can buy with each extra dollar that I spend depends on how much beer or pizza I've already bought. But given the information in Tables 5-1 and 5-2, I can figure out the best thing I should do with each dollar:

- » **Dollar 1:** What should I do with the first dollar? From the fourth column of Table 5-1, you can see that if I spend that dollar on pizza, I can buy 24 utils of utility. On the other hand, the fourth column of Table 5-2 tells you that if I spend that first dollar on beer (along with a second dollar because pints cost \$2), I get only 11 utils of utility. So, the obvious thing to do with the first dollar is to buy pizza rather than beer.
- » **Dollar 2:** If I use my second dollar to buy a second slice of pizza, I get 12 utils of utility. If I buy beer with that second dollar (along with a third dollar because the price of a pint is \$2), I get only 11 utils for that second dollar because I spent it on buying the first pint. So, it's better to spend the second dollar on pizza rather than beer.
- » **Dollars 3 and 4:** At dollar number three, everything changes. That's because if I spend a third dollar on pizza, it brings 10 utils. But if I spend that third dollar (along with a fourth dollar since each pint costs \$2) on a pint, I get an MU per dollar of 11 utils (for each dollar). So I should spend dollars three and four on buying the first pint of beer.

- » **Dollar 5:** I want to spend the fifth dollar on pizza rather than beer because I get 10 utils of marginal utility if I use this dollar to buy the third slice of pizza but only 9 utils of marginal utility if I use this dollar (along with a sixth dollar since pints cost \$2 each) to purchase a second pint of beer.
- » **Dollars 6 and 7:** I should spend dollars six and seven on beer because I get an MU per dollar of nine utils for my second pint, whereas I get only eight utils if I spend the sixth dollar on a fourth slice of pizza.
- » **Dollars 8, 9, and 10:** The MUs per dollar are tied for dollar number eight. If I use this dollar to buy a fourth slice of pizza, I get eight utils. I get the same by spending that eighth dollar (plus the ninth dollar, because pints cost \$2 each) on a third pint of beer. So I should spend my last three dollars buying a fourth slice of pizza and a third pint of beer.

In Table 5-3, I list where I should spend each of my ten dollars. Notice that the total utility I can purchase with my \$10 is 110 utils. That's much better than the 50 utils I'd get spending all the money on pizza or the 80 utils I'd get spending all of it on beer. By spending each dollar in sequence on whichever good brings the most utility, I've done much better than I could by spending the money on only one good or the other.

TABLE 5-3

How I Optimally Spend Each Dollar in my \$10 Budget When Pizza Costs \$1 per Slice and Beer Costs \$2 per Pint

Dollar	Good Chosen	MU per Dollar
1	Pizza (1 st slice)	24
2	Pizza (2 nd slice)	12
3	Beer (1 st pint)	11
4	Beer (1 st pint)	11
5	Pizza (3 rd slice)	10
6	Beer (2 nd pint)	9
7	Beer (2 nd pint)	9
8	Pizza (4 th slice)	8
9	Beer (3 rd pint)	8
10	Beer (3 rd pint)	8
Total utils		110

Also notice that I end up buying four slices of pizza and three pints of beer. Given this budget and these prices, my quantity demanded of pizza is four slices, and my quantity demanded of beer is three pints. The process of maximizing utility is also the basis of demand curves and the relationship between quantity demanded and price. (I discuss demand curves in Chapter 4 and return to them later in this chapter, in the section “Deriving Demand Curves from Diminishing Marginal Utility.”) In the next section, I present the magic formula for choosing where to spend your money.

Aiming for equal marginal utility per dollar

In this section, I explain a simple formula that guides people to maximize the total utility they can get out of spending any budget — no matter how many goods they can choose from or how much the goods cost.

To keep things simple, I begin by showing you the version of the formula that applies to deciding how to best spend your budget when you’re choosing from only two goods or services. When you get the hang of the two-good version, the multi-good version is effortless.

Call the two goods X and Y , and say that their respective prices are P_X dollars for each unit of X and P_Y dollars for each unit of Y . Also, their respective marginal utilities are MU_X and MU_Y . The formula looks like this:

$$\frac{MU_X}{P_X} = \frac{MU_Y}{P_Y} \quad (1)$$



REMEMBER

What the formula means is that if a person has allocated their limited budget optimally between the two goods, then at the optimal quantities of X and Y , the marginal utilities per dollar of X and Y are equal.

This relationship holds in the example in the preceding section. Look back at Table 5-3. When I optimally spend my \$10 on beer and pizza, the optimal amounts of each are four slices of pizza and three pints of beer. From the third column of Table 5-3, you can see that marginal utilities per dollar for the fourth slice of pizza and the third pint of beer are indeed equal at 8 utils per dollar, just as the formula in Equation (1) dictates.

Seeing why the marginal utilities per dollar must be equal

If marginal utilities per dollar aren’t identical, keep rearranging your purchases until they are. In this section, I demonstrate why marginal utilities per dollar have to be equal if you want to maximize utility when spending a limited budget.

First, imagine that I choose some other quantities of each good, so that for the final unit of X and the final unit of Y that I purchase,

$$\frac{MU_X}{P_X} > \frac{MU_Y}{P_Y} \quad (2)$$

Let pizza be X and beer be Y. From the respective fourth columns of Tables 5-1 and 5-2, you can see that if I purchase four pints of beer and two slices of pizza, the MU per dollar for the fourth pint of beer is 7, and the MU per dollar for the second slice of pizza is 12. Clearly, the MU per dollar of pizza is much bigger than the MU per dollar of beer if I spend my limited budget in this way.

But this way of spending my budget isn't optimal. That's because the money I'm spending on what's currently the final unit of X (pizza) buys more marginal utility than the money I'm currently spending on the final unit of Y (beer). Suppose I can get more utility by spending a dollar on X than Y. In that case, I should take money away from spending on Y to spend it on X. And as long as the inequality in Equation (2) holds, I should continue to take money away from Y to increase spending on X.

Applying the formula to multiple goods and services

To maximize total utility, you should rearrange your purchases so that for the final units of each good, the marginal utilities per dollar are equal. If that isn't true, one of the goods offers you a higher amount of happiness for each dollar spent, meaning that you want to rearrange your purchases to spend more on that good. Only when Equation (1) (from earlier) holds will you not want to rearrange any further because neither good offers you more happiness per dollar than the other.

Also realize that Equation (1) can be generalized for many goods. For example, in the case of three goods, you would arrange your buying so that for the last unit of each of the three goods X, Y, and Z:

$$\frac{MU_X}{P_X} = \frac{MU_Y}{P_Y} = \frac{MU_Z}{P_Z} \quad (3)$$

If any of the three goods has a higher marginal utility per dollar than the others, you rearrange your purchases to buy less of the others and more of that good. And you keep rearranging until Equation (3) holds true.

Deriving Demand Curves from Diminishing Marginal Utility

Diminishing marginal utility is one reason demand curves slope downward. You can get a hint of this from Figure 5-1, where you see that the marginal utility that comes with each successive slice of pizza decreases. Suppose your goal is to use your money to buy up as much utility as possible to make yourself as happy as possible. In that case, you're willing to pay less and less for each successive slice of pizza because each successive slice brings less utility than the previous slice.

However, Figure 5-1 isn't a demand curve:

- » **Price:** It doesn't consider the effect of (own) price on quantity demanded.
- » **Other goods:** It looks at only one good in isolation, whereas the quantity demanded of a good is determined by finding the solution to the more general problem of allocating a limited budget across all available goods to maximize total utility. In other words, you can't look at each good in isolation. How much of it you want to buy depends not only on its price but also on the prices of everything else and how their marginal utilities vary as you buy more or less of them.

This section builds up demand curves from first principles, showing how changes in a good's own price cause movements along that good's demand curve if you hold all other possible influencing factors constant. It also shows you how changes in any of those other possible influencing factors — including the prices of other goods as well as changes in preferences — can shift the entire demand curve to a new position.

Seeing how price changes affect quantities demanded

Price changes affect the quantity demanded of each good. As the prices of different goods vary with respect to each other, so do the quantity demanded of each good to obtain as much utility as possible given the new prices.

In the example I've been using in this chapter, I've had to decide how to best spend \$10 when my choices are slices of pizza or pints of beer. I want to make one change to that example: Say that pizza now costs \$2 per slice rather than \$1 per slice. What I want to show you is how this price change affects the quantity demanded of both pizza and beer.

The changes in quantities demanded result from the fact that the new, higher price of pizza reduces the marginal utility per dollar of pizza. Doubling the price of pizza means that the marginal utility per dollar generated by each slice of pizza is exactly half of what it was before. You can see this by comparing the fourth and fifth columns of Table 5-2. Because the increase in price lowers the marginal utility that each dollar spent on pizza buys, it naturally affects where I spend my limited budget of \$10.

As you may expect, a higher price of pizza leads me to eat less pizza and drink more beer. I can prove this by spending, in order, each of my dollars so that I buy whichever good has the higher marginal utility. (The earlier section “Purchasing the best combination of two goods to maximize total utility” walks through the process.) The results are summarized in Table 5-4.

TABLE 5-4

How I Optimally Spend My \$10 Budget When Pizza Cost \$2 per Slice and Beer Costs \$2 per Pint

Dollar	Good Chosen	MU per Dollar
1	Pizza (1 st slice)	12
2	Pizza (1 st slice)	12
3	Beer (1 st pint)	11
4	Beer (1 st pint)	11
5	Beer (2 nd pint)	9
6	Beer (2 nd pint)	9
7	Beer (3 rd pint)	8
8	Beer (3 rd pint)	8
9	Beer (4 th pint)	7
10	Beer (4 th pint)	7
Total utils		94

By comparing Table 5-4 with Table 5-3, you can see that raising the price of pizza from \$1 to \$2 has affected not only my quantity demanded of pizza but also my quantity demanded of beer. For pizza, my quantity demanded has fallen from four slices down to only one. For beer, my quantity demanded has increased from three pints to four pints.

The increase in the price of pizza has also made me poorer in the only sense that really matters: I'm less happy. Due to the price increase, the total number of utils that I can purchase with my \$10 budget has fallen from 110 to only 94. Despite rearranging my quantities consumed of beer and pizza to make the most of the new situation, the price increase still hurts me overall.

Graphing the price and quantity changes to form a demand curve

You can use information about how quantity demanded changes when price increases to plot points on a demand curve. For instance, in the pizza example, you can plot these two points: four slices demanded at a price of \$1 and one slice demanded at a price of \$2. Figure 5-3 plots these two points and sketches in the rest of the demand curve. Keep in mind two things:

- » The downward slope of the pizza demand curve derives in part from the diminishing marginal utility of pizza.
- » As the price of pizza changes, the quantity demanded of pizza doesn't change in isolation; it changes as a result of rearranging the quantities demanded of pizza *and* beer to maximize total utility.

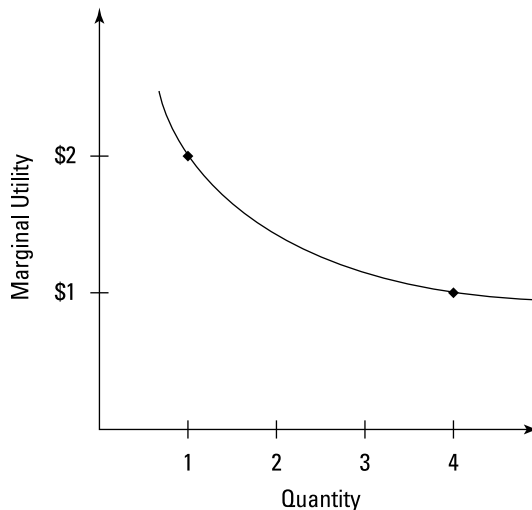


FIGURE 5-3:
My demand curve
for slices of pizza.

© John Wiley & Sons, Inc.



REMEMBER

Demand curves for individual goods aren't made in isolation. Certainly, a relationship exists between a good's price and its quantity demanded. However, when the good's price changes, that change affects the entire budgeting decision — not just for that good, but for *every* good. The resulting change in the good's quantity demanded is just part of the overall rearrangement of spending that strives to keep maximizing total utility given the new price.

Consider how the increase in the price of pizza affects the demand curve for beer. My quantity demanded of beer went from three pints to four pints when the price of pizza increased from \$1 to \$2. But the price of beer was unchanged. This means that the demand curve for beer must have shifted (which I explain in Chapter 4). I illustrate this shift in Figure 5-4. Point A on demand curve *D* shifts over to become Point A' on demand curve *D'*.

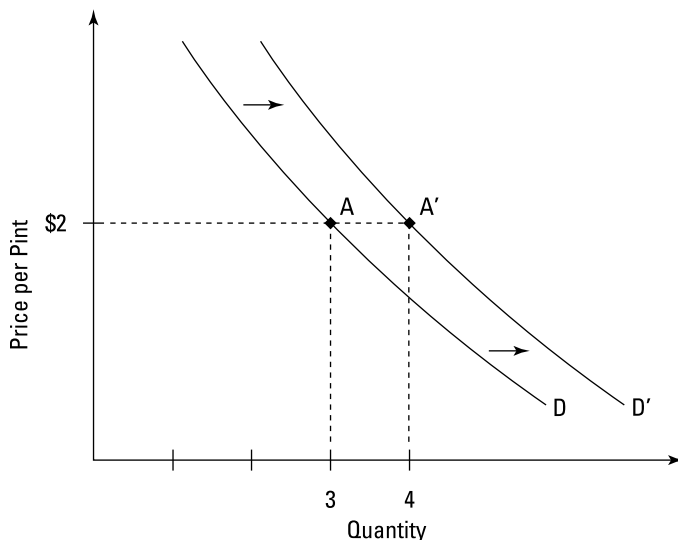


FIGURE 5-4:

My demand curve for beer shifts to the right when the price of pizza increases.

© John Wiley & Sons, Inc.



REMEMBER

Events where changes in the price of one good affect the quantity demanded of another good are called *cross-price effects*. By contrast, when a change in a good's own price affects its quantity demanded, you have *own-price effects*. Please note that although cross-price effects cause demand curves to shift, own-price effects cause movements along given demand curves.



TIP

The direction of a cross-price effect depends on the situation. In this chapter, I allow consumers to purchase only two goods: beer and pizza. The result is that when the price of pizza goes up, consumers switch some of their purchasing power to buying beer — or, as economists say, they *substitute* from one good to the other. (See the sidebar “Examining complementary and substitute goods.”) That’s why when the price of pizza goes up, the demand curve for beer in Figure 5-4 shifts to the right.

But in the real world, where many other consumption goods are available, the demand curve might very well shift in the other direction. For instance, some people like drinking beer only when they eat pizza. For them, an increase in the price of pizza may decrease both the amount of pizza eaten and the amount of beer drunk.

Such people think of beer and pizza as a bundle. An increase in the price of one member of the bundle increases the price of the entire bundle. These people buy less of each member of the bundle to free up money to spend on the many other available consumption goods. For consumers with these preferences and with the option of buying goods besides beer and pizza, when the price of pizza goes up, the demand curve for beer shifts left.

EXAMINING COMPLEMENTARY AND SUBSTITUTE GOODS

Some things just go together. Hot dogs and hot dog buns. Hamburgers and ketchup. Strawberries and shortcake. In each of these pairs, the goods in question are more useful or more pleasing when consumed along with the other member of the pair.

Because such goods complement each other, economists refer to them as *complementary goods*. An interesting thing about complementary goods is that changes in the price of one complement affect the other complement. For instance, if hot dogs go on sale, not only do people buy more hot dogs, they also buy more hot dog buns. And more mustard is sold too.

By contrast, consider *substitute goods* — goods that serve similar functions so that if the price of one goes up, people switch to the other one. For instance, if the price of train travel goes up, more people drive cars. And if the cost of postal mail increases, more people use email.

(continued)

(continued)

Both complementary goods and substitute goods result from cross-price effects. An increase in the price of a complement causes the quantity demanded of its pair to fall, whereas an increase in the price of a substitute causes the quantity demanded of its pair to rise.

As you look around the economy, make sure you think of it as one great big organic whole where things don't happen in isolation. When the price of one good changes, it affects not just that good but many other goods that are either substitutes or complements. And if the prices of the substitutes or complements change as a result of the initial price change, all their substitutes and complements are also affected. It's a gigantic ripple effect.

IN THIS CHAPTER

- » Understanding why firms choose to maximize profits
- » Deconstructing a firm's cost structure
- » Determining a firm's profit-maximizing output level
- » Seeing how costs determine a firm's supply curve
- » Understanding how firms react to losing money

Chapter 6

The Core of Capitalism: The Profit-Maximizing Firm

Business enterprises make nearly everything you eat, drink, wear, drive, ride, fly, and use. So, naturally, economists devote a lot of effort to studying how businesses behave.

In this chapter, I show you how economists model a firm that's a member of a competitive industry, meaning a firm that's just one of many firms competing against each other for your business. It's essential to understand how firms behave in competitive industries for two reasons:

- » Most firms in the real world face a lot of competition because they're members of either perfectly competitive industries (which I talk about in this chapter) or monopolistically competitive industries (which I discuss in Chapter 9).

» All firms — even those that don't face much competition — behave in remarkably similar ways.

Above all, firms like to maximize profits. And even more importantly, all firms maximize profits in the same way: by producing precisely the output level at which the cost of producing one more unit just equals the additional revenue that the firm gets from selling that unit.

In this chapter, I show you why firms behave this way. When you know that, you'll understand how *all* firms work, whether they face intense competition from rivals or have no competitors at all.

A Firm's Goal: Maximizing Profits

People bring firms into existence so they can produce things. That statement should make you want to ask a fundamental question: *Why* do people bother creating firms to make things? One reason could be altruism. Another could be that making things is fun. A third could be that the people who start a firm are bored doing other things, like scrolling through ridiculous TikTok videos. But economists think the answer is much more straightforward: They like cold, hard cash.



REMEMBER

Economists assume that the overriding goal of all corporations is to make as big a profit as possible. Economists make this assumption for two reasons:

- » If you ask around, profit maximization is near the top of every firm's to-do list.
- » No matter what other goals a firm may have, it still wants to maximize profits after taking steps to achieve those other goals.

For instance, a firm that wants to have a factory that emits no greenhouse gases still wants to make as much money as possible after it builds such a factory. After it's taken the steps necessary to protect the environment, why not make a big profit?

As another example, consider the ice cream company Ben & Jerry's. When it started, it donated a large percentage of its profits to charity, making you feel good about your Chunky Monkey indulgences before bedtime. Given such a policy, the best way to help worthy causes was for Ben & Jerry's to make as big a profit as possible.

Many noneconomists object to firms focusing so intently on profits, but profits ensure that firms are able to recruit, and then sensibly deploy, the economic input

of entrepreneurial ability and one of its key characteristics: the prudent and calculated willingness of entrepreneurs to engage in risk-taking.

Think of someone who's working for someone else and receiving a respectable, steady wage. What is their incentive to strike out on their own and risk starting a business that may fail? The incentive is that this person will receive the profits if the business succeeds. Without potential profits, few people would be willing to risk leaving a safe job to innovate, and consumers would be hurt because the supply of new products and services would come to a halt. (In Chapter 14, I explain in detail how "entrepreneurial ability" receives profits in exchange for assuming the financial risks associated with innovation and invention.)

Facing Competition

Firms may or may not face a lot of competition from other firms. At one extreme lies *monopoly*, in which a firm faces no competition because it's the only firm in its industry. At the other extreme lies what economists call *perfect competition*, a situation in which a firm competes against many other firms to produce an identical good. Between those polar opposites lie two intermediate situations: *oligopoly*, whereby two, three, or (at most) a few firms compete in an industry; and *monopolistic competition*, which is marked by many competitors, but each produces a slightly different good. (See Chapters 8 and 9 for details on monopolies, oligopolies, and monopolistic competition.)

In this chapter, you find out how firms behave under perfect competition. In addition to explaining how essential markets like the stock market behave, this type of market structure is the most straightforward case to understand. It's simple because when an industry has many competitors and every firm produces identical products, none of them has control over the prices it charges.

Listing the requirements for perfect competition

To see why firms engaging in perfect competition have no control over the prices they charge, you have to understand that perfect competition assumes three things about the firms in an industry:

1. There are many firms.
2. Each represents a tiny part of the industry.
3. All sell a commodity, by which economists mean an identical or nearly identical product.

Wheat farming is an example of an industry that satisfies the three criteria of perfect competition. Wheat farmers in the United States number in the tens of thousands. None of them produces more than a small percentage of the total wheat produced each year, and all their wheat is essentially identical, making it a commodity product.

To see why these things together mean that individual farmers have no control over the price of wheat, start with the fact that the farmers are producing an identical product. Because the wheat from one farm looks like the wheat from any other farm, the only way a Kansas wheat farmer can entice me to buy from him rather than from a Texas wheat farmer is to offer me a lower price. Because all the wheat is identical, I only care about price, meaning that farmers must compete on price and price alone.

With price jumping to the fore as the key factor in the wheat market, you can use supply and demand analysis to determine the price. The price is determined by where the *market demand curve* for wheat crosses the *market supply curve* for wheat. (See Chapter 4 for details.) How are these curves determined?

- » **Market demand curve:** Determined by adding up the individual demand curves of all the people who want to buy wheat
- » **Market supply curve:** Comes from adding up the individual supply curves of all the individual wheat farmers

This is where the first two assumptions of perfect competition come into play. Because of the thousands of wheat farmers, and because each produces such a tiny part of the total supply of wheat, the market supply curve for wheat is unaffected by the presence or absence of any individual supply curve of any particular farmer. If a billion bushels of wheat are sold every year, the market price is unaffected by whether a small farmer with only 1,000 bushels to sell bothers showing up to the market. He's just too small a player to cause the market price to change.

If every player is too small to cause the market price to change, each one has to accept whatever price is generated by market demand interacting with market supply.

Taking prices but setting quantities

If the three assumptions of perfect competition are met, they produce a situation in which individual firms have no control over the prices they can charge. In fact, under perfect competition, economists refer to firms as *price takers* because they have to take the price as given and deal with it.



REMEMBER

When you come right down to it, even the most powerful firm can hope to control only two things: how much of its product to make and what price to charge. Because firms have no control over their prices under perfect competition, that narrows the list to one. The only thing that price-taking firms can control is how much to produce.

Firms choose to make whatever quantity maximizes their profits. This fact is mathematically convenient because the quantity of output that a firm decides to produce controls each of the two things that determine profits: total revenues and total costs.



REMEMBER

To see this fact more clearly, you must know that a firm's profit is defined as its total revenue minus its total costs. Put into math, you get the following equation, where TR stands for total revenue and TC stands for total costs:

$$\text{Profit} = TR - TC \quad (1)$$

For a competitive firm, total revenue, TR , is simply the quantity, q , of its output that it chooses to sell times the market price, p , that it can get for each unit:

$$TR = pq \quad (2)$$

That means if I can sell apples for \$1 each and I sell 37 apples, my total revenue is \$37 (\$1 per apple $\times$ 37 apples = \$37).

But notice that because the price at which I can sell (p) is out of my hands if I'm a price taker, the only way I can control my total revenue is by deciding how many apples to sell. So a firm can determine its total revenue by its decision about how big or small to make q .

Much of this chapter shows that the firm's total costs, TC , are also determined by how big or small q is. But the interesting thing here is that although each extra unit of q sold brings in a revenue of p dollars, the cost of each unit of q manufactured depends on how many units of q have already been made. Costs tend to increase as firms produce more and more, so each successive unit costs more than the previous unit. This fact limits the number of units a firm wants to produce.

Suppose I can sell as many apples as I want for \$1 each. The first apple costs 10 cents to produce, the second costs 20 cents, the third costs 30 cents, and so on. In such a case, I'm willing to produce no more than ten apples. Why? Because for each of the first nine apples, I'll make a profit, but for apple ten (which costs \$1 to produce), I'll break even. If I produce any more apples, I'll sustain a loss. (Apple number 11, for instance, would cost \$1.10 to make, but I'd get only \$1 for selling it.)

Consequently, you can see that the firm's choice of q determines both the TR and the TC terms in profit equation (1). The only thing left to figure out is precisely

how big to make q to maximize profits. It turns out that there's a straightforward formula that gives the solution. Pay attention because you just may, uh, profit from reading this chapter.

Distinguishing between accounting profits and economic profits

To an economist, the terms *profit* and *loss* refer to whether the revenue from selling a firm's output is bigger or smaller than the costs that must be incurred to produce that output. If the revenue exceeds the costs, the firm is *running a profit*, whereas if the costs exceed the revenue, the firm is *running a loss*. If the two are equal, the firm is *breaking even*.

Consider a business that sells lemonade. The accountant and the economist agree that the firm's revenue is how much money it makes from selling its product. However, they differ on what to count as "costs" when calculating profit:

» **Accounting profit:** The accountant considers costs to be only the amounts of money spent in running the business: how much the firm pays its workers, how much it pays to buy supplies, and so on. If the firm has revenues of \$10,000 and it spends \$9,000 to make those revenues, the accountant concludes that the firm has a profit of \$1,000. This number is the firm's *accounting profit* — the profit reported daily in financial statements and newspaper articles.



REMEMBER

» **Economic profit:** Economic profit considers not just the money costs directly incurred by running a business but also the opportunity costs incurred. Think about the entrepreneur who starts this lemonade business. After paying for her materials and her employees' wages, her accounting profits are \$1,000. But unless you're ten and have a stand set up in your front yard, is that really a good deal?

Suppose that this entrepreneur left a job as a computer programmer to open the lemonade business, and in the same amount of time that it took the lemonade business to turn a \$1,000 profit, this person would have made \$10,000 in wages if they'd stayed at their old job. That is, the entrepreneur gave up the opportunity to earn \$10,000 in wages to open up a business that makes them only a \$1,000 accounting profit. This person actually sustains an *economic loss* of \$9,000. When you know this fact, the decision to switch careers doesn't seem like such a good idea.

Economists like to concentrate on economic profits and losses rather than accounting profits or losses because the economic profits and losses are what motivate people. By considering economic profits and losses, you get directly at what motivates firms to produce not only the types of goods they choose to make

but the quantities of those goods as well. In this example, you can imagine that when other computer programmers see what happened to this entrepreneur who switched careers, they won't follow the same path.

For the rest of this chapter, whenever you see costs listed, assume that they're *economic costs*; that is, they include not only money directly spent operating a business but also the costs of other opportunities forgone to operate the business. Likewise, whenever you see a profit or a loss, assume that it's an economic profit or an economic loss — the factor that motivates entrepreneurs to want to do something or avoid it.

The most important application of this concept is to determine how much output a firm should produce. If producing the 12th unit of a product generates an economic profit, obviously the firm wants to make it. But if increasing production to a 13th unit would result in an economic loss, the firm doesn't want to produce it.

Analyzing a Firm's Cost Structure

To see how costs and revenues interact to determine economic profits or losses, economists like to break up a firm's total costs into two subcategories:

- » **Fixed costs:** Fixed costs are costs the firm must pay even if it isn't producing anything. For instance, after a rental contract is signed for the firm's headquarters, the firm must pay that rent whether it produces anything or not. Similarly, if the firm has taken out a loan, it's legally required to make its debt payments, whether it's producing zero units of output or a billion.
- » **Variable costs:** Variable costs are costs that vary with the amount of output produced. For instance, if you're in the lemonade-making business and choose to produce nothing, you don't have to buy lemons. But the more lemonade you do produce, the more you spend buying lemons. Similarly, producing more lemonade requires more workers, so your labor costs also vary with the amount of output you produce.

Fixed costs can be represented as *FC* and variable costs as *VC*. Together, they sum up to a firm's total costs, or *TC*:

$$TC = FC + VC \quad (3)$$



TIP

As you look at Equation (3), remember that it deals with the economic costs facing the firm and therefore captures the opportunity costs of the firm's expenditures on both fixed and variable costs. (All expenditures, whether fixed or variable, involve opportunity costs — the other things you gave up buying so you could spend the money on your fixed and variable costs.)

Focusing on costs per unit of output

Economists distinguish between fixed and variable costs because they have very different effects on a firm's decision regarding how much to produce. Look at Table 6-1, which gives data on LemonAid Corporation, a lemonade producer.

TABLE 6-1 The Cost Structure of LemonAid Corporation

Workers	Output	Fixed Costs	Average Fixed Costs	Variable Costs	Average Variable Costs	Total Costs	Average Total Costs	Marginal Costs
0	0	100	—	0	—	100	—	—
1	50	100	2.00	80	1.60	180	3.60	1.60
2	140	100	0.71	160	1.14	260	1.86	0.89
3	220	100	0.45	240	1.09	340	1.55	1.00
4	290	100	0.34	320	1.10	420	1.45	1.14
5	350	100	0.29	400	1.14	500	1.43	1.33
6	400	100	0.25	480	1.20	580	1.45	1.60
7	440	100	0.23	560	1.27	660	1.50	2.00
8	470	100	0.21	640	1.36	740	1.57	2.67

When LemonAid Corporation gets started, it buys a juicer machine for \$100, which gives it fixed costs of \$100. It then has to decide how much to produce, which in turn determines how many workers it needs to hire. In the first column, the number of workers varies from zero to eight. If the firm employs no workers, you can see in the top entry of the second column that no output is produced. But if it hires workers, output increases as you move down the second column. More workers mean more output.

Studying increasing and decreasing returns

Pay attention to the fact that the amount of additional, or marginal, output produced by each additional worker isn't constant. In the LemonAid Corporation example, output increases from nothing to 50 bottles of lemonade if you go from no workers to one worker. However, as you go from one worker to two workers, output increases from 50 bottles to 140 bottles. Put into economic jargon, the second worker's *marginal output* is 90 bottles, whereas the first worker's marginal output is only 50 bottles. (*Note:* Economists sometimes refer to a firm's output as its *product*, so a worker's marginal output can also be referred to as their *marginal product*.)

Now look at these facts in terms of costs and benefits. If you have to pay each worker the same wage of \$80 per day (\$10 per hour for 8 hours of work), you're going to like the fact that although the first worker produces 50 bottles for their \$80 pay, the second worker produces 90 bottles for their \$80 pay.

Determining the cause of diminishing returns

I go into detail about what causes diminishing returns in Chapter 3, but I briefly explain here. What's going on is that LemonAid Corporation bought only one juicer machine for squeezing the juice out of lemons.

The first worker can use the machine to squeeze enough juice for 50 bottles by carrying lemons to the machine and then operating the machine. But it turns out that two workers can do even better by dividing up the work: One can bring lemons to the machine, and the other can operate it. Working together, they can produce 140 bottles — more than double the 50 bottles that one worker could produce working alone.

However, a third worker doesn't increase output nearly as much as a second because the two primary tasks — carrying and operating — have already been taken care of. At best, the third worker can just help the first two workers do these tasks a little faster. The same holds for all successive workers: Having them is helpful, but each one adds less to output than the previous one because things start getting crowded, and there isn't much room left for improvement.

Examining average variable costs

Variable costs are affected by the fact that additional workers first bring increasing returns (when the firm's capital is understaffed) but then bring decreasing returns (when the capital is overstaffed). In the case of the LemonAid Corporation example in Table 6-1, the variable costs are all labor costs, with each worker having to be paid \$80 per day. You can see that these variable costs increase as you move down the fifth column.



REMEMBER

What's much more interesting than looking at variable costs is looking at *average variable costs* (AVC), which are defined as variable costs divided by quantity (VC/q). For instance, because one worker produces 50 bottles of output at a variable cost of \$80, the average variable cost is $\frac{\$80}{50} = \1.60 per bottle. When two workers together cost \$160 in variable costs but produce 140 bottles, the average variable cost for two workers is only $\frac{\$160}{140} = \1.14 per bottle.

The decrease in average variable costs results from increasing returns. When moving from one worker to two workers, variable costs double (from \$80 to \$160), but output more than doubles (from 50 bottles to 140 bottles).

When diminishing returns set in, average variable costs start to rise, which you can see as you move down the sixth column of Table 6-1. This happens because although each additional worker costs an extra \$80, each additional worker after the second worker brings a smaller increase in output than their predecessor. Each successive \$80 wage payment brings with it fewer and fewer additional bottles produced, so the average variable cost per bottle must rise.

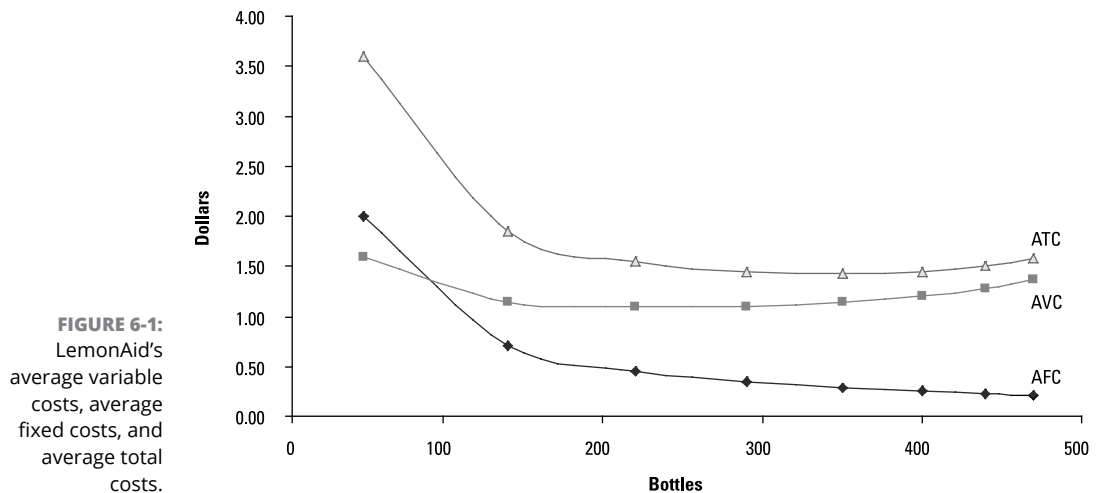
LemonAid Corporation's average variable costs show up as a subtle U-shape when you plot them on a graph, which I do in Figure 6-1. (I also show the company's average fixed costs and average total costs.) Keep this average variable cost curve in mind because it has a massive influence on how many bottles the firm's managers want to produce to maximize firm profits.

Watching average fixed costs fall



REMEMBER

Average fixed costs (AFC) are defined as fixed costs divided by quantity (FC/q). Average fixed costs always decline because the same fixed cost is divided up over a greater and greater number of units of output as output increases.



© John Wiley & Sons, Inc.

The fixed costs of LemonAid Corporation are always the \$100 it paid for the juicer machine, no matter what amount of output it produces. As a result, the more lemonade it produces, the less average fixed costs are. That's why *AFC* falls (see the fourth column of Table 6-1) from a value of \$2.00 per bottle when 50 bottles are produced using one worker down to only \$0.21 per bottle when 470 bottles are produced using eight workers. When you plot out average fixed costs per bottle, as in Figure 6-1, you get a downward-sloping *AFC* curve. Keep this fact in mind because it helps explain the shape of the average total costs (*ATC*) curve, as I explain in the next section.

Tracking the movement of average total costs

Average fixed costs always decline as output increases, while average variable costs first fall (due to increasing returns) and then rise (due to diminishing returns); see the preceding sections for details. Because total costs are the sum of fixed costs and variable costs, *average total costs* obviously depend on how average fixed costs and average variable costs sum up.

Average total costs (*ATC*) are defined as total costs divided by quantity (TC/q). Now look back at Equation (3) earlier in the chapter. If you divide every term in Equation (3) by q , you get the following:

$$\frac{TC}{q} = \frac{FC}{q} + \frac{VC}{q} \quad (4)$$

You can simplify this equation by realizing that average costs equal the fixed and variable costs divided by the quantity, q . In other words, $ATC = \frac{TC}{q}$, $AFC = \frac{FC}{q}$, and $AVC = \frac{VC}{q}$. What you get is

$$ATC = AFC + AVC \quad (5)$$

You can see clearly from this equation that average total costs depend on how average fixed costs and average variable costs interact. There are two key points to understand here:

» **Average total costs (*ATC*) must always be greater than average variable costs (*AVC*) because you have to add in average fixed costs (*AFC*).** Look at Figure 6-1, which shows that the *ATC* curve is above the *AVC* curve. The vertical distance between them at any particular output level equals the *AFC* at that output level. As you move from lower to higher output levels, the *ATC* and *AVC* curves converge because *AFC* becomes smaller and smaller. (In other words, the vertical distance between the *ATC* and *AVC* curves also gets smaller and smaller.)

» **Average total costs (ATC) will reach a minimum value at a higher output level than average variable costs (AVC).** For instance, Table 6-1 shows that average variable costs reach their minimum value of \$1.09 when three workers are hired and 220 bottles are produced. Average total costs, however, reach their minimum of \$1.43 when five workers are hired and 350 bottles are produced.

This happens because average fixed costs are always falling, meaning that in Equation (5), the *AFC* part on the right side of the equation is always shrinking. This constant decline helps to temporarily offset the increases in average variable costs that happen when diminishing returns set in. Consequently, although average variable costs bottom out at three workers, average total costs don't bottom out and start increasing until the fifth worker.

Focusing on marginal costs

The preceding sections demonstrate how the manager of a firm can use data on total costs, fixed costs, and variable costs to calculate, respectively, average total costs, average fixed costs, and average variable costs. But what the manager really wants to know is what quantity, q , of output to produce to maximize profits. To solve this problem, the manager needs one more cost concept: marginal cost.



REMEMBER

Marginal cost is how much total costs increase when you produce one more unit of output. Crucially, the marginal cost of one more unit of output depends on how much output has already been produced.

To see this, examine the total costs column of Table 6-1. Notice that total costs increase from \$100 in the first row to \$180 in the second row as output increases from 0 bottles to 50 bottles when the firm hires the first worker. In other words, costs go up \$80 while output goes up by 50 bottles. So each of these extra, or *marginal*, 50 bottles on average increases costs by $\frac{\$80}{50} = \1.60 each. The marginal cost per bottle, *MC*, is defined as follows:

$$MC = \frac{\text{change in } TC}{\text{change in } q} \quad (6)$$

As you move down the marginal costs column of Table 6-1, you can see that marginal costs first fall and then rise. This is yet another reflection of the fact that LemonAid Corporation's production process exhibits increasing returns followed by diminishing returns. Because the second worker produces much more than the first worker but costs the same, the marginal cost falls when the second worker is added. For successive workers, costs keep increasing but marginal output keeps declining, which means marginal costs must rise.

Noticing where marginal cost equals average cost



REMEMBER

Here's a fun fact that economists love: If you plot out marginal costs to create a marginal cost (MC) curve, that curve will cross both the average variable cost (AVC) curve and average total cost (ATC) curve at their minimum points — that is, at the bottom of their respective U-shapes. (What, you don't see the cause for celebration?)

Look at Figure 6-2, where I plot the AVC , ATC , and MC curves you get by plotting out the data in Table 6-1. The MC curve goes through the minimum points of both the AVC and the ATC curves. This happens because the marginal cost at each unit determines whether the AVC and ATC curves are increasing or decreasing.

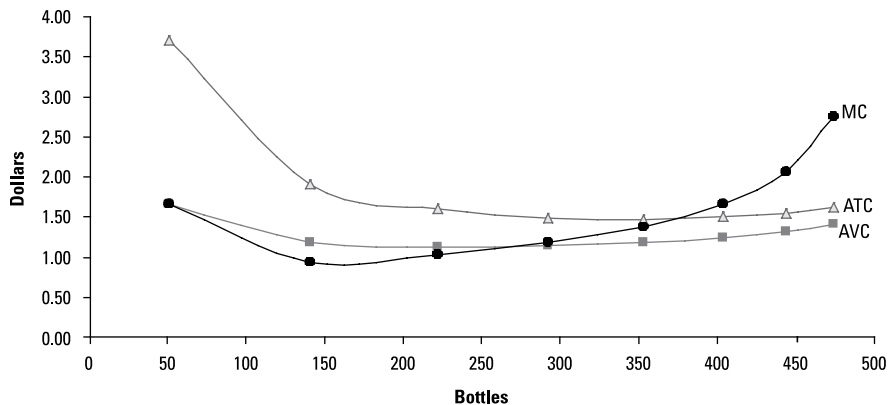


FIGURE 6-2:
LemonAid Corporation's AVC , ATC , and MC curves.

© John Wiley & Sons, Inc.

If that has you scratching your head, consider another example. Instead of thinking about costs, think about heights. Suppose you have a room with ten people in it, and you determine that the average height of the people in the room is 5'6". Now think about what's going to happen to that average if another person walks into the room:

- » If the 11th person is taller than 5'6", the average will rise.
- » If the 11th person is shorter than average, the average will fall.
- » If the 11th person is exactly 5'6" tall, the average will stay the same.

The same sort of reasoning applies to marginal costs and average costs. After q units of output, you can compute AVC and ATC , just like you can compute the average height after the first ten people enter the room. After that, AVC and ATC either

rise or fall depending on the *MC* of the next unit of output, just as the average height of the people in the room increases, decreases, or stays the same depending on the height of the next person entering the room.



TIP

Here's how the marginal cost of the next unit of output will change an average (variable or total) cost:

- » If the *MC* is less than the previous value of an average cost, that average will fall.
- » If the *MC* is greater than the previous value of an average cost, that average will rise.
- » If the *MC* is the same as the previous value of an average cost, that average will stay the same.

You can see these effects graphically by looking at various parts of Figure 6-2. First, look at the output level of 140 bottles. At that output level, the *MC* of producing one more bottle is less than both *ATC* and *AVC*, meaning that *ATC* and *AVC* will decrease if output is increased by one more bottle. That's why the *AVC* and *ATC* curves are downward sloping at that output level. The average curves are being pulled down by the low value of *MC*.

Next, look at the output level of 440 bottles. You can see that the *MC* at that output level is higher than the *ATC* and the *AVC*. Consequently, both *AVC* and *ATC* must be increasing. This increase is reflected geometrically by the upward slopes of both the *AVC* and *ATC* curves. The curves slope upward because the high value for *MC* is pulling them up.

Now put some pieces together. Notice that the *MC* curve causes both the *AVC* and the *ATC* curves to be U-shaped (albeit subtly). On the left side of Figure 6-2, the fact that *MC* is less than the average curves means that the average curves slope downward. On the figure's right side, the fact that *MC* is greater than the average curves means the average curves slope upward.

So you've come full circle to the fact that the *MC* curve has to cross the two average curves at their respective minimum points — at the bottoms of their respective U-shapes. To the left of such a crossing point, the average must be falling because *MC* is less than the average. And to the right, the average must be rising because *MC* is larger than the average. But at any point where the *MC* curve crosses an average curve, the average curve is neither rising nor falling because the *MC* of that unit of output is equal to the current average. (In other words, a 5'6" person has walked into a room that already has a 5'6" average height, so the average doesn't budge.)

Economists love to go on and on (and on) about this fact, but it's really just a reflection of the effect that increasing and then decreasing returns have on cost

curves. Costs first fall and then rise. And there's some point in the middle at which they momentarily stay the same, frozen for an instant while transitioning from falling to rising. That point must be where marginal cost equals average cost because only when MC equals average cost can average cost be stationary.

Comparing Marginal Revenues with Marginal Costs

Here's a sad but true fact: Firms can't always make a profit. In a perfectly competitive industry, that happens because a firm in a perfectly competitive industry can't control the price at which its output sells, and sometimes that price is too low for a firm to make a profit no matter what quantity it produces. When that happens, the best the firm can do is minimize its losses and hope the price changes. If the price drops low enough, the best thing to do may be to shut down production immediately because, that way, the firm will lose only its fixed costs. (I discuss the difference between fixed and variable costs in the earlier section, "Analyzing a Firm's Cost Structure.")

Later in the chapter, I discuss this gloomy situation in more detail. But first, I focus on a happier situation — one in which the market price is high enough that a firm wants to produce a positive amount of output. This may or may not mean that a firm is making a profit, but even if it isn't, its losses aren't large enough to halt production.

Finding where marginal revenue equals marginal cost

In the typical case where market prices are high enough that a firm wants to make a positive amount of output, a straightforward formula is used to determine the optimal quantity of output, q , that the firm should produce. The firm wants to produce at the level of output where marginal revenue equals marginal cost ($MR = MC$).

Producing where $MR = MC$ does two things:

- » It minimizes the firm's loss if it has to take a loss due to a low selling price for its output.
- » It maximizes the firm's profit if it's able to make a profit because the selling price is high enough.



The idea behind $MR = MC$ comes down to a cost-benefit analysis. If producing and selling a bottle brings in more revenue than it costs to make the bottle, then make it. If not, then don't make it. Easy, right?

Imagine that LemonAid Corporation can sell each bottle of lemonade it produces for \$2. Economists like to say that the marginal revenue of each bottle is \$2 because each bottle, when sold, brings in an extra \$2. In terms of algebra, you can say that $p = MR$, where p is the market price that this competitive firm must charge for each unit of output.

The firm's managers must decide how much to produce based on whether any bottle will cost more or less than the \$2 marginal revenue that the firm would get by selling it.



Be careful at this point. You have to remember that the relevant cost that the managers look at is an individual bottle's marginal cost, MC . That's because if they're deciding on making that particular bottle, they need to isolate that bottle's production cost from the costs of all previously produced bottles so they can compare that particular bottle's marginal cost to the revenue that the bottle will bring if it's produced and sold. MC does just that by ignoring all previous bottles and focusing on what the next bottle will cost to make.

If the MC of that bottle is less than \$2, there's a gain to be made by making it, and the managers will choose to make it. On the other hand, if the MC is bigger than \$2, producing the bottle would cause a loss, and the managers would choose not to produce it.

By looking at the MC of every possible bottle (the 1st, the 5th, the 97th, and so on) and comparing it with the marginal revenue the firm could get by selling it, the managers can determine exactly how many bottles to produce. The necessary comparisons can be made by looking at a table of costs, such as Table 6-1, but it's even easier to make the comparisons graphically.

In Figure 6-3, I've drawn in the marginal cost (MC), average variable cost (AVC), and average total cost (ATC) curves for LemonAid Corporation. I've also drawn a horizontal line at \$2, which is the marginal revenue for selling any and all bottles that the firm may choose to produce. I've labeled the line $p = MR = \$2$ to indicate the fact that the selling price of the bottle is \$2, which is also the marginal revenue.

Look at the quantity q^* , which corresponds to where the horizontal $p = MR = \$2$ line crosses the MC curve. As you can see, $q^* = 440$ bottles. This is the level of output that the firm will choose to produce to maximize profits.

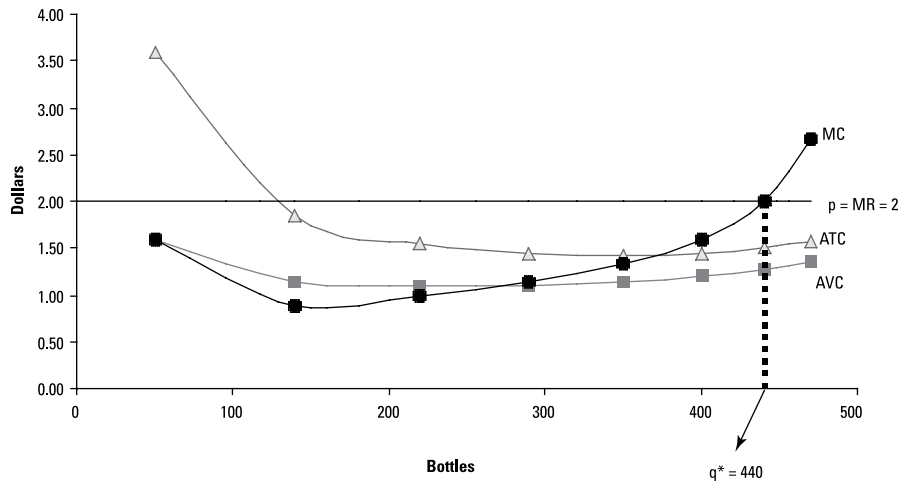


FIGURE 6-3:
The firm's optimal
output level, q^* ,
happens where
 MC and MR cross.

© John Wiley & Sons, Inc.

To understand why adhering to $MR = MC$ maximizes profits, look at Table 6-1 and consider each unit of output, q , for which $q < 440$. For all those units, the marginal revenue is greater than the marginal cost ($MR < MC$), meaning that producing and selling each of those bottles brings in more money than it costs to make. For instance, look at bottle number 140. It has a marginal cost of only \$0.89 but can be sold for \$2.00. Clearly, LemonAid Corporation should make that bottle because the firm makes more selling it than it costs to produce. The same is true for all the bottles for which $q < 440$; the firm should produce all of them because they all bring in more money than they cost to make.

On the other hand, for all units above the q^* level of output ($q > 440$), the situation is reversed: The marginal revenue is less than the marginal cost ($MR < MC$). The firm would lose money if it produced and sold those bottles. For instance, at an output level of 470 bottles, the MC is \$2.67, but the MR is only \$2.00. If the firm produced at that output level, it would lose 67 cents on bottle number 470. Clearly, the company doesn't want to do this.

By comparing the marginal revenues and marginal costs at all output levels, you can see that the managers of LemonAid Corporation want to produce precisely $q^* = 440$ units, the number of units where the MR and MC lines cross.



REMEMBER

Producing where $MR = MC$ doesn't guarantee you a profit, but it does at least make sure that you produce only units that bring in more money than they cost to make. This formula by itself can't guarantee a profit because it doesn't consider the fixed costs you have to pay no matter what output level you're producing. Even though you produce only units for which marginal revenue is at least as great as marginal cost, you still may not make enough of a gain from those units to pay off your fixed costs.

Visualizing profits

There's a quick and easy way to visually use the cost curves to determine whether the firm is making a profit or a loss. The trick is to realize that the two components of profits, total revenue (TR) and total costs (TC), can each be represented by rectangles whose areas are equivalent to their respective sizes. As a result, you can immediately tell whether profits are positive or negative by seeing whether the TR rectangle is larger or smaller than the TC rectangle. If the TR rectangle exceeds the size of the TC rectangle, profits are positive. And if the TR rectangle is smaller than the TC rectangle, profits are negative — the firm is running a loss.

To see how this works, look at Figure 6-4, where I've drawn a generalized set of average total cost (ATC), average variable cost (AVC), and marginal cost (MC) curves, plus a horizontal line labeled $p = MR$ to indicate that price equals marginal revenue for this competitive firm. Switching to this generalized set of curves will (I hope!) convince you that the geometric way of determining the size of a firm's profits holds true for *any* set of cost curves.

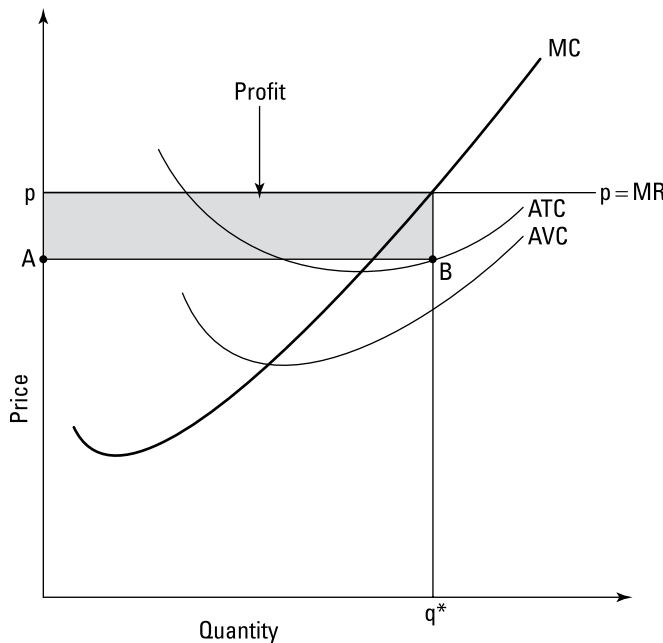


FIGURE 6-4:
A firm making
a profit.

© John Wiley & Sons, Inc.



REMEMBER

The big trick behind expressing total revenue as a rectangular area is to remember that a firm's total revenue when it's producing the profit-maximizing output level, q^* , is simply price times that quantity, or $TR = pq^*$. Just as you can define the area of a rectangular room as length times width, you can define total revenue on

a graph as a rectangle determined by price times quantity. In Figure 6-4, total revenue (TR) is a rectangle of height p and width q^* . Its four corners are located at the origin, at p , at the point where the $p = MR$ line crosses the MC curve, and at q^* .

You can also use a rectangle to represent the total costs that the firm incurs when producing q^* units of output. To determine where to draw this rectangle, you have to use a little math trick to convert the information that the average total cost (ATC) curve gives you into what you want to graph, total costs (TC). To see how to apply this math trick, first look at Point B in Figure 6-4. It shows the average total cost (ATC) *per unit* when the firm produces output level q^* . The reason the trick is handy is because it shows that the rectangle whose width is q^* and whose height is given by the ATC at output level q^* is actually equal to the firm's total costs. In other words, TC is equal to the area of the rectangle whose four corners are the origin, the point I've labeled A, the point I've labeled B, and q^* .



TIP

The heart of the math trick is realizing that when the firm is producing at q^* , $ATC = \frac{TC}{q^*}$. If you multiply both sides of this equation by q^* , you find that $ATC \times q^* = TC$. This equation tells you that TC (total cost) is indeed equal to the product of ATC (the average per-unit cost) and q^* (the number of units), or to the area of a rectangle of height ATC and width q^* — exactly the rectangle that I just showed you!

The firm's profits, which are by definition equal to $TR - TC$, can also be represented by the area of a specific rectangle. In fact, the profit is equal to the area of the shaded rectangle in Figure 6-4. That's because profits are simply the difference between TR and TC . Because the TR rectangle is larger than the TC rectangle in this case, the firm is making a profit whose size is equivalent to the area of the shaded rectangle that's defined by the area of the larger TR rectangle minus the area of the smaller TC rectangle.



TIP

Imagine what would happen if the price, p , increased. First, notice that the optimal output, q^* , would increase because the place where the horizontal $p = MR$ line crosses the MC curve would move up and to the right. Simultaneously, the total revenue rectangle would increase, as would the total cost rectangle. But which one would grow faster? Do profits rise or fall?

Go ahead and draw in some lines to convince yourself that profits will increase — that is, the shaded profit rectangle will grow as the price increases. As you'll discover, a rising price increases the firm's profits.



TIP

You can prove that a rising price increases profits by noting that as q^* shifts right with increases in p , both the width and the height of the profit rectangle swell. The width of the profit rectangle grows with each increase in q^* because the width of the profit rectangle is, by definition, exactly equal to q^* . The height also increases because as you move farther to the right, the distance between the MC curve and

the ATC curve increases because the slope of the MC curve is steeper than the slope of the ATC curve. Thus, with both the width and the height of the profit rectangle increasing simultaneously, the area of the profit rectangle must also be increasing.

Visualizing losses

So what happens if the price falls? This section explains how profits can go negative if the price falls far enough. Consider the situation illustrated in Figure 6-5, where the cost curves are the same as in Figure 6-4, but the price (and therefore the marginal revenue [MR]) at which the firm can sell its product is much lower.

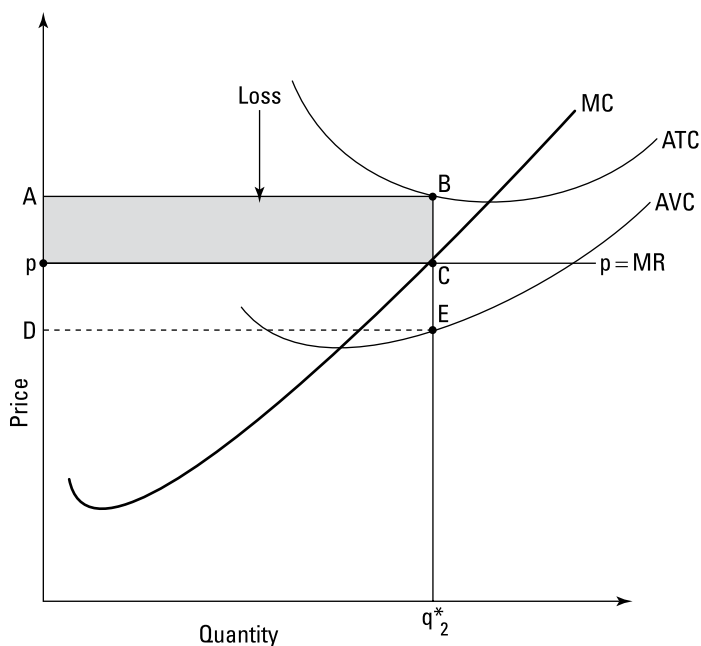


FIGURE 6-5:
A firm running
a loss.

© John Wiley & Sons, Inc.

Following the $MR = MC$ rule for selecting the optimal output level, the firm will choose to produce at the output level q_2^* , where the new lower $p = MR$ line crosses the MC curve. But because of the low price at which the firm is forced to sell its output, it won't be able to make a profit. (I've labeled the optimal output level for the firm in Figure 6-5 as q_2^* to clarify that the optimal output level in this case, where the price is lower, is different from the optimal output level q^* in Figure 6-4, where the price is higher.)

You can see the size of the loss geometrically by comparing the TR and the TC rectangles that occur in this situation. Because $TR = pq_2^*$, total revenue equals the area

of a rectangle of height p and width q_2^* . Consequently, the TR is equal to the area of the rectangle whose four corners lie at the origin, at p , at C , and at q_2^* . It's smaller than the TC rectangle defined by the origin, Point A, Point B, and q_2^* . Because the area of the total cost rectangle exceeds the area of the total revenue rectangle, the firm is running a loss equivalent to the size of the shaded area in Figure 6-5.

Pulling the Plug: When Producing Nothing Is Your Best Bet

You may wonder why a firm would stay in business if it's running a loss rather than a profit. The usual answer is that it hopes that things will turn around soon. Either it expects the price at which it can sell its products to rise, or it expects that it can somehow reduce its production costs.

Notice that those hopes are based on outcomes that aren't necessarily under the firm's control. Indeed, you can best see the firm's general lack of control over its destiny by the fact that, as a competitive firm, it has to take the market price as given. This implies that the firm's decision about whether to continue operating is, in some sense, out of its hands.

To see this, first note that if the price is high enough, the firm will be making a profit and should stay in business to keep collecting the profit. Graphically, this happens whenever the horizontal $p = MR$ line crosses the MC curve at a point above the bottom of the U-shaped ATC curve, as in Figure 6-4. If, however, the horizontal $p = MR$ line crosses the MC curve at a point below the bottom of the U-shaped ATC curve, the firm will be making a loss, as in Figure 6-5. Naturally, the firm won't want to keep operating at a loss and will want to shut down if it expects that low price to persist.

However, the firm may be better off staying in business and continuing to operate at a loss for a while. The existence of fixed costs implies that even if a firm shuts down production, it'll still operate at a loss. (The firm must still pay for its fixed costs, even though it won't be generating any revenue after shutting down production.) Whether it's better to continue operating for the time being or to shut down immediately depends on whether the loss that the firm would incur by continuing to operate is bigger or smaller than the loss that the firm would incur if it shut down immediately.

This section explains how the size of a firm's fixed costs is the key to understanding whether a loss-making firm will want to shut down immediately or continue operating until those fixed-cost commitments expire.

Distinguishing between the short run and the long run in microeconomics



REMEMBER

In microeconomics, the *short run* is the period of time between the present and when a firm's fixed-cost contracts expire. It's when the continuing existence of fixed-cost commitments prevents the firm from being able to terminate its purchases of productive inputs.

For example, consider a firm that has signed a one-year rent contract. Even if the firm is losing money and decides to fire all its employees and sell all its inventory, it will still have to pay rent until the contract expires.

Thus, fixed-cost commitments imply that a loss-making firm can't completely halt all its operations and exit its line of business until all its fixed-cost contracts have expired. Because the length of those fixed-cost commitments varies from firm to firm, there's no set definition of how long the "short run" will last. For any particular firm, it lasts *until* the firm's fixed-cost commitments expire.

Extending that logic, the *long run* begins as soon as the short run ends. The start of the long run is important because it marks the moment when the firm's fixed-cost commitments are completed and the firm is fully able to end all business operations and exit its industry.

The short-run shutdown condition: Variable costs exceed total revenues



REMEMBER

A loss-making firm chooses to shut down production immediately if the size of the loss that it'd make by shutting down immediately is less than the size of the loss that it'd make by continuing in operation and producing output until its fixed-cost contracts expire. Economists call this situation the *short-run shutdown condition*.

Suppose you're in charge of a firm that has a fixed-cost commitment consisting of a monthly rent of \$1,000. That rent commitment implies that even if you produce nothing, you'd sustain a loss of a whopping grand per month until the short run ends and you can get out of your rent commitment. But that potential loss doesn't mean that you should *definitely* start producing stuff to try to make back some of the rent expenditure. You should choose to produce only if doing so results in either an outright profit or a loss of less than the \$1,000 you stand to lose by doing nothing.

Consider Figure 6-6, where the price at which the firm can sell its output is so low that the marginal revenue ($p = MR$) line and the marginal cost (MC) curve intersect at a point *below* the average variable cost (AVC) curve. What does this mean? (Total

revenues are represented by the rectangle whose four corners are at the origin and Points p , B , and q_3^* , where q_3^* represents the optimal output level at this price. Variable costs are represented by the rectangle whose four corners are the origin and Points C , D , and q_3^* .) In this case, the total revenues are less than variable costs.

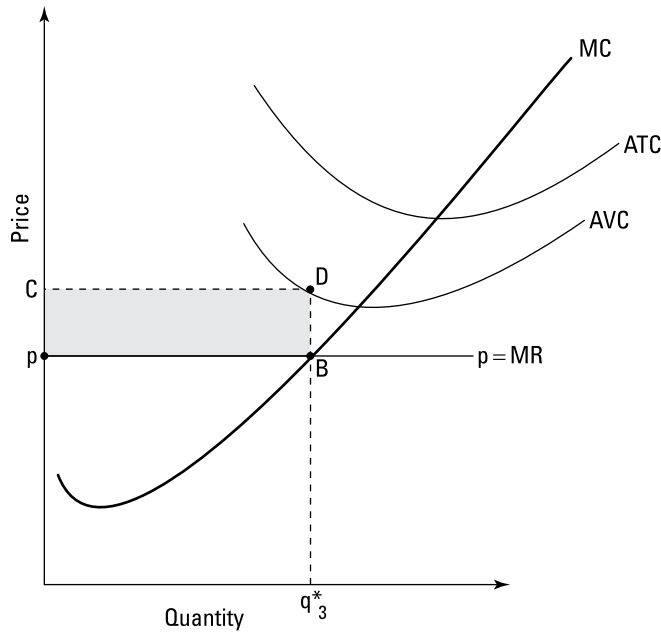


FIGURE 6-6:
A firm running
such a huge loss
that it can't even
cover its variable
costs.

© John Wiley & Sons, Inc.

This means that by producing q_3^* units, the firm doesn't even bring in enough total revenue to cover the variable costs of producing that many units. Not only is the firm going to lose its fixed costs, but it's losing even more money by not being able to cover the variable costs associated with producing q_3^* . The logical thing to do in such a situation is to shut down production immediately and produce nothing. By producing zero units, you lose only your fixed costs. By producing q_3^* , you lose even more money because you can't even cover your variable costs.

As a more concrete example, suppose that fixed costs are \$1,000 and that, by producing q_3^* units, the firm makes total revenues of \$400 and incurs variable costs of \$500. Because total revenues cover only \$400 of the \$500 in variable costs, the firm loses \$100 in variable costs by producing. Add to that the \$1,000 of fixed costs the firm will incur no matter how much it produces, and the firm will lose a total of \$1,100 by producing q_3^* units of output. By contrast, if it shuts down and produces nothing, it loses only the \$1,000 in fixed costs. Given those options, the firm should shut down production immediately (in the short run) instead of waiting until the long run when its fixed-cost commitment expires. Sometimes you've just gotta cut your losses.



REMEMBER

For the general case, the short-run shutdown condition works as follows. If a firm's total revenues at q_2^* are less than variable costs, it's better to shut down immediately (that is, in the short run). Graphically, this happens anytime the horizontal $p = MR$ (marginal revenue) line intersects the MC (marginal cost) curve at a point below the U-shaped AVC (average variable cost) curve. In all such situations, total revenues will be less than variable costs — implying that it's better to shut down immediately than to continue to produce until the firm's fixed-cost commitments have expired in the long run.

The long-run shutdown condition: Total costs exceed total revenues



REMEMBER

The *long-run shutdown condition* occurs when a loss-making firm is better off waiting until its fixed-cost commitments have expired before shutting down production. It's better off waiting until the long run to shut down production if the firm's total revenues currently exceed its variable costs but are less than its total costs. Graphically, this happens in any situation where the horizontal $p = MR$ (marginal revenue) line intersects the MC (marginal costs) curve at any point on the segment of the MC curve that lies above the bottom of the U-shaped AVC (average variable costs) curve but below the bottom of the U-shaped ATC (average total costs) curve, as in Figure 6-5.

In such a situation, the firm is guaranteed to lose money. But as long as the firm is stuck with its current set of fixed cost commitments, the firm is better off producing rather than shutting down immediately. If it produces, its total revenue will exceed its variable costs, meaning that it can use the excess to pay off at least part of its fixed costs. On the other hand, if it shuts down and produces nothing, it'll lose all its fixed costs and thereby do worse.

Look at Figure 6-5. In this case, the firm is more than covering its variable costs because total revenues (represented by the box whose four corners are the origin and Points P , C , and q_2^*) exceed variable costs (represented by the box whose four corners are the origin and Points D , E , and q_2^*). Although this firm is losing money, it's better off producing q_2^* rather than $q = 0$ because total revenues exceed variable costs. The firm can take the extra money left over after paying variable costs and use it to pay off some of its fixed costs.

As a more concrete example, suppose that the firm's fixed costs are \$1,000 and that when producing output level q_2^* , it has a total revenue of \$800 and variable costs of \$700. The first \$700 of the \$800 in total revenues can go to paying off the variable costs, leaving \$100 to pay off a portion of the \$1,000 in fixed costs. The result is an overall loss of \$900 rather than a \$1,000 loss if the firm produces nothing. Thus, this firm should continue operating in the short run and wait until the long run — when its fixed-cost commitments have ended — before shutting down production and exiting its industry.

IN THIS CHAPTER

- » Measuring the social benefits of different output levels
- » Demonstrating that free markets maximize total surplus
- » Reducing total surplus with taxes and price controls
- » Producing at the lowest possible cost to society
- » Adjusting to changes in supply and demand

Chapter 7

Why Economists Love Free Markets and Competition

Economists love *competitive free markets* — markets in which numerous buyers freely interact with numerous competitive firms. Indeed, economists firmly believe that when competitive free markets work properly, they're the best way to convert society's limited resources into the goods and services that people want to buy. (Please note that, for brevity, I sometimes just say *free markets* or *markets* in this chapter rather than writing *competitive free markets* each time. I'm trying to maximize my resources here.)

Why do economists place such great confidence in competitive free markets? Because the interaction of supply and demand (which I discuss in Chapter 4) leads

to an outcome in which every unit of output that's produced satisfies two excellent conditions:

- » It's produced at the minimum cost possible, meaning that there's no waste or inefficiency.
- » Its benefits are always at least as big as its costs. Competitive free markets produce only output that makes the world better off or at least no worse off.

Economists also love competitive free markets because they provide a gold standard against which all other economic institutions can be judged. In fact, economists refer to many economic problems as *market failures* precisely because they're instances where, if markets could function properly, the problems would quickly go away.

In this chapter, I show you that competitive free markets ensure that benefits exceed costs for all the output produced. I also show you that competitive free markets produce the *socially optimal quantity* of output — the level that maximizes the benefits that society can get from its limited supply of resources. Finally, I show you how competitive industries adjust to changes in supply and demand to ensure that everything is produced at the lowest possible cost to society.

Ensuring That Benefits Exceed Costs: Competitive Free Markets

Society has only a limited amount of land, labor, and capital to make things with. Consequently, society must pay attention when figuring out how to best convert its limited resources into the goods and services that people most greatly desire.



REMEMBER

Economists love competitive free markets because, if they're operating properly, they make sure that resources are allocated optimally. In particular, such markets ensure that resources go toward producing only output for which the benefits exceed the costs.

This point can be easily demonstrated using nothing more complicated than a supply and demand graph (such as the type I introduce in Chapter 4). But first, this section explains the conditions under which competitive free markets can function properly and thereby deliver such ideal results.

Examining the traits of a properly functioning market



REMEMBER

Free markets guarantee optimal outcomes if these six conditions are met:

- 1. Buyers and sellers have access to the same full and complete information about the good or service in question.** This guarantees that both parties will be willing to negotiate without having to worry that the other has some secret information. (In Chapter 11, I explain how markets break down if one side or the other has more information.)
- 2. Property rights are set up so that the only way buyers can get the good or service in question is by paying sellers for it.** This ensures that sellers have an incentive to produce output. As a counterexample, consider trying to sell tickets to an outdoor fireworks display: Because everyone knows that they can see the display for free, nobody wants to pay for a ticket. And with nobody willing to pay for a ticket, producers have no incentive to put on a display. (In Chapter 11, I discuss situations like these and how society must deal with them given that markets can't.)
- 3. Supply curves capture all the production costs associated with making the good or service in question.** This requirement helps ensure that markets can make the proper cost-benefit calculations. For instance, if a steel factory can pollute for free, there's no way that the price of steel will incorporate the damage that the factory's pollution does to the environment. On the other hand, if the government forces the factory to continuously pay for cleanup costs, these costs will be reflected in the market price, thereby allowing society to properly weigh the costs and benefits of the company's output. (Chapter 10 deals with ways to help markets along if supply and demand curves don't reflect all costs and benefits.)
- 4. Demand curves capture all the benefits derived from the good or service in question.** This requirement also ensures proper cost-benefit analysis. If these first four conditions for free markets are met, market forces can reach a social optimum — but only if they're free from interference. That's why we need two more conditions: one that limits buyers and sellers, and another that constrains government intervention.
- 5. There are both numerous buyers and numerous sellers, such that nobody is big enough to affect the market price.** This is often called the *price-taking assumption* because everybody just has to take prices as given. This requirement eliminates problems such as monopolies, in which individual buyers or sellers are so powerful that they can manipulate the market price in their own favor.

- 6. The market price is completely free to adjust to equalize supply and demand for the good or service in question.** The sixth requirement stipulates that supply and demand must be allowed to freely determine the market price and market quantity unimpeded by government-imposed price ceilings or floors. (In Chapter 4, I explain the problems with price ceilings and floors and discuss how they hurt society.)

Basically, these six conditions accomplish two broad goals: They guarantee that people will want to buy and sell in a market environment, and they ensure that markets will consider all the costs and all the benefits of producing and then consuming a given amount of output.

Analyzing the efficiency of free markets

Economists use supply and demand curves to demonstrate that free markets produce socially optimal levels of output. But the simple insight behind this result is that a unit of output can be socially beneficial to produce and consume only if the benefits that people derive from consuming it exceed the costs of producing it.

This simple idea is, in fact, why demand curves and supply curves are so useful in analyzing the social optimum. Demand curves quantify the benefits that people get from consumption by showing what they'd be willing to pay to consume every particular unit of output. (See Chapter 4 for details.) Similarly, supply curves quantify the cost of producing each particular unit of output. (See Chapter 6.)

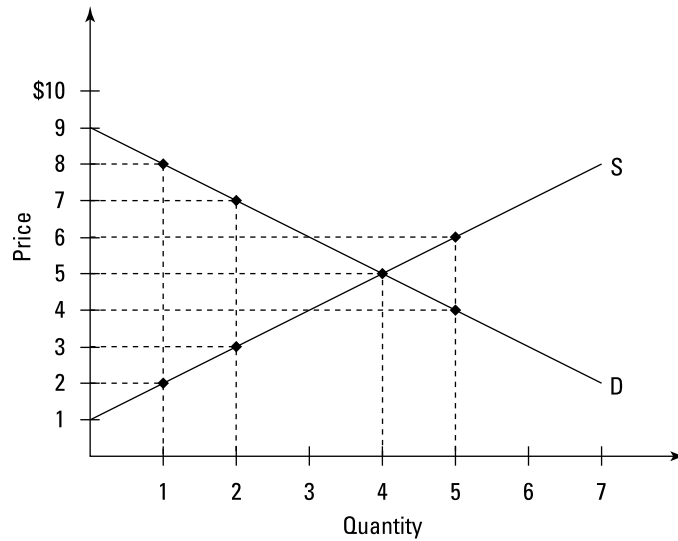
Using supply and demand to compare costs and benefits

By drawing the demand and supply curves for a good or service on the same graph, you can easily compare the benefits and costs of producing each unit of output. To see how, take a look at Figure 7-1, where I've drawn a demand curve, *D*, and a supply curve, *S*.

To start, look at one unit of output on the horizontal axis. At that output level, go up to the demand curve and see that people are willing to pay \$8 for one unit of output. At the same time, by going up to the supply curve, you can see that firms are willing to supply one unit at a cost of \$2.

So you can see that it's socially beneficial to produce this first unit of output because it's worth more to buyers (\$8) than it costs sellers to produce (\$2). Put slightly differently, although the resources that it takes to make that unit of output cost society only \$2, they bring \$8 in benefits when they're converted into this particular good or service. Because the benefits exceed the costs, this is a unit of output that *should be* produced.

FIGURE 7-1:
Comparing costs
and benefits
using supply and
demand curves.



© John Wiley & Sons, Inc.

Now look at the second unit of output. Going up to the demand curve tells you that people are willing to pay \$7 for that unit, while going up to the supply curve tells you that the second unit costs \$3 to produce. Again, benefits exceed costs. Again, this unit of output *should be* produced.

By contrast, look at the fifth unit of output. The costs as given by the supply curve for producing the fifth unit are \$6, but the benefits as given by the demand curve are only \$4. Because the costs of producing this unit exceed what anyone is willing to pay for it, this is a unit of output that *shouldn't be* produced.

In other words, producing the fifth unit of output would destroy value. Why? Because making it involves converting \$6 worth of resources into something that's worth only \$4 to consumers. Producing it would make the world worse off. Better to redirect the resources that would be necessary to make a fifth unit of this product toward the production of something that will generate net benefits to consumers.

Determining the socially optimal output level

A graph showing both the supply curve and the demand curve can tell you precisely what quantity (q) of output should be produced. That's because the supply and demand curves let you quickly compare costs and benefits for every possible output level.

There are only three cost-benefit relationships:

- » **Benefits exceed costs.** In Figure 7-1, this occurs for every bit of output such that $q < 4$.
- » **Benefits equal costs.** This occurs at exactly $q = 4$ units in the graph.
- » **Costs exceed benefits.** This occurs for all output levels where $q > 4$.

Economists look at this and conclude that the socially optimal level of output to produce is $q = 4$ units because for these units, benefits either exceed costs or at least are equal to costs. By producing the first four units of output, society either gains or at least isn't made any worse off.



TIP

The socially optimal output level is always incredibly easy to identify on any supply and demand graph: It's just the quantity produced where the demand and supply curves cross.



REMEMBER

Economist Adam Smith's big insight was to realize that competitive free markets produce exactly the socially optimal output level on their own without anyone having to direct them to do the right thing. This insight was the basis of Smith's metaphor of an invisible hand that seems to guide markets toward the socially optimal output level, despite the fact that each individual in the market may well be looking out only for their own interests. This result greatly simplifies life because it eliminates the need to have a government official or any other sort of central planner constantly checking whether the right amount of output is being produced.

The proof of Smith's assertion is almost trivial. All you have to do is look at Figure 7-1 and realize that the socially optimal output level (determined, as we just said, by where the demand and supply curves cross) is identical to the *market equilibrium quantity* that will occur if this product is produced and sold in a competitive market. That is true because in a competitive market, the market equilibrium quantity is also determined by where the demand and supply curves cross. (To understand why, see Chapter 8.) Thus, the market equilibrium quantity is exactly equal to the number of units you'd want to produce if you were using the demand and supply curves to compare benefits and costs.

Nobody planned it that way, but because the intersection of demand and supply defines both the quantity that will be produced by a competitive market as well as the quantity for which benefits at least equal costs, competitive markets will produce the socially optimal output level (assuming that all six conditions noted earlier are satisfied).

Measuring everyone's gains with total surplus

Economists use a concept called *total surplus* to add up the gains that come from producing the socially optimal output level. The gain, or *surplus*, comes from the fact that benefits exceed costs for the units of output that are produced.

The total surplus turns out to be divided between consumers and producers. The part of the total surplus that goes to consumers is called *consumer surplus*, and the part that goes to producers is called *producer surplus*. (Clever names, eh?)

In the subsections that follow, I tackle consumer surplus first and then move on to producer surplus. After I explain each separately, I add them together to explain total surplus. (And I hope that when you're done with this section, you feel like you've received at least a *little* consumer surplus.)

Measuring the consumer surplus of a discrete good

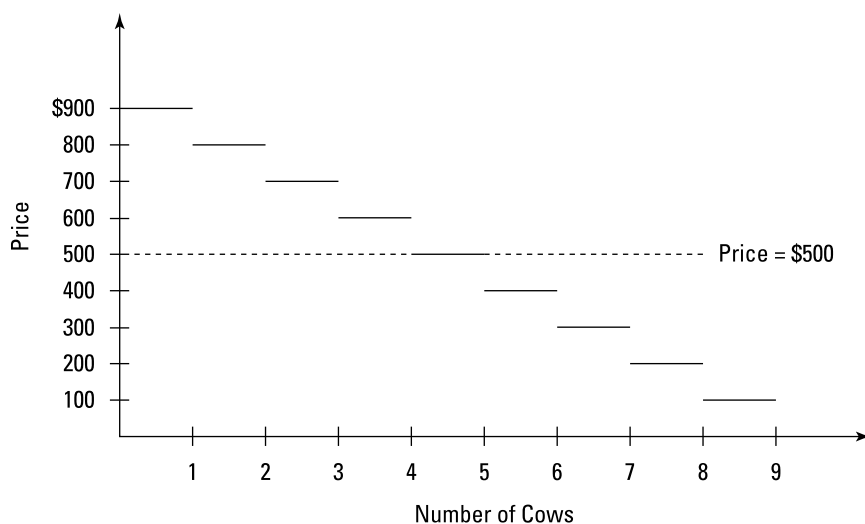
Consumer surplus is the gain people receive when they can buy things at a price that's lower than what they're willing to pay. The easiest way to understand consumer surplus is by looking at a discrete good. A *discrete good* is a good that comes only in whole units. For instance, you can buy 1 car or 57 cars, but you can't buy 2.33 cars. You can purchase 1 horse or 13 cows but not fractional amounts of livestock — at least if you want them alive!

Look at Figure 7-2, which shows the demand for cows. Because cows come in discrete units, you don't get a smooth, downward-sloping curve. Rather, you get what mathematicians call a *step function*. The way to understand it is that people are willing to pay \$900 for the first cow, \$800 for the second cow, \$700 for the third cow, and so on.

Now imagine that the market price of cows is \$500, which is why I've drawn a horizontal dotted line at that price. Compare that price with what people are willing to pay for each cow.

For the first cow, people are willing to pay \$900. Because the market price of cows is only \$500, these buyers come out ahead because they're able to purchase a cow for \$400 less than they were willing to pay. Or, as economists like to say, the *consumer surplus* on the first cow is \$400.

FIGURE 7-2:
The demand
curve for a good
that is sold in
discrete units
looks like
stair steps.



© John Wiley & Sons, Inc.

Next, look at the second cow. People are willing to pay \$800 for it, but because the market price is only \$500, they receive a consumer surplus for that cow of \$300. Similarly, for the third cow, people get a consumer surplus of \$200 because they're willing to pay \$700 for it but have to pay only the market price of \$500. You see where this is going, right?

For the first four cows, there's a positive consumer surplus, but on the fifth cow, people just break even because they're willing to pay \$500 and the cow costs \$500. This means that people will want to buy only five cows. (Economists always assume that when the price exactly equals your willingness to pay, you go ahead and buy.)

To calculate consumer surplus for a discrete good such as cows, you need to total the surpluses that people get on each unit that they choose to buy. In this case, the total is \$1,000 (\$400 for the first cow, plus \$300 for the second cow, plus \$200 for the third cow, plus \$100 for the fourth cow, plus \$0 for the fifth cow).

I show this \$1,000 of consumer surplus in the graph in Figure 7-3 by shading in the area below each step and above the horizontal price line at \$500. The staircase-shaped area equals \$1,000.

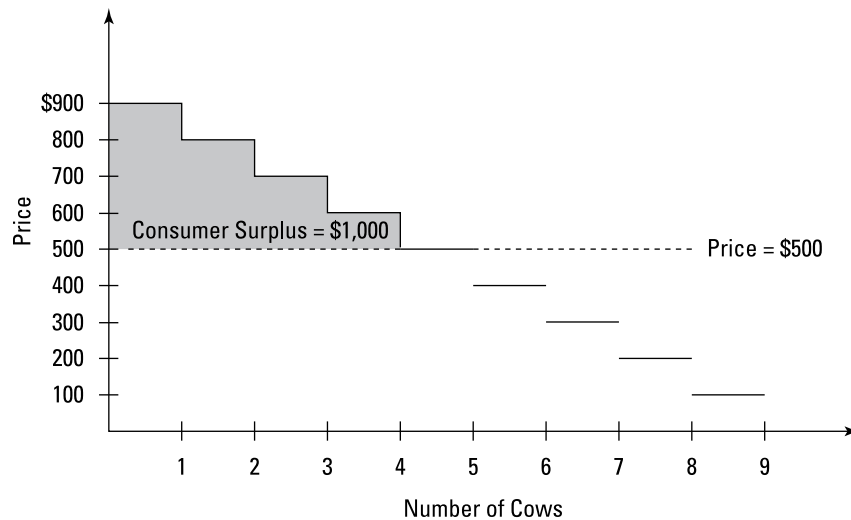


FIGURE 7-3:
Consumer
surplus for a
discrete good.

© John Wiley & Sons, Inc.

Measuring the consumer surplus of a continuous good

Consumer surplus can be computed for continuously measured goods and services — things like land or cooking oil or hours of music lessons, which aren't necessarily sold in discrete units. In other words, you can buy fractional amounts of continuously measured goods, such as 78.5 acres of land, 6.33 gallons of cooking oil, or 2.5 hours of music lessons.

The demand curves for continuously measured goods are much nicer than the step functions you get for discretely measured goods. In fact, the demand curves for continuously measured goods are smooth, downward-sloping lines (such as the ones in Chapter 4).

Because of the smoothness of such demand curves, when you graph consumer surplus for a continuously measured good, you get a triangular area that lies below the demand curve and above the market price. You can see this wedge illustrated in Figure 7-4, which depicts the cooking oil market.

In Figure 7-4, the price of cooking oil is \$5 per gallon. At that price, people want to buy 1,000 gallons of cooking oil. For each of those 1,000 gallons, the demand curve lies above the horizontal, \$5 price line. That implies that buyers are made better off by buying those 1,000 gallons because each of those 1,000 gallons is worth more to its buyer than the \$5 per gallon that it costs to buy it.

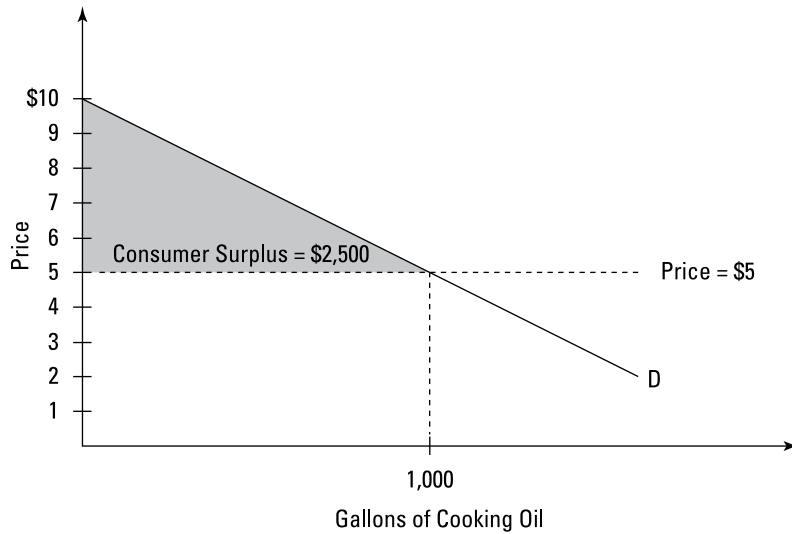


FIGURE 7-4:
Consumer
surplus for a
continuous good.

© John Wiley & Sons, Inc.



TIP

To calculate consumer surplus for a continuous good, you total up all the gains that people receive when buying for less money than they're willing to pay — just as you would do for a discrete good. But because you're now dealing with a triangle, totaling up requires a bit of geometry. Simply use the formula for the area of a triangle ($\frac{1}{2} \times \text{base} \times \text{height}$) to find the total surplus. In this case, you multiply $\frac{1}{2} \times 1,000 \times 5 = \$2,500$.

Measuring producer surplus

Producer surplus measures the gain that firms receive when they can sell their output for more than the minimum price that they would've been willing to accept. You can calculate producer surplus for both discrete and continuous goods, just as you can calculate consumer surplus for each. In this section, I offer an example of calculating producer surplus for a continuous good.

You can get a good handle on producer surplus by looking at Figure 7-5, which shows the supply curve, *S*, for cooking oil. This supply curve is crucial for determining producer surplus because each point on the supply curve tells you the minimum that you would have to pay suppliers for them to give you the associated amount of output. By comparing each minimum value with the higher market price that firms actually receive when they sell their output, you can compute producer surplus by finding the area of a triangle. (For more on supply curves and how to interpret them, see Chapter 4.)

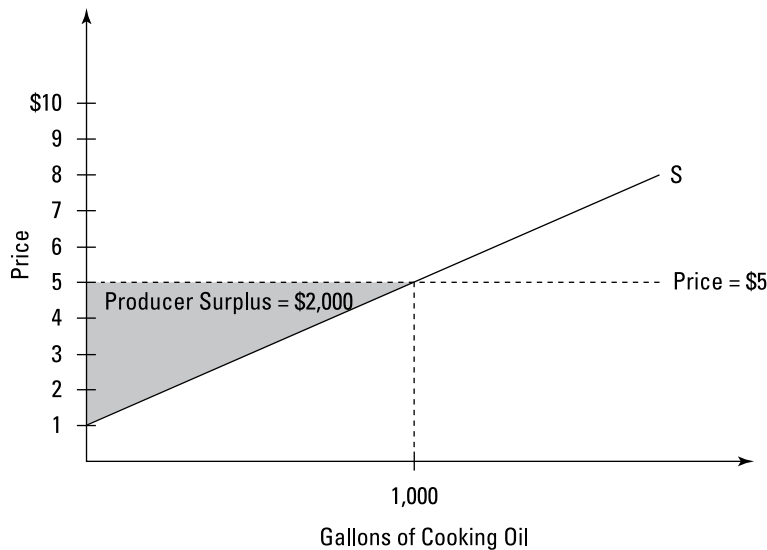


FIGURE 7-5:
Producer
surplus for a
continuous good.

© John Wiley & Sons, Inc.

Suppose the price of cooking oil is \$5 per gallon. You can see in Figure 7-5 that producers are going to want to supply exactly 1,000 gallons of cooking oil at that price. They want to supply that much because for each drop of oil up to and including the very last drop of the 1,000th gallon, the production costs as given by the supply curve are less than the \$5 per gallon that producers get when they sell the oil.

But crucially, producers are willing to supply almost all that cooking oil for *less than* the \$5 per gallon market price. You can see this by the fact that the supply curve lies below the horizontal price line up to the very last drop of the 1,000th gallon. The fact that they receive \$5 per gallon for all of it despite being willing to produce it for less is the source of the producer surplus, which is represented by the area of the shaded triangle.



TIP

Using the formula for the area of a triangle ($\frac{1}{2} \times \text{base} \times \text{height}$), you can compute that the producer surplus in this example is \$2,000. Producers are \$2,000 better off after selling the 1,000 gallons of oil because the total cash they get from selling the 1,000 gallons is \$2,000 greater than the minimum amount that they'd have been willing to accept to produce those units.

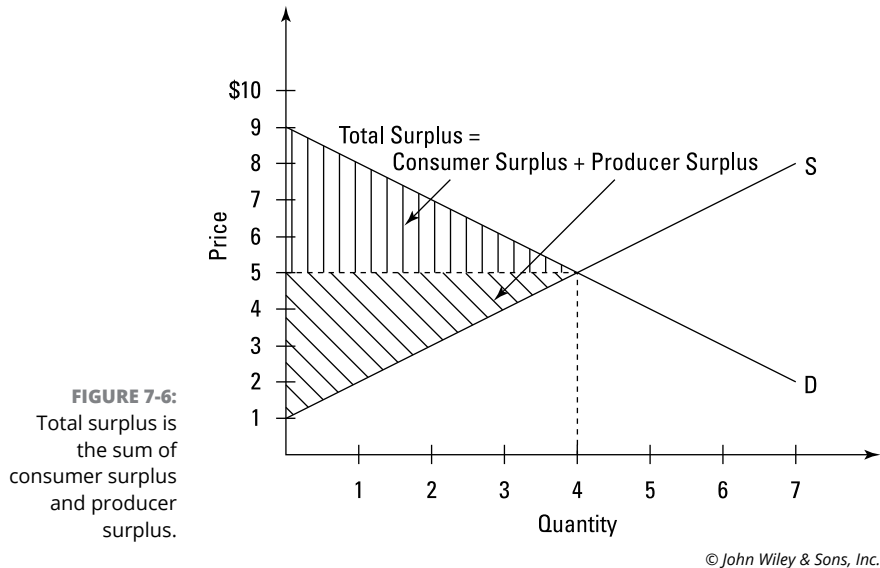
Computing total surplus



REMEMBER

The *total surplus* that society receives from producing the socially optimal level of output of a certain good or service is simply the sum of the consumer surplus and producer surplus generated by that output level.

Figure 7-6 illustrates total surplus for a market in which the equilibrium price and quantity are, respectively, $p^* = \$5$ and $q^* = 4$. (If this graph looks familiar, that's because it's just like Figure 7-1.)



I've drawn the total surplus area so you can clearly see that it's made up of consumer surplus (the vertically striped area) plus producer surplus (the diagonally striped area). The two are separated by the horizontal line extending from the market equilibrium price (\$5).

By again using the formula for the area of a triangle, you multiply $\frac{1}{2} \times 4 \times 8$ to figure out that for this graph the total surplus is \$16. The total gain to society of producing at this output level is \$16.

Contemplating total surplus

Total surplus is important because it puts a number on the gains that come from production and trade. Firms make things to generate a profit. People spend money on products because consuming those products makes them happy. And total surplus tells you just how much better off both consumers and producers are after interacting with each other.



REMEMBER

By putting a number on the gains made by their interaction, total surplus also provides a benchmark by which economists can measure the harm that comes from government policies that interfere with the market. It's one thing to say that, for instance, price subsidies hurt consumers. It's another thing to be able to say by exactly how many dollars consumers are harmed.

When Free Markets Lose Their Freedom: Dealing with Deadweight Losses

Anything that interferes with the market's ability to reach the market equilibrium and produce the market quantity reduces total surplus. Economists refer to the amount by which total surplus is reduced using the colorful term *deadweight loss*.

In this section, I give you detailed examples of deadweight losses caused by price ceilings and taxes. These types of market interference are under the government's control, but you shouldn't think that deadweight losses are caused *only* by government policy. Anything that reduces output below the market quantity causes a deadweight loss. Monopolies and oligopolies can be to blame, as can asymmetric information and public goods problems — all things that I discuss in the next few chapters.

Coming up short: The deadweight loss from a price ceiling

Price ceilings are maximum prices at which sellers can legally sell their product. (See Chapter 4 for details.) Generally, price ceilings are intended to help buyers obtain a low price, but they cause a lot of harm. As an example of a deadweight loss, look at Figure 7-7, in which the government has imposed a price ceiling at P^C .

To see the damage that price ceilings inflict, first notice that at a maximum price of P^C , suppliers are going to want to sell only q^L units of output. (The L stands for *low*.) In other words, at that price, only the first q^L units of output are profitable to produce. By contrast, if no price ceiling existed and the market were left to its own devices, suppliers would choose to produce the market equilibrium quantity of output, q^* .

Consequently, if this were a free market, the total surplus would be represented graphically by the triangle defined by Points A, B, and C. But because only q^L units of output can be produced, the total surplus area is reduced to the shaded area with corners at A, B, F, and E.

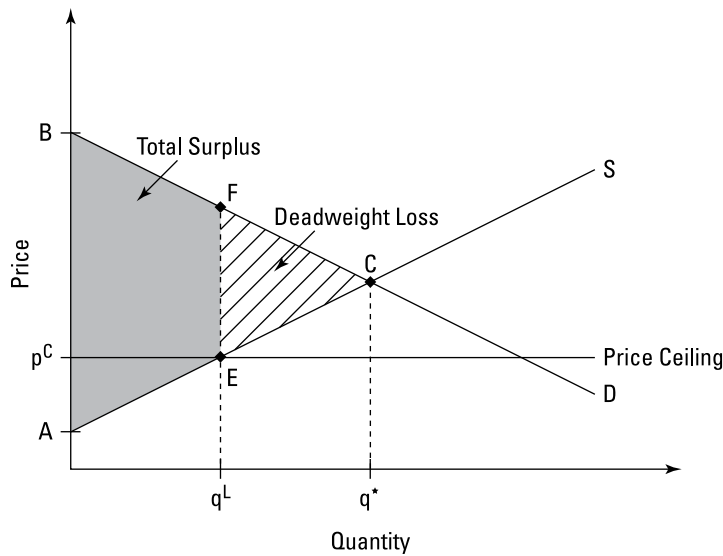


FIGURE 7-7:
Reduced total
surplus caused by
a price ceiling.

© John Wiley & Sons, Inc.

The difference between the total surplus generated by producing q^* versus q^L units of output is the diagonally striped triangle defined by Points E, F, and C. The area of this triangle illustrates the deadweight loss that comes from reducing output below the socially optimal level, q^* .



REMEMBER

The price ceiling is harmful because for all units between q^L and q^* , benefits exceed costs, meaning that such units should be produced. By tallying up the gains that should've come from producing and consuming these units, the deadweight loss triangle can precisely measure the harm that results from interfering with the market.

Death and taxes: Finding the deadweight loss of a tax

Taxes on goods and services can also cause deadweight losses. This happens because such taxes raise the costs of producing and consuming output. When these costs are artificially raised by the tax, people respond by producing and consuming fewer units of output than they did before the tax was imposed. Because each unit that had been consumed before the tax was imposed was a unit for which benefits had exceeded costs, the reduction in output that results from the tax necessarily reduces total surplus and causes a deadweight loss.

Seeing how taxes shift the supply curve

Taxes on production shift supply curves upward. Why's that? Well, to the firms that are being taxed, taxes on production feel like increases in production costs. The firms know that the only way they can at least break even is if they pass the cost of the tax on to buyers — just as they know that they must pass on to buyers the costs of any labor, capital, or other resources. So just as any increase in input costs shifts supply curves vertically, so does any production tax. Imposing a tax on sellers shifts the supply curve vertically by the amount of the tax.

Consider a concrete example — the supply of beef in a beef market in which the government is going to impose a tax of \$1 per pound. Figure 7–8 shows two curves. (Actually, they're straight lines, but humor me here.) The lower one, S , is the supply curve for beef. The higher one, labeled $S + tax$, is the supply curve after the tax is imposed. The important thing to realize is that the curve $S + tax$ is simply the original supply curve shifted up vertically by the amount of the tax, which in this case is \$1.

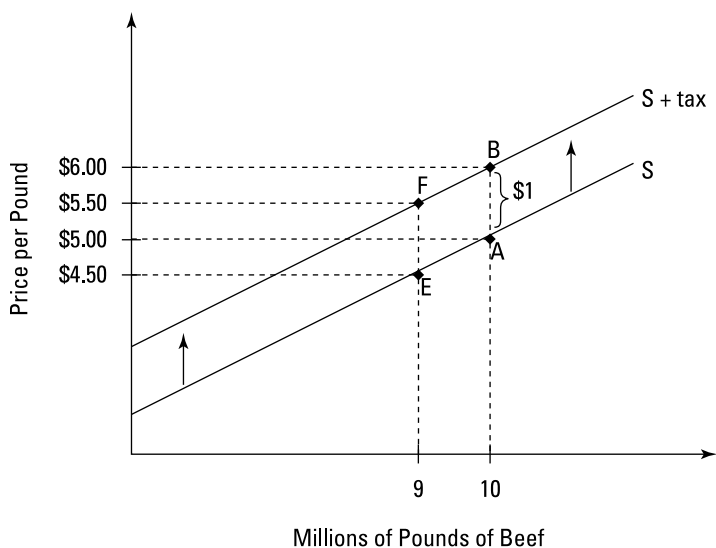


FIGURE 7-8: Imposing a \$1 tax on beef shifts the supply curve vertically by \$1, from S to $S + tax$.

© John Wiley & Sons, Inc.

The supply curve shifts up vertically by the amount of the tax because of motivating suppliers. Each point on the supply curve tells you the minimum amount that suppliers must receive to get them to supply a particular quantity. (See Chapter 4 for details.) For instance, look at Point A. Because Point A is on the supply curve, you know that suppliers must receive \$5 per pound if you want them to provide 10 million pounds of beef. Similarly, Point E tells you that suppliers must receive \$4.50 per pound if you want them to supply 9 million pounds of beef.

If the government comes in and imposes a tax of \$1 per pound, it affects how much consumers must pay to get any given amount produced. In particular, there has to be enough money left over after paying the tax to motivate the suppliers. For instance, if consumers still want 10 million pounds of beef, they'll have to pay the original amount required to motivate the suppliers to produce that much beef (\$5 per pound for each of the 10 million pounds of beef), as well as enough money to pay the taxes on that much beef (\$1 per pound for each of the 10 million pounds of beef). Thus, to get 10 million pounds of beef produced, the selling price would have to be \$6 per pound (= the original \$5 per pound needed to motivate the producers plus \$1 per pound to cover the taxes).

Graphically, this means that Point A on supply curve S shifts up by the \$1 amount of the tax to become Point B on the $S + \text{tax}$ curve. For the same motivational reasons, Point E on the supply curve must shift up to Point F on the $S + \text{tax}$ curve. That is, if you have to pay suppliers \$4.50 per pound to motivate them to give you 9 million pounds of beef in a world in which \$1 per pound must go to the government in taxes, you have to collect a total of \$5.50 per pound (= \$4.50 per pound to motivate the suppliers and \$1 per pound to pay the tax man). And that's exactly what happens at Point F.

Every point on the supply curve, S , must shift up vertically in the same way that Points A and E do, so the $S + \text{tax}$ curve captures what the supply curve looks like after the tax is imposed. With this shift in mind, you're ready to discover how this sort of taxation causes deadweight losses.

Seeing how taxes cause deadweight losses

Deadweight losses are losses in the sense of annihilation. You can't say, "Your loss is my gain" because you aren't talking about a situation in which something passes from one person to another. Some of the gains that would have resulted if the full market quantity had been produced simply vanish; their loss is a "dead weight" that society is forced to bear as output declines to an amount that is less than the socially optimal level that would have occurred in the absence of the tax.

Figure 7-9 adds a demand curve, D , to Figure 7-8 so you can see the deadweight loss that results when the government imposes a \$1 per pound tax on the beef that's sold in the beef market.

Before the tax, the market equilibrium happens at Point A, where supply curve S crosses demand curve D . At that point, producers supply 10 million pounds of beef at a price of \$5 per pound. The total surplus in this case is given by the triangle defined by Points C, D, and A.

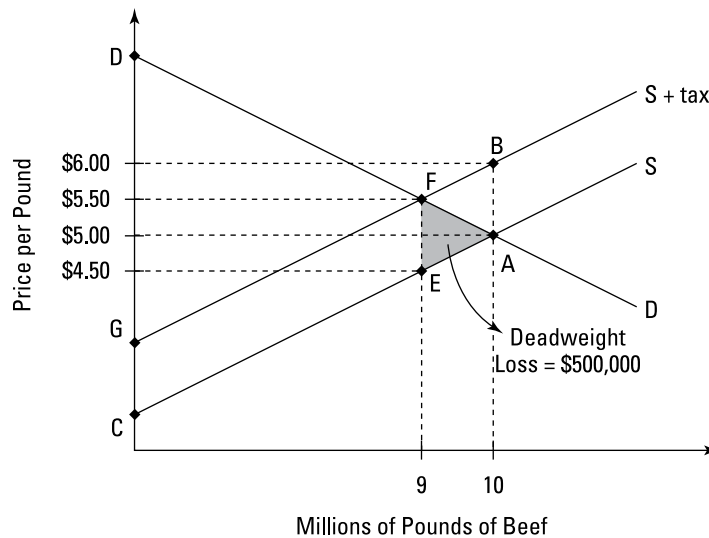


FIGURE 7-9:
The deadweight
loss caused by a
\$1 tax on beef.

© John Wiley & Sons, Inc.

After the tax is imposed, however, the equilibrium happens at Point F, where the $S + \text{tax}$ curve crosses the demand curve. At that point, the price of beef is \$5.50 per pound, and 9 million pounds are supplied. (Only 9 million pounds are supplied because after the government takes its \$1 per pound in taxes, only \$4.50 remains to motivate suppliers. You can see from the supply curve S that at that much money per pound, suppliers want to supply only 9 million pounds.)

Because of the tax, the amount of beef supplied falls from 10 million pounds to 9 million pounds. Furthermore, the total surplus is reduced to the triangle whose three corners are G , D , and F .



TIP

You can immediately see that this new total surplus is much smaller than the old one. But before you start ranting about the evils of government, you need to take account of the fact that taxes are being collected. Taxes (theoretically, at least) benefit society, so you need to include this amount when calculating the total surplus of this good sold at this price. At the new equilibrium, \$9 million in taxes will be collected because the 9 million pounds of beef sold will be taxed at \$1 each.

The \$9 million in tax collections are represented graphically by the parallelogram whose corners are C , G , F , and E . This area was previously contained in the old total surplus triangle whose corners were C , D , and A . Consequently, this area that used to be part of the old total surplus hasn't been destroyed; it's merely been transferred to the government.

However, part of the old total surplus *has* been destroyed. This part is shown graphically by the shaded deadweight loss triangle (with corners at E, F, and A). This area captures the fact that society has been made worse off by the reduction in beef output from 10 million pounds to 9 million pounds. (Okay, *now* you can start ranting about the evils of government.)

Measuring the size of the deadweight loss using the formula for the area of a triangle ($\frac{1}{2} \times \text{base} \times \text{height}$) tells you that the tax leads to a deadweight loss of \$500,000. That's a big number representing a huge reduction in total surplus deriving from the fact that for each of the 1 million pounds of beef that are no longer being produced, benefits had exceeded costs. All those gains are lost when the tax is imposed.

Note: In certain special cases, a tax can reduce (rather than cause) a deadweight loss. The intuition is that if a product is being inefficiently *overproduced*, then imposing a tax can make things better: By shifting the supply curve upward, the tax causes production to decrease, thereby reducing the overproduction. I discuss these special cases in detail in Chapter 10.

Hallmarks of Perfect Competition: Zero Profits and Lowest Possible Costs

A wonderful thing about free markets and competition is that output is produced at the lowest possible cost. This fact is extremely important because it means that free markets are as efficient as possible at converting resources into the goods and services that people want to buy.

In addition, markets save society a lot of money because they produce efficiently without requiring human intervention. People don't have to pay big salaries to experts to make sure that markets run efficiently; markets do the job for free.

Understanding the causes and consequences of perfect competition



REMEMBER

To ensure that markets function efficiently, you need really strong competition among firms, a situation that economists refer to as *perfect competition*. Perfect competition exists when there are many firms within a given industry, and all those firms are producing identical or nearly identical products. (See Chapter 6 for details.) The following things are also true when perfect competition exists:

- » Every firm is a *price taker* (meaning that the firm must accept the market equilibrium price for what it produces) because its output is a small fraction of the industry's total output. (See Chapter 6.)
- » Every firm has identical production technology.
- » Firms are free to enter or leave the industry as they please.



REMEMBER

When these requirements are met, perfect competition leads to two excellent outcomes:

- » **Every firm in the industry makes zero economic profits.** As you may recall from Chapter 6, a firm's *economic profit* is the amount of money that flows to a firm's entrepreneurs after they account not only for the costs of land, labor, and capital but for the opportunity cost of their entrepreneurial ability in its most lucrative alternative use. Under that definition, economic profit is only positive if the entrepreneurs are earning more at their current firm than they could earn anywhere else. Conversely, a negative economic profit implies that they could be earning more somewhere else.

So the fact that perfect competition leads to zero economic profits means that the entrepreneurs working at competitive firms just barely want to keep their current firms operating rather than trying their luck at starting a firm in another industry.

It also means that nobody in the industry is getting filthy rich at anyone else's expense. Rather, they're doing just well enough to supply the output that society wants them to. This situation is great for society because it'd be wasteful to pay entrepreneurs more than necessary to get them to do what society wants.

- » **Every firm produces output at the minimum possible cost.** This outcome is good for society because it means that the least possible amounts of resources are consumed while producing the output that society wants.

Peering into the process of perfect competition



REMEMBER

How does perfect competition actually work? The following four steps explain it:

1. **The market price of the output sold by every firm in the industry is determined by the interaction of the industry's overall supply and demand curves.**

2. Each firm takes the market price as given and produces whatever quantity of output will maximize its own profit (or minimize its own loss if the price is so low that it's not possible to make a profit).
3. Because each firm has an identical production technology, each will choose to produce the same quantity and will consequently make the same profit or loss as every other firm in the industry.
4. Depending on whether firms in the industry are making profits or losses, firms will either enter or leave the industry until the market price adjusts to the level where all remaining firms are making zero economic profit.

The fourth point in this process — firm entry and exit — is critical. To understand it clearly, I break it into two cases: one where every firm in the industry is making a profit because the market price is high and another where every firm in the industry is taking a loss because the market price is low:

» **Attracting new firms by making profits:** If every firm in an industry is making a positive economic profit, new firms are attracted to enter the industry in hopes of sharing in the profits. But when they enter, total industry output increases so much that the market price begins to fall. As the price falls, so do profits, thereby lowering the incentive for further firms to enter the industry.

The process of new firms entering the industry continues until the market price falls so low that economic profits drop to zero. When that happens, the incentive to enter the industry disappears, and no more firms enter.

» **Losing existing firms when taking losses:** If every firm in an industry starts out taking losses because the market price is low, some of the existing firms exit the industry because they can't stand losing money. When they do, total industry output falls. That reduction in total supply, in turn, causes the market price to rise. And as the market price rises, firms' losses decrease.

The process of firms leaving and prices rising continues until the remaining firms are no longer losing money.



REMEMBER

The fact that firms can freely enter or leave the industry means that after all adjustments are made, firms always make zero economic profit. In other words, if there's perfect competition, you don't have to worry about firms exploiting anyone; they just barely make enough money to stay in business.

The other important result of perfect competition — that competitive firms produce at minimum cost — becomes apparent if you flesh out the four-stage process of perfect competition by using cost curves. I encourage you to take a look at Chapter 6 before moving on to the next section. If you haven't read that chapter, this section may cause your eyes to cross (and you know what your mother said about the dangers of crossing your eyes).

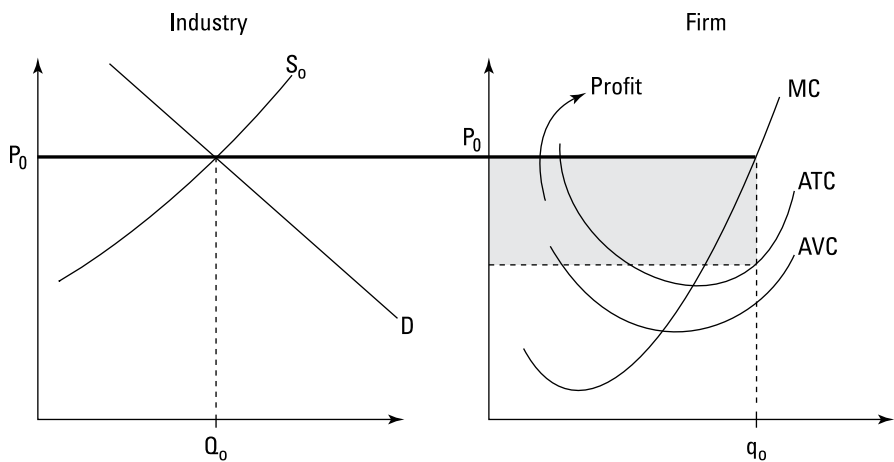
Graphing how profits guide firm entry and exit

In this section, I use firm cost curves to demonstrate how market forces automatically cause firms to produce output at the lowest possible cost. To make this process clear, I present two cases. In the first, firms begin by making profits. In the second, firms begin by taking losses. Either way, adjustments happen so that they end up making zero economic profits and producing at minimum costs.

Visualizing firm entry when there are profits

To see how an industry adjusts when it starts off making profits, look at Figure 7-10, which consists of two graphs. The one on the left gives the market demand curve, D , and the initial market supply curve, S_0 , for tennis balls. The one on the right gives the cost curves for one of the many identical firms that make tennis balls.

FIGURE 7-10: Industry demand and supply curves (left) determine a market price that a competitive firm (right) must take as given.



© John Wiley & Sons, Inc.

Because the firms in this industry are identical, they have the same cost structures. In particular, they have the same marginal cost curve (MC). This point is important because a competitive firm's marginal cost curve is its supply curve, as I show in Chapter 6.

The example firm takes the market price, P_0 , that's determined by supply and demand in the left graph and uses it to figure out its profit-maximizing output level in the right graph. (To emphasize that P_0 is the same in both graphs, I've drawn a solid horizontal line that goes all the way across both graphs.)



TIP

Each firm chooses to produce the output level at which the horizontal price line intersects the MC curve. In the right-hand graph, I label the output level q_o . In the left-hand graph, you can see that the industry's total supply is Q_o . The industry's total supply is simply each firm's output, q_o , times the total number of firms in the industry.

Next, focus on the fact that each firm runs a profit when the market price is P_o . The profit is shown by the shaded rectangle in the graph on the right.

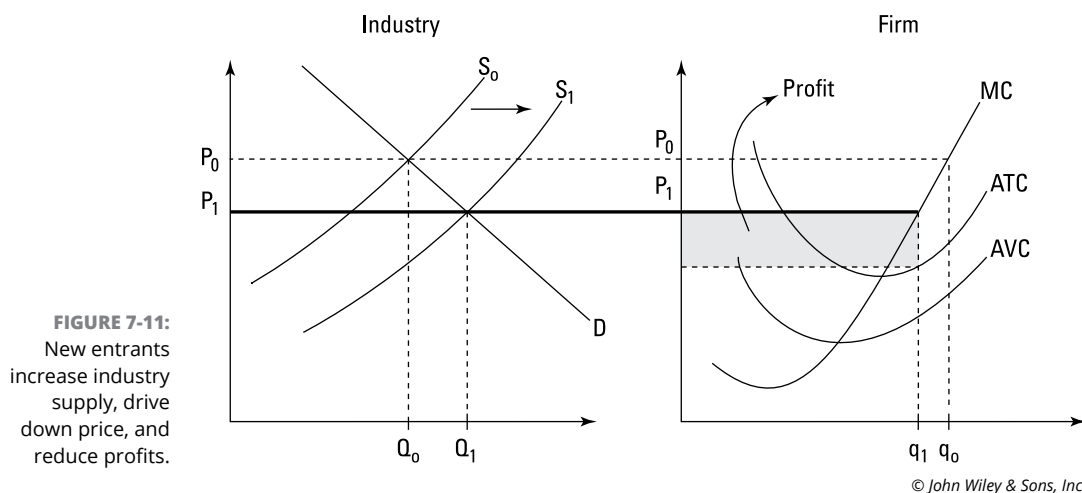


REMEMBER

This profit is important because it attracts entrepreneurs to enter the industry. They realize that they can set up yet more identical firms and make some nice profits. As economists like to say, profits attract entrants.

SEEING HOW NEW ENTRY REDUCES PROFITS

Figure 7-11 shows what happens when the new entrants to the industry arrive. Their new production increases overall production so that the total supply curve shifts from S_o to S_1 in the left-hand graph. That lowers the market equilibrium price from P_o to P_1 .

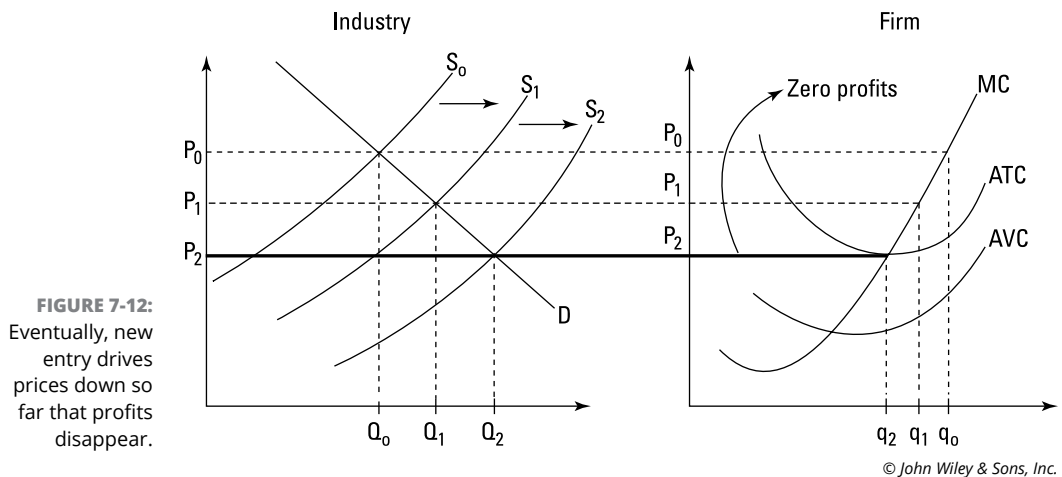


Each of the price-taking firms reacts to the lower price by producing a lower output level, q_1 , which you can see illustrated in the right-hand graph. More importantly, the firms' profits decrease, which you can see by comparing the shaded profit rectangles in Figures 7-10 and 7-11.

The new entry results in smaller profits. The smaller profits are less attractive to entrepreneurs. So although the fact that some profits are still available will still cause new entry, there won't be as much new entry as when profits were larger.

SEEING HOW ENOUGH ENTRY DRIVES PROFITS TO ZERO

As more firms enter the market, entry continues until prices fall so far that all profits are driven away. This situation is illustrated in Figure 7-12, in which new entry has increased supply still more, to S_2 . The result is that the market price falls to P_2 , which results in zero profits. (Note that there's no shaded profit rectangle in the right-hand graph.) Because profits fall to zero, entry ceases.



Realizing that zero profits also means minimum-cost production

When profits are driven to zero by the entry of new firms, the cost per unit at which output is produced is minimized. You can see this fact in the right-hand graph of Figure 7-12 by noticing that when faced with price P_2 , firms choose to produce at the quantity that minimizes per-unit production costs.

You can tell this is true because the output that firms choose to produce, q_2 , lies exactly at the minimum point of the U-shaped average total cost (ATC) curve. When output is produced at that level, the average cost per unit is lower than at any other output level. (In other words, any other output level results in a higher average total cost.)

This is a wonderful thing because it means that each firm is being as efficient as possible, producing output at the lowest possible cost per unit. Moreover, each firm is voluntarily choosing to produce at that level without the need for coercion.



REMEMBER

What's going on here is that profits serve as a self-correcting feedback mechanism. High profits automatically attract new entrants that automatically increase supply and drive prices down. That process continues until there are no more profits and no more new entrants. But more importantly, it continues until each firm is producing output at the most efficient, least-cost output level. This is truly Adam Smith's invisible hand at work.

Visualizing firm exit when there are losses

A competitive industry in which firms are losing money is an industry that should reduce its output level. Firms lose money only when the output they produce is worth less than the cost of the inputs used to produce it. Reducing production solves this problem because a reduced quantity of output has a higher average value to consumers (as you can see by sliding down any demand curve and noting that successive units are worth successively less to consumers).

A fortunate thing about perfect competition is that it provides a feedback mechanism that leads to zero profits and efficient production whenever an industry starts out taking losses. To see this, take a look at Figure 7-13, where the initial supply curve, S_3 , interacts with the demand curve, D , to produce a low market price of P_3 .

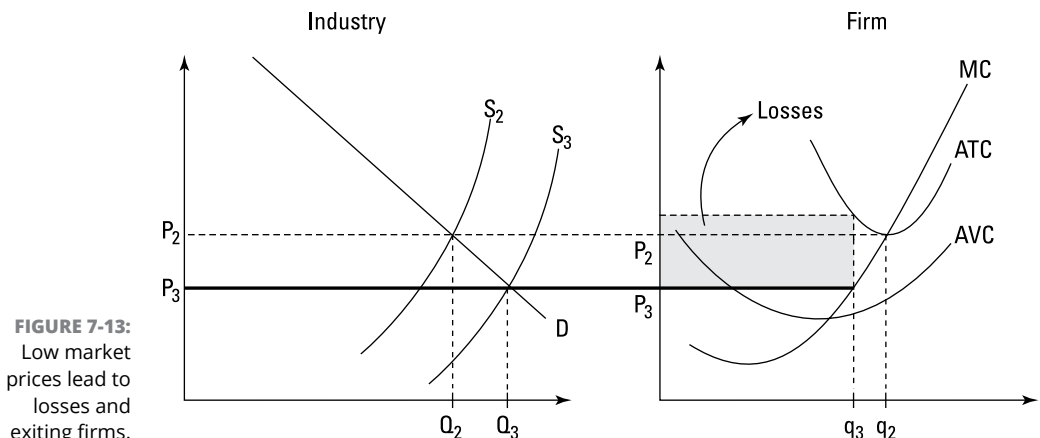


FIGURE 7-13: Low market prices lead to losses and exiting firms.

© John Wiley & Sons, Inc.

At this market price, you can see in the right-hand graph that each firm in the industry is making a loss, which is shown by the shaded rectangle.



REMEMBER

This loss discourages all the firms in the industry, and those in the weakest financial condition begin to exit. As that happens, the industry supply curve in the left-hand graph shifts left because supply decreases. That shift raises the market price and reduces the losses made by firms remaining in the industry. But as long as there are losses, firms continue to exit until the supply curve moves all the way back to S_2 , at which point the market price is P_2 , and firms are making zero profits as in Figure 7-12.

When the market price reaches P_2 and firms are making zero profits, firm exits stop and, more importantly, each firm is producing at the least-cost output level, q_2 .

Understanding that entry and exit don't happen instantly

Market pressures always push perfectly competitive firms to produce at the lowest possible per-unit cost. (See the preceding sections.) Keep in mind that this nice result doesn't happen overnight. When firms are making profits or sustaining losses, it takes time for new firms to enter (if there are profits) or for existing firms to leave (if there are losses).

Depending on the industry, these adjustment processes may take anywhere from a few weeks to a few years. For example, setting up new power plants takes a while because building a new power plant takes at least a year. Similarly, even if agricultural prices fall and farmers are taking losses, those farmers who drop out of the industry won't do so until the next growing season. On the other hand, if producing U.S. flags suddenly becomes really profitable, you can be sure that scores of new firms will pour into the industry within weeks.

The wonderful thing about perfect competition is that there are always market forces acting to drive firms to produce at the minimum possible cost. As I show you in the next few chapters, this lovely result falls apart when monopolies, oligopolies, public goods, and other problems prevent perfect competition.

- » Producing less and charging more than competitive firms
- » Maximizing profit
- » Benefiting society (in certain situations)
- » Abiding by regulations

Chapter 8

Monopolies: Bad Behavior When Competition Is Lacking

Economists use the word *monopoly* to describe either an industry that has only a single firm or a firm that has no competitors. Both of those definitions make sense because the word monopoly is Greek for “one seller.”

Monopolies are much maligned because their profit incentive leads them to raise prices and lower output to squeeze more money out of consumers. As a result, governments typically go out of their way to break up monopolies and replace them with competitive industries that generate lower prices and higher output.

At the same time, however, governments also intentionally create monopolies in other situations. For instance, governments issue patents, which give monopoly rights to inventors to sell and market their inventions. Similarly, in many places, local services such as natural gas delivery and trash collection are monopolies created and enforced by local governments.

In this chapter, I explain why society forbids monopolies in some situations and promotes them in others. First, I show you that profit-maximizing monopolies

compare unfavorably with competitive firms because they set higher prices and produce less output than competitive firms. Then I explain how these problems may, in certain cases, be outweighed by other factors — the need to promote innovation, for example, and the odd fact that in some cases having a lot of competitors is just too annoying.

Examining Profit-Maximizing Monopolies

Essentially, this chapter is one big exercise in cost-benefit analysis. Monopolies aren't all evil. Neither are they utterly good. Whether you want to have one in any particular instance depends on whether, in that situation, the benefits outweigh the costs. This section goes into detail about the costs associated with monopolies.

Zeroing in on the problems monopolies cause



REMEMBER

In an industry that has only one monopoly firm rather than lots of small competitive firms, three socially harmful things occur:

- » The monopoly firm produces less output than a competitive industry would.
- » The monopoly firm sells its output at a higher price than the market price would be if the industry were competitive.
- » The monopoly's output is produced less efficiently (i.e., at a higher cost per unit) relative to the efficiency level of a competitive industry.

Although all these things are harmful to consumers, keep in mind that monopolies don't do these things to be jerks. Rather, these outcomes are simply the result of monopolies' acting to maximize their profits — which is, of course, the same thing that competitive firms try to do.



REMEMBER

The difference in outcomes between a competitive industry and a monopoly industry doesn't have anything to do with bad intentions. Rather, it results from the fact that monopolies are free from the pressures that lead competitive industries to produce the socially optimal output level. (See Chapter 7 for info on these pressures.) Without these pressures, monopoly firms can increase prices and restrict output to increase their profits — things that competitive firms would also love to do but can't.

The lack of competitive pressure also means that monopoly firms can get away with costly, inefficient production. This is a real problem that you should take seriously when considering whether the benefits of a monopoly outweigh its costs. I talk more about this issue later in the chapter.

Identifying the source of the problem: Decreasing marginal revenues



REMEMBER

All the bad outcomes a monopoly generates derive from the same source: The monopolist faces a downward-sloping marginal revenue curve. *Marginal revenue* is the increase in total revenue that comes from selling each successive unit of a product. This simple fact causes monopolies to charge more, produce less, and produce at higher costs than competitive firms.

How can one little curve cause such mayhem? Because a downward-sloping marginal revenue curve implies that each additional unit that the monopoly sells brings less revenue than the previous unit. For instance, although the 10th unit sold may bring in \$8, the 11th brings in only \$3. Obviously, such a situation reduces the incentive to produce a lot of output.

This situation stands in stark contrast to the marginal revenue situation facing competitive firms. (See Chapter 6.) Competitive firms face horizontal marginal revenue curves, meaning that whether they sell 11 units or 11,000, each unit brings in the same amount of money. Naturally, that's much more of an inducement to produce a lot of output.

Facing down demand

Why is there such a difference between the marginal revenue curves facing monopolists and competitive firms? A monopoly is free to choose the price it wants to charge along the demand curve it faces for its product. A competitive firm, on the other hand, has to take the market price as given.

A monopoly firm can choose its price because, as the only firm in its industry, it controls all the output in that industry. As a result, it can create a relatively high price by producing only a few units, or it can induce a relatively low price by flooding the market. By contrast, each firm in a competitive industry is such a small part of its industry that its choice of output makes too small a difference in total output to cause price changes. (See Chapter 7 for more on why competitive firms can't affect prices.)



REMEMBER

The monopolist's ability to control the price by altering its output level means that the firm has to step back and consider what output level to produce. The firm's goal is profit maximization, so it has to figure out what level of output will maximize its profits. It turns out that a monopolist's profit-maximizing output level is defined by the same condition as that of a competitive firm: Produce at the output level where the marginal revenue curve crosses the marginal cost curve.

Deriving marginal revenue from the demand curve



REMEMBER

For a monopoly, the marginal revenue of each successive unit of output is less than the marginal revenue of the preceding unit because demand curves slope downward. The monopoly's marginal revenue curve has a precise relationship with the demand curve for the monopoly's output: If the demand curve is a straight line, the slope of the marginal revenue curve is twice as steep as the slope of the demand curve, meaning that marginal revenue falls quite quickly as output increases.

To see how this works, take a look at Figure 8-1, which draws out a demand curve and its associated marginal revenue curve.

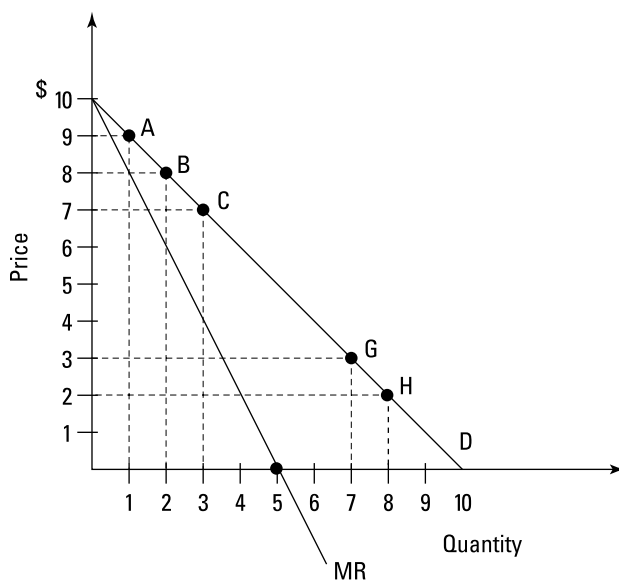


FIGURE 8-1:
The MR curve for a monopolist facing a straight-line demand curve has a slope twice as steep as that of the demand curve.

© John Wiley & Sons, Inc.

I provide the data needed to draw these two curves in Table 8-1. The first column contains output levels ranging from zero to ten units. The second column shows the price per unit that can be charged at each output level. The third column shows the total revenue that the monopoly would get for producing and selling each output level — the price per unit times the number of units. And the final column gives the marginal revenue — the change in total revenue — that happens as you increase output by one unit.

TABLE 8-1

Price and MR for Various Output Levels

Output	Selling Price	Total Revenue	Marginal Revenue
0	\$10	\$0	
			\$9
1	\$9	\$9	
			\$7
2	\$8	\$16	
			\$5
3	\$7	\$21	
			\$3
4	\$6	\$24	
			\$1
5	\$5	\$25	
			-\$1
6	\$4	\$24	
			-\$3
7	\$3	\$21	
			-\$5
8	\$2	\$16	
			-\$7
9	\$1	\$9	
			-\$9
10	\$0	\$0	

To clarify that marginal revenue represents the *change* in total revenue, the entries in the marginal revenue column are displayed between the two total revenue figures to which they correspond. For instance, total revenue increases from \$0 to \$9 as you move from producing no output to one unit of output. That's why I place the marginal revenue of \$9 between the total revenue entries of \$0 and \$9.

As you see in Figure 8-1, the marginal revenue (*MR*) curve starts at the same point as the demand curve, but it falls with twice the slope. It hits the horizontal axis at an output level of $q = 5$ instead of the $q = 10$ output level at which demand hits the horizontal axis (where q stands for quantity produced).

Relating marginal revenue to total revenue

You can get a handle on why the monopolist's marginal revenue curve falls so quickly if you first examine total revenue, or *TR*. The total revenue that the monopolist can get is simply the output it produces times the price at which it can sell its output. That is, $TR = p \times q$. However, the price at which a monopolist can sell depends on how much the firm produces.

The relationship between the output produced and the price at which the product can be sold depends on the demand curve. For instance, consider Point A on the demand curve in Figure 8-1. At that point, one unit is being produced, and it can be sold for \$9. Consequently, the total revenue at that point is \$9. Next, look at Point B, at which two units of output are being sold. At that output level, each unit can be sold for \$8. Consequently, total revenue is $\$8 \times 2 = \16 . And at Point C, where three units can be sold for \$7 each, total revenue is \$21.

The important thing to notice is how total revenue changes as you move from A to B to C and output increases from one to two to three units. Total revenue goes from \$9 to \$16 to \$21. Obviously, total revenue increases. But look more deeply. Moving from A to B, *TR* increases by \$7 (from \$9 to \$16). But moving from B to C, it increases by only \$5 (from \$16 to \$21). Each successive increase in total revenue is smaller than the preceding increase. In other words, marginal revenue declines as the monopoly increases production.

Looking at Table 8-1, you can see that marginal revenue continues to fall for each successive unit. In fact, it becomes negative for all units after the fifth. You can see why by looking at Points G and H in Figure 8-1 as examples. At Point G, the monopolist can sell seven units of output for \$3 each. That makes for a total revenue of \$21. But if the monopolist increases output to eight units at Point H, they can sell these units for only \$2 each, implying a total revenue of \$16.

Increasing output from seven units to eight units means decreasing total revenue from \$21 to \$16. That's the same thing as saying that marginal revenue is *negative* \$5 as you move from seven to eight units of output.

Sliding down the demand curve: Higher output, lower prices

The reason marginal revenue keeps declining and even becomes negative is that the demand curve slopes downward, meaning that the only way to get people to buy more stuff is to offer them a lower price per unit. You have to offer them a lower price per unit not just on additional units but on all previous units as well. In other words, if the monopolist wants to sell only one unit (see Point A), they can get \$9 for it. But if the monopolist wants to sell two units (see Point B), they have to lower the price to \$8 per unit for *both* the first unit *and* the second unit.

Because total revenue equals price times quantity ($TR = p \times q$), you can see that the monopolist faces a trade-off when increasing production and sliding down the demand curve. As the monopolist produces more, the quantity q obviously goes up, but the price per unit p must fall. What happens to TR depends on whether the increases in q (output effects) are bigger than the decreases in p (price effects).

You can see from Table 8-1 that as the monopoly increases production through the first four units, total revenue keeps increasing, meaning that the gains from selling more units more than offset the declines from getting less money per unit. At an output of five units, the two effects cancel each other out. And for higher outputs, total revenue falls because the negative effect of less money per unit overwhelms the positive effect of selling more units.



REMEMBER

Because marginal revenue (MR) tells you how total revenue (TR) changes as you increase output, the changes in TR caused by increasing output show up in MR as well. If you look at Figure 8-1, you can see that MR is always declining. That's because the negative price effect of getting less per unit keeps getting stronger relative to the positive quantity effect of selling more units.



TECHNICAL
STUFF

For straight-line demand curves (like the one you see in Figure 8-1), the MR curve is a straight line that has twice as steep a slope as the demand curve. If you know calculus, you can prove that the MR curve falls twice as fast as the demand curve. First take the equation of the demand curve shown in Figure 8-1, $p = 10 - q$, and substitute it into the total revenue equation, $TR = p \times q$. That gives you $TR = (10 - q)q = 10q - q^2$. Then take the first derivative with respect to output, q . Because marginal revenue is dTR/dq , you find that $MR = 10 - 2q$, meaning that MR has the same y -intercept as the demand curve but twice as steep a slope.

When you understand the marginal revenue situation facing a monopolist, you can combine it with the firm's marginal cost curve to figure out its profit-maximizing output level. As I show you in the next section, this level is less than that chosen by a competitive firm — a behavior that leads to a social harm that can be quantified by calculating the dollar amount of the deadweight loss that results from producing less than the socially optimal output level.

Choosing an output level to maximize profits

A monopoly is no different from a competitive firm when it comes to the costs of producing output. Just like a competitive firm, a monopoly has fixed costs, variable costs, and marginal costs. (See Chapter 6.) More importantly, these costs behave the same way, whether a firm is competitive or a monopoly. This means you can use costs to help analyze the decision-making process of a monopoly in the same way that you use them to analyze the decision-making process of competitive firms.

The key difference, however, is that the monopoly faces a downward-sloping marginal revenue curve. In this section, you see how this factor causes a profit-maximizing monopoly to produce less output than would a profit-maximizing competitive firm.

Setting a monopoly's marginal revenue equal to marginal cost

The monopoly goes about maximizing profits in much the same way as a competitive firm. To see this, take a look at Figure 8-2, which draws a monopoly's average total cost (*ATC*) and marginal cost (*MC*) curves on the same graph as the monopoly's demand curve and marginal revenue (*MR*) curve.

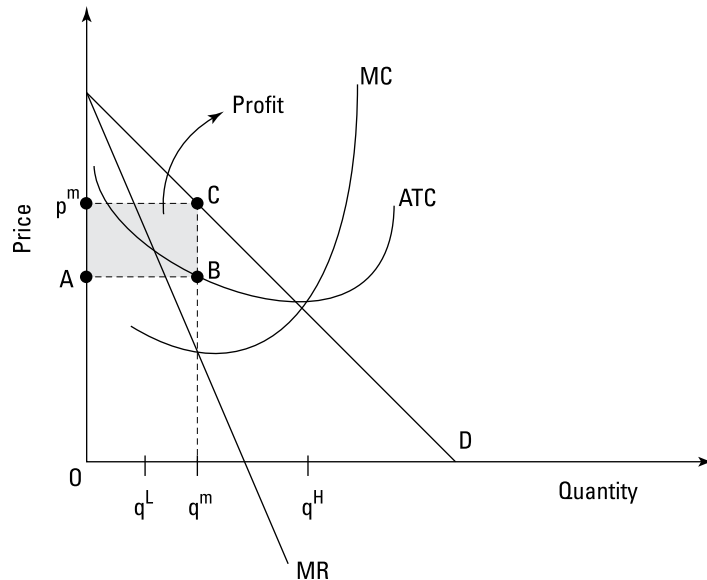
For every output level, q , the *ATC* curve gives the average total cost per unit of producing q units of output. This curve is U-shaped because average total costs first fall due to increasing returns and then increase due to diminishing returns. The *marginal cost curve* gives the cost of producing one more unit of output; that is, it tells you how much total costs rise if you increase output by one unit.



REMEMBER

The profit-maximizing monopolist's optimal output level, q^m , is determined by where the *MR* and *MC* curves cross. As with a competitive firm, choosing to produce where marginal revenue equals marginal cost ($MR = MC$) either maximizes profits or minimizes losses, depending on whether demand is strong enough for the firm to be able to make a profit. (See Chapter 6 for details.)

FIGURE 8-2:
A monopoly produces where its MC curve crosses the MR curve.



© John Wiley & Sons, Inc.

You can see why q^m is optimal by looking at two different output levels, q^L and q^H , where L stands for low and H stands for high:

- » **Low output:** At output level q^L , you see that MR at that output exceeds MC , meaning that if you produce and sell that unit, it will generate more in revenue than it costs to produce. Clearly, this is a good unit to produce. Because a similar relationship holds true for all output levels less than q^m , the monopolist should keep increasing output until it reaches q^m .
- » **High output:** On the other hand, the monopolist doesn't want to increase output beyond q^m . To see why, examine output level q^H . At that output level, marginal costs are much bigger than marginal revenues, meaning that if you produce that unit of output, the cost of producing it will exceed the money you could get selling it. In other words, if you produce that unit, you'll lose money.

If you're a monopolist, you want to produce exactly q^m units because for all units up to q^m , marginal revenues exceed marginal costs, meaning that you receive more money selling such units than you spend producing them.



TIP

Next, note that you can use the demand curve in Figure 8-2 to determine what price the monopolist will charge at output level q^m . That's possible because, as we saw in Chapter 3, demand curves indicate consumers' maximum willingness to pay for any particular amount of output.

In this case where the monopolist is going to want them to purchase the profit-maximizing output level q^m , you can calculate the maximum price per unit that consumers would be willing to pay for that amount by starting out from output level q^m on the horizontal axis, going up vertically until you hit the demand curve, and then moving horizontally to the left until you hit the vertical axis. By doing so you can see that at output level q^m , the monopolist can charge a maximum price of p^m per unit (which it will do because it has no competitors and thus no reason to charge anyone less than the maximum amount that they're willing to pay).

Eyeing the monopoly's profit

A monopoly's profit is the difference between total revenue and total costs. In Figure 8-2, the profit that the monopolist makes is shown by the shaded rectangle with corners at A , p^m , C , and B . As I discuss in Chapter 10, such profit rectangles are derived by comparing the two rectangles that give, respectively, total revenues and total costs.



TIP

The basic trick is to remember that the area of a rectangle is defined as a product — the product of its length times its width. Total revenue and total costs are also products — the price (or average cost) per unit multiplied by the number of units. Therefore, on a graph showing price along the y -axis and quantity along the x -axis, total revenue and total costs are represented as rectangles:

- » **Total revenue:** For the monopolist maximizing profits by producing q^m units and selling them for p^m dollars, total revenue is price times quantity: $TR = p^m \times q^m$. Consequently, total revenue is the area of the rectangle whose length is equal to the price and whose width is equal to the quantity. That is, TR is the area of the rectangle that has corners O , p^m , C , and q^m .
- » **Total costs:** Total costs are also a product — a product of the average cost per unit times the number of units. If you go up from Point q^m until you hit the ATC curve, you get to Point B . The vertical distance up to Point B gives the average cost per unit of producing output q^m . So if you multiply that amount by the output q^m , you get total costs. Geometrically, that means total costs are given by the rectangle whose corners are O , A , B , and q^m .

In Figure 8-2, the total revenue rectangle (O , p^m , C , q^m) is bigger than the total cost rectangle (O , A , B , q^m), meaning that the monopoly is earning a profit. That profit is given by the shaded rectangle whose points are A , p^m , C , and B , which represents the difference in areas between the total revenue and total cost rectangles.

Understanding that monopoly doesn't guarantee profitability

Just because a firm has a monopoly doesn't mean that it's guaranteed a profit. If demand is weak relative to costs, then the monopoly price will be too low to generate a profit.

To see an example of this situation, look at Figure 8-3, where the demand curve, D_1 , leads to a marginal revenue curve, MR_1 . The monopoly again sets marginal revenue equal to marginal cost to find the optimal output level, q_1^m . But because of demand being low relative to costs, the monopoly operates at a loss represented by the area of the shaded rectangle.

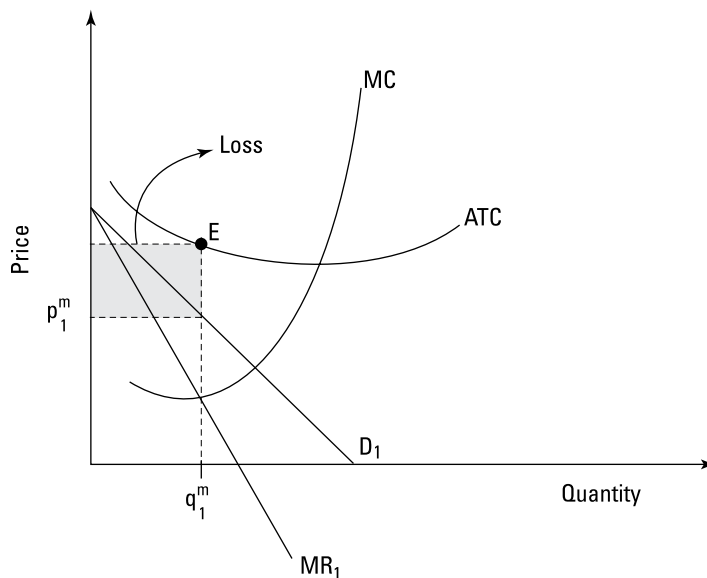


FIGURE 8-3:
A monopoly facing weak demand can sustain a loss.

© John Wiley & Sons, Inc.

One way to see that the shaded rectangle gives a loss is to compare the total revenue rectangle with the total cost rectangle, as I did for Figure 8-2. Here in Figure 8-3, the total cost rectangle exceeds the total revenue rectangle by the amount of the shaded rectangle.



TIP

A different way to understand where the loss comes from is by comparing the monopoly's average total cost per unit with the price per unit it gets when producing and selling at output level q_1^m . At that output level, you find the price per unit, p_1^m , by starting on the horizontal axis at q_1^m and then going up to the demand

curve. As you can see, you have to go up even farther to get to the ATC curve, meaning that the average total cost per unit to make q_i^m units exceeds the price per unit you get from selling these units. This fact implies that the firm will lose money producing at output level q_i^m .

As I show in Chapter 6, a firm in such a situation can't do any better. That is, any other output level besides q_i^m would produce an even bigger loss. If the monopoly can't figure out a way to either reduce costs or increase demand, it will quickly go bankrupt.

Comparing Monopolies with Competitive Firms

In this section, I compare a profit-maximizing monopoly with a profit-maximizing competitive firm. This comparison comes off poorly for the monopoly because competitive firms deliver socially optimal output levels. Because monopolies always end up producing less than competitive firms, their output levels are always less than socially optimal.

Looking at output and price levels

Monopolies produce less than competitive firms because they have different marginal revenue curves. As I show earlier in this chapter, monopolies face downward-sloping marginal revenue curves. By contrast, competitive firms face horizontal marginal revenue curves.

You can see the comparison in Figure 8-4, where I've drawn in both the downward-sloping marginal revenue curve of a monopoly, MR^m , and the horizontal marginal revenue curve of a competitive firm, MR^c . The graph also has an average total cost curve, ATC , as well as a marginal cost curve, MC .

Figure 8-4 assumes that the competitive firm and the monopoly have the same cost structure, which is why I show only one MC curve and one ATC curve. By assuming that both firms have the same cost structure, I can isolate the effect that the difference in marginal revenue curves has on each firm's output decisions.

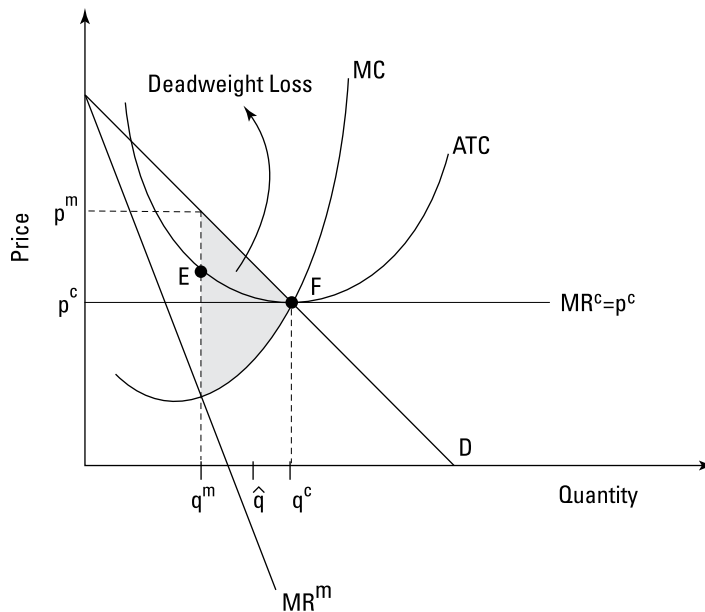


FIGURE 8-4: If a monopoly and a competitive firm have the same cost structure, the monopoly produces less, which causes a deadweight loss.

© John Wiley & Sons, Inc.

Maximizing profits for each firm

The marginal revenue curve for a competitive firm, MR^c , is a horizontal line set at the market price, p^c . This is the case because a competitive firm is such a small part of its industry that the firm can't affect the market price. As a result, it can sell as many or as few units as it wants at p^c , meaning that the marginal revenue it gets for every unit it chooses to produce is p^c . As I show in Figure 8-4, $MR^c = p^c$ for a competitive firm.

Market forces adjust supply and demand until the market price is equal to the minimum average total cost at which a firm can produce. Geometrically, this means that the horizontal $MR^c = p^c$ line just touches the bottom of the U-shaped ATC curve. (Refer to Chapter 7 for more information.)



REMEMBER

Monopolies and competitive firms follow the same basic rule to maximize profits: They each produce where their marginal revenue curve intersects their marginal cost curve. But because they have different marginal revenue curves (as in Figure 8-4), they produce different outputs. The competitive firm produces q^c , and the monopoly produces q^m .

Understanding why the monopoly produces less

A competitive firm produces more than the monopoly because the competitive firm doesn't have to worry about reducing its revenue per unit if it increases

output. No matter how much it produces, it always receives $MR^C = p^C$ on every unit sold because its output is too small relative to total output to affect the market price.

By contrast, the monopolist faces the market demand curve, meaning that every additional unit it sells lowers the price per unit that it receives on all units sold. Graphically, this appears as a downward-sloping MR^m curve that leads the monopoly to restrict output because it knows that the more it produces, the less money per unit it gets.

Because the monopoly restricts output compared to the competitive firm, the monopoly price, p^m , is also higher than the competitive price, p^C . This fact really irks consumers, but as I show you in the next section, the real harm comes from the reduction in output.

Deadweight losses: Quantifying the harm caused by monopolies

Monopolies cause harm because they reduce output below the socially optimal level produced by competitive firms. Take a look at Figure 8-4 and consider whether it would be good for society if all the units of output between the monopoly output level, q^m , and the competitive output level, q^C , were produced. For instance, look at unit q^A . At that level of output, the demand curve is above the marginal cost curve. That implies that people are willing to pay more for that unit of output than it costs to make it. In other words, benefits exceed costs for that unit of output. Because this is true for all units between q^m and q^C , monopolies hurt society by failing to produce units of output for which benefits exceed costs.

The harm caused to society when the monopoly fails to produce output level q^A can be quantified as the vertical distance between the demand curve and the marginal cost curve at output level q^A . That vertical distance is a dollar amount — the number of dollars by which benefits would exceed costs for that unit if it were produced and consumed. If you go through the same exercise for each and every unit between q^m and q^C , you can total up the harm caused by the failure of the monopoly to produce all those units. Graphically, the total harm measured in dollars is equal to the area of the shaded deadweight loss region in Figure 8-4.

The deadweight loss region demonstrates that when monopolies restrict output to maximize their profits, they fail to produce units for which benefits exceed costs. That harms society.

Losing efficiency

Another problem with monopolies is that they aren't efficient producers. You can see this in Figure 8-4. Competition leads competitive firms to produce at q^c , the output level that puts them at the bottom of the average total costs (ATC) curve. You can see this output level and amount of average total cost at Point F.



TIP

The competitive output level q^c minimizes production costs per unit of output, which you can see by comparing q^c with any other output level. Whether you produce more or less than q^c , average costs per unit will be higher due to the U-shape of the ATC curve.

In particular, look at the monopoly output level, q^m . At that output level, you get to Point E on the ATC curve. Because the vertical distance between the horizontal axis and E is greater than the vertical distance between the horizontal axis and Point F, you know for certain that total costs per unit when producing the monopoly output level, q^m , are higher than those when producing the competitive output level, q^c . Consequently, the monopoly is less efficient than the competitive firm despite having an identical overall cost structure.



REMEMBER

This bad result is yet another consequence of monopoly firms facing downward-sloping marginal revenue curves. The competitive firm's horizontal marginal revenue curve implies that it can sell as many units as it wants at the competitive price p^c . Consequently, the competitive firm has an incentive to increase output all the way to q^c because doing so lowers per-unit production costs without reducing per-unit revenue. The same cost-reducing incentive exists for a monopolist to increase its output; but the cost reduction is more than offset by declining marginal revenue as the monopoly moves down along its downward-sloping marginal revenue curve. As a result, the monopolist's profits are maximized at q^m , even though q^c is the lowest-cost output level.

Considering Good Monopolies

In some cases, the benefits of monopolies actually outweigh their costs. This section takes a closer look at these instances, where the goal is to encourage innovation, cut redundancy, and keep costs low.

Encouraging innovation and investment with patents

The most obvious place where monopolies do society a lot of good is patents. Patents give inventors the exclusive right to market their inventions for 20 years, after which time their inventions become public property. So inventors have the right to run a monopoly for 20 years.



REMEMBER

Without patents, inventors are unlikely to ever see financial rewards for their hard work because copycats will steal their ideas and flood the market with rip-offs, thereby collapsing the price. Consequently, in a world without patents, far fewer people would bother to put in the time, effort, and money required to come up with new inventions.

To remedy this situation, nations all over the world have established patent offices to issue patents to inventors. The result is faster innovation, much more rapid economic growth, and much faster increases in living standards. Indeed, it's hard to think of any more socially beneficial monopolies than those that arise from patents.

Reducing annoyingly redundant competitors

Societies have also stepped in to create monopolies when competition means annoying redundancies. Consider the following examples:

- » **Trash hauling:** Garbage trucks are extremely loud and annoying. If one company has a monopoly on hauling trash, you have to endure a loud, annoying truck only once per week. But if, say, seven different trash-hauling companies compete, you may have to endure one each day if you and six of your neighbors each choose to use a different company that picks up on a different day of the week.
- » **Fiber-optic internet access:** If ten different companies offering fiber-optic internet access competed for your business, neighborhoods would have to have ten different sets of fiber-optic cables running through them — at much greater expense than running just one set of cables.
- » **Natural gas:** Laying the pipes that deliver natural gas is expensive, and laying down multiple grids of gas pipe in one area would be wasteful.

Consequently, most towns and cities have decided that there will be only one trash-hauling company, one company laying fiber-optic cables, and one natural gas company. Each company is given a monopoly and is then regulated to make sure that it doesn't exploit customers. (See the upcoming "Regulating Monopolies" section.)

Keeping costs low with natural monopolies

Another place where society may decide it's better to have a monopoly rather than competition is in the case of what economists refer to as *natural monopoly industries*, or *natural monopolies*. An industry is a natural monopoly if one large producer can produce output at a lower average cost per unit than many small producers. A good example is electric power generation. Due to engineering constraints, a 10-megawatt power plant can produce energy at a far lower per-unit cost than a 1-megawatt power plant can.

To see how this leads to a natural monopoly, imagine that a town needing 10 megawatts of power is initially served by ten of the small, 1-megawatt power plants. But then a big corporation comes along and builds a 10-megawatt power plant. Because the big plant can produce at a lower per-unit cost than the smaller, less efficient plants, the big plant offers lower prices and steals all the customers — meaning that the smaller plants quickly go bankrupt.

Such an industry is called a *natural monopoly* because it naturally becomes dominated by a single, low-cost producer. The perplexing problem here for policymakers is what to do with a natural monopoly. On one hand, everyone welcomes the fact that the big plant is much more efficient: It burns less fuel and causes less environmental damage. But because it has crushed all competition, people now have to worry that the new monopoly will charge high prices and produce less than the socially optimal output level.

These conflicting good and bad points typically mean that governments allow the natural monopoly to stay in business as the only firm in its industry, but at the same time, regulate the monopoly so that people don't have to worry about high prices or low output levels. By doing so, society benefits from the most efficient production method without having to worry about the problems that would otherwise result if the monopoly were left unregulated.

Regulating Monopolies

Governments have to decide when to support and when to suppress monopolies. For instance, patents support an inventor's monopoly right to produce and sell their invention for 20 years. After that, the production and sale of the invention is thrown open to competition.

In other situations, various regulatory institutions have been developed either to destroy a monopoly by breaking it apart or to regulate it after deciding to let it continue to be the only firm in its industry. In this section, I present several of these regulatory schemes and explore what they do to improve the behavior of monopolies.

Subsidizing a monopoly to increase output

A profit-maximizing monopoly produces less than the socially optimal level. In particular, a profit-maximizing monopoly produces where its downward-sloping marginal revenue curve, MR^m , intersects its upward-sloping marginal cost curve, MC , as shown in Figure 8-4. This output level, q^m , is less than the socially optimal output level that would be produced by a competitive firm, q^c . (For why, see "Comparing Monopolies with Competitive Firms," earlier.)



REMEMBER

One way to get the monopoly to produce more is to subsidize its production costs so that the marginal cost curve shifts down vertically. Doing so causes the marginal cost and marginal revenue curves to meet at a higher level of output. And if the subsidy is big enough, the monopoly can be induced to increase output all the way to q^c .

Some governments use this type of subsidy to get gas, electric, and phone companies to serve more people, especially those without money. If the monopoly firms' costs of hooking up customers are subsidized, the firms are willing to hook up more customers than they would without the subsidy. Some people object to subsidizing a monopolist, so this sort of solution isn't necessarily the most popular politically. But it's effective in increasing output.

Imposing minimum output requirements

Another way to get a monopoly to produce more is simply to order it to produce more. For instance, in some areas, cell phone companies are required to provide emergency 911 service to everyone — even to people who have otherwise had their wireless service cut off for nonpayment of their monthly bills. The same is often

true for companies that provide heating in the winter; in many jurisdictions, it's illegal to disconnect a client's heat for nonpayment of bills.

Minimum output requirements can force a monopoly to produce the socially optimal output level. They're often politically popular because many people think of monopolists as evil and exploitative and don't mind seeing them ordered to produce more.

Any forced increase in output also means a reduction in the monopoly's profit. Therefore, such programs are also popular because many people consider a monopoly's profits to be ill-gotten given the fact that the firm doesn't have to compete to earn them.

Regulators have to be careful, though, not to bankrupt the monopolies they're regulating. Depending on a monopoly's cost curves, it's quite possible to force a monopoly to produce at an output level where it loses money. Because regulators don't want to bankrupt monopolies and deny consumers access to the products they produce, regulators are careful to take a monopoly's cost structure into account when considering minimum output requirements.

Regulating monopoly pricing

Perhaps the most common way to regulate a monopoly is to set the price at which it can sell each unit of output that it produces. This approach works because it changes the monopoly firm's marginal revenue curve from sloping downward to being horizontal. Therefore, it eliminates the normal constraint that the more a monopoly sells, the less it can charge per unit.

That being said, regulators have to pay close attention to the regulated monopoly's cost structure so that they don't accidentally bankrupt the monopoly by setting an overly low regulated price.

Marginal cost pricing: Benefit society but risk bankruptcy



REMEMBER

A monopoly doesn't produce as much of its product as would be socially optimal. (See the earlier section "Comparing Monopolies with Competitive Firms" for details.) Therefore, a well-intentioned regulator may want to get a monopoly to produce every unit of output for which benefits exceed costs.

One option is *marginal cost pricing*, in which a regulated price is set at the dollar amount determined by where the marginal cost curve crosses the demand curve.

Consider the monopoly whose cost curves are given in Figure 8-5. Left unregulated, the monopoly will choose to produce the profit-maximizing output level q^m , defined by where MR (marginal revenue) crosses MC (marginal cost). From the demand curve, you can see that the firm will be able to charge price p^m per unit for that amount of output.

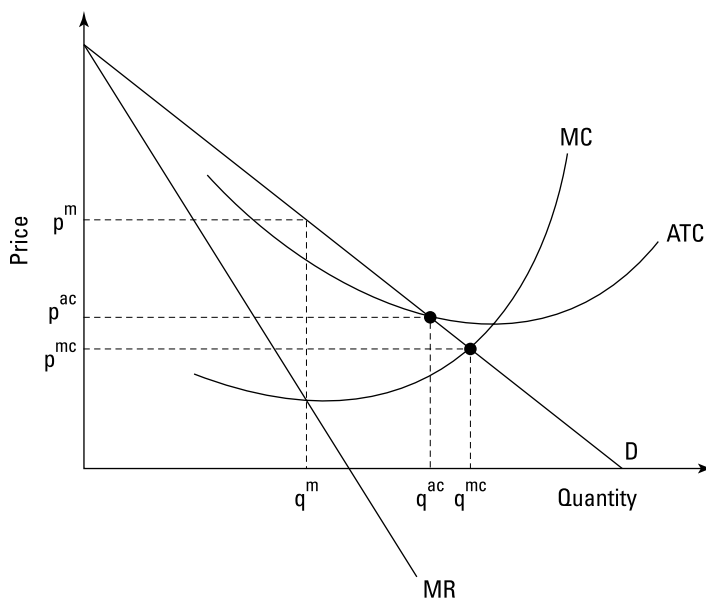


FIGURE 8-5: Regulating a monopoly using average cost pricing and marginal cost pricing.

© John Wiley & Sons, Inc.

A regulator may want to get the monopoly to produce output level q^{mc} , defined by where the downward-sloping demand curve intersects the MC curve. It's socially beneficial to produce each unit up to and including q^{mc} because for each of those units, benefits exceed costs. We know that because the amount that people are willing to pay to consume one of those units (given by the vertical distance from the horizontal axis up to the demand curve) exceeds the marginal cost of producing it (given by the vertical distance from the horizontal axis up to the MC curve).

To motivate the monopoly to produce q^{mc} units of output, the regulator can set the price at p^{mc} . At that price, the demand curve tells you that consumers will want to purchase q^{mc} units of output. Better yet, the monopoly will want to supply that level of output because the marginal revenue from selling each unit exceeds the marginal cost of producing it.

There is, however, a big problem with this policy given this particular monopoly's cost structure: The monopoly will go bankrupt. That's because at output level q^{mc} , the firm's total costs will exceed its total revenues.

You can see this problem on a per-unit basis by noting that the average total cost per unit at output level q^{mc} (given by the vertical distance from the horizontal axis up to the ATC curve) is more than the regulator-imposed revenue of p^{mc} per unit. Because the average total cost per unit exceeds the revenue per unit, the monopoly will be operating at a loss. And unless the regulator relents and allows a higher price, the monopoly will eventually go bankrupt unless the government steps in to subsidize the firm by giving it a cash grant equal to the amount of its loss.



REMEMBER

Average cost pricing

Because marginal cost pricing can cause a monopoly to lose money, a more common alternative is *average cost pricing*, which sets the regulated price where the average total cost curve (ATC) intersects the demand curve.

In Figure 8-5, a regulator using average cost pricing would set the price at p^{ac} . At that price, you can see from the demand curve that consumers demand q^{ac} units of output. The monopoly is happy to supply that output level because for each and every unit up to q^{ac} , marginal revenue (the regulated price per unit, p^{ac}) exceeds marginal cost — meaning that the monopoly gains financially by producing every unit.

The main benefit of this system is that you don't have to worry about the monopoly's going bankrupt (or where to get the money to subsidize a monopoly that would go bankrupt under marginal cost pricing). Average cost pricing guarantees that the monopoly will break even.

You can see this fact by comparing the average total costs per unit at output level q^{ac} with the revenue per unit at that output level. You get the average total cost per unit by going up until you hit the average total cost curve. Because that vertical distance is equal to the regulated price per unit, p^{ac} , you know that the average total cost per unit is equal to the regulated price per unit — so the firm must be breaking even.

The downside to average cost pricing is that the socially beneficial units between q^{ac} and q^{mc} aren't produced. That loss of output is problematic, but the deadweight loss associated with the loss of those units needs to be compared with the potential difficulties associated with following a different policy, such as marginal cost pricing.

In particular, if you were to implement marginal cost pricing, the only way to keep the monopoly in business would be to subsidize the monopoly so that it wouldn't go bankrupt. But doing that would create another problem: Where do you get the money for the subsidy? That money would have to come from tax revenues and whatever taxes are imposed elsewhere in the economy to fund the subsidy might cause a bigger deadweight loss over there than the one that would result over here from average cost pricing.

So while it's certainly problematic that average cost pricing will lead the monopoly to not produce the beneficial units between q^{ac} and q^{mc} , that harm may still be less than the harm that might be caused by other regulatory strategies.

Breaking up a monopoly into several competing firms

One solution to the problem of a monopoly is to destroy it by breaking it up into many competing firms. In the United States, the most famous example of this solution was the division of American Telephone and Telegraph Corporation (AT&T) into a bunch of smaller competitors in 1984.

Before 1984, AT&T was a nationwide monopoly. If you wanted to make a telephone call anywhere in the United States, you had to use AT&T because it was the only telephone company in the country. It was highly regulated, with both quantity requirements to provide everyone with a phone and price requirements that encouraged it to provide a high quantity of telecommunication services. But it was still a monopoly, and a judge ruled in 1984 that it should be broken up into numerous local firms to foster competition.

The policy change worked extremely well. There was soon a competitive market for telephone services between the numerous local firms that had previously been part of AT&T. More recently, the telephone service industry has become even more competitive due to the arrival of cell phone companies, internet telephony companies, and space-based mobile broadband providers like Starlink. This robust competition eliminates the problems associated with monopolies and ensures that telecommunication services are provided at low cost and in large quantity.

Creating competition is also a handy way to deal with a monopoly because it eliminates the costs associated with having to continually monitor a regulated monopoly. Competition gets you to the socially optimal output level without any sort of central control. (For info on why, see Chapter 7.) That stands in stark contrast to regulated monopolies, which typically require expensive government bureaucracies to develop and enforce laws and regulations.

IN THIS CHAPTER

- » Deciding whether to compete or collude in an oligopoly
- » Examining why some collusive pacts work and others don't
- » Regulating firms so they can't collude
- » Using product differentiation to elude perfect competition
- » Limiting profits in monopolistic competition

Chapter 9

Oligopoly and Monopolistic Competition: Middle Grounds

The two most extreme forms that an industry can take are perfect competition (with many small competitive firms) and monopoly (marked by only one firm and no competition). I cover those cases in Chapters 6 and 8. This chapter concentrates on two interesting intermediate cases: oligopoly and monopolistic competition.

An *oligopoly* is an industry in which there are only a small number of firms — two, three, or a handful. The word itself is Greek for “few sellers.” A diverse group of industries looks like this, including soft drinks and oil production. For instance,

Coke and Pepsi dominate the soft drink market, vastly outselling other carbonated beverages. Similarly, just four or five countries produce the majority of the world's oil.

Oligopoly industries are interesting because, depending on specific circumstances, the firms can either compete ruthlessly with each other or unite to behave almost exactly like a monopoly would. This means that in some cases, oligopolies can be left alone because competition ensures that they produce socially optimal output levels; in other cases, government regulation may be needed to prevent them from acting like monopolies and behaving in socially undesirable ways.

The second intermediate case is *monopolistic competition*, a sort of hybrid between perfect competition and monopoly. The key thing that sets firms in this type of industry apart from firms in a perfectly competitive industry is *product differentiation* — the fact that each firm produces a slightly different product relative to the other firms in its industry.

This chapter starts with a detailed look at oligopolies and the decisions that firms in this type of industry have to make. It then moves on to monopolistic competition and shows why product differentiation doesn't necessarily translate into tidy profits.

Oligopolies: Looking at the Allure of Joining Forces

In industries where only a few firms operate, firms have a choice about whether to compete or cooperate. This situation is very different from perfect competition (which I discuss in Chapter 7). Perfectly competitive markets have so many firms, and each firm is such a small part of the market that their individual outputs don't affect the market price. As a result, competitive firms just take the market price as given and adjust their output levels accordingly to make as large a profit as possible.

By contrast, in a market with only a few sellers, each one produces enough of the total output to be able to affect the market price. Each company can increase output and lower its own price in an attempt to gain more customers at the expense of its rivals, or the firms can work together to limit production and keep prices high to maximize their combined profits. This section explores the firms' incentives to work together and the possible benefits of acting like a monopoly. Later, in "Understanding Incentives to Cheat on a Cartel Agreement," I explain why collusion usually doesn't go as planned.

Sharing power over prices

When an industry is made up of only a few firms, the decisions that any one firm makes about how much to produce or charge for its own output almost always influence the outputs and prices of the other firms in its industry.

As an example, consider that only two major producers of cola-flavored sodas operate in the United States: the Coca-Cola Company and PepsiCo. These two corporations produce such large fractions of the total output that if either one were to suddenly increase supply, the market price of cola-flavored soda would drop dramatically not only for itself but for all the other companies in its industry too.

For instance, suppose that Pepsi doubles production and floods the market. The price of Pepsi would drop dramatically due to the large increase in the supply of Pepsi. But most soda drinkers aren't 100 percent loyal to one brand or the other, so if the price of Pepsi drops dramatically, a lot of regular Coke drinkers are going to switch brands and drink Pepsi. As a result of this decline in the demand for Coke, the price of Coke will drop too.

Cartel behavior: Trying to imitate monopolists



REMEMBER

When firms are involved in a situation in which each of their supply decisions affects not only their own sales but those of their competitors, too, economists say that the firms are in a *strategic situation* because the firms involved have to decide which type of strategy to pursue. In particular, they have to decide whether to compete or collude:

- » **Compete:** If they compete, they'll try to increase production to undercut each other on price and capture as many customers as possible.
- » **Collude:** If they collude, they'll jointly cut back on production to drive up prices and increase their profits. For producers, collusion is better than competition because it leads to profits that last as long as the firms keep colluding.

A group of firms that colludes and acts as a single coordinated whole is known as a *cartel*. Because a cartel acts essentially as one gigantic firm, it effectively turns a bunch of individual firms into a single, big monopoly.

This fact makes understanding the profit-maximizing behavior of a cartel easy. It's just like that of a monopoly. In fact, you can see what a cartel wants to do by looking at the figures in Chapter 8, which illustrate what a monopoly likes to do. In particular, a profit-maximizing cartel chooses to produce the monopoly's

profit-maximizing output level of q^m units (where marginal cost equals marginal revenue), as in Figure 8–2. Producing that output level maximizes the cartel's collective profit. And better yet for the cartel, that monopoly profit will persist as long as the participating companies keep cooperating and producing a combined output of q^m .

What's really interesting, though, is that cartels often fail to maximize their collective profit because they can't sustain the cooperation needed to keep their collective output at the monopoly output level q^m .

Considering the criteria for coordinating a cartel

For consumers, collusion is worse than competition because it leads to higher prices and lower output. You may assume that government intervention is called for to protect consumers from collusion. But such intervention is needed only if firms actually collude.

A fascinating thing about the real world is that collusion doesn't happen in a lot of industries where you might expect it to. For instance, Coke and Pepsi are fierce competitors that spend billions of dollars a year on advertising to try to steal each other's customers. Similarly, most cities have only a handful of competing cell phone networks. But instead of colluding, they compete so ruthlessly that many of them constantly flirt with bankruptcy. The same holds true for the airline industry, where bankruptcies are routine.



REMEMBER

The big question that economists have to answer is, “Why do we see so little collusion in industries where you would expect more of it?” The answer is that it's often difficult to get all the firms to coordinate so they're collectively producing the monopoly output level, q^m . To get the individual firms to cooperate and produce exactly q^m units of combined output, you have to get them to agree on two related things:

- » **How to share the profits:** Obviously, every firm wants as large a share as possible.
- » **Output quotas:** The firms must agree, and abide by, how much of the total output (q^m) each firm will produce. Each firm will constantly be tempted to produce more than its quota because doing so would bring it higher revenues.

Understanding Incentives to Cheat on a Cartel Agreement

To the extent that cartels work and create monopoly profits, those profits tempt cartel members to cheat. You can understand those temptations with game theory.

Game theory is a branch of mathematics that studies how people behave in strategic situations — situations in which people take account of others' actions or anticipated actions and then modify their own actions accordingly. For instance, chess and checkers are strategic because what I do on my current move changes what my opponent does in subsequent moves. Even more importantly, what I think my opponent will do in response to each of the moves I might make helps me calculate which of those moves is likely to lead to the best outcome.

Cartels are strategic because each firm must consider what it thinks all the other firms are going to do before deciding what it should do. Consequently, game theory models are probably the best way to understand the motivations and temptations that guide the behavior of cartel members. This section introduces a game theory model called the prisoner's dilemma to explore decisions about whether to cheat. It then applies the model to the OPEC oil cartel.

Fleshing out the prisoner's dilemma

Understanding why cartels cheat is easy if you apply the prisoner's dilemma. The *prisoner's dilemma* is a game theory model in which two criminal partners have to individually decide whether to cheat on their previous agreement to not talk to the police.

It works in this way: Imagine Jesse and James have just robbed a bank. The police know this but don't have any hard evidence against them. Rather, the only way of getting a conviction is to get one or both of the bank robbers to confess to the crime and give evidence against the other.

The police have some leverage because they managed to catch Jesse and James committing other unrelated, minor crimes. These other crimes carry with them a one-year prison sentence. The police are hoping to use the threat of a year in prison to get one or both of the bank robbers to implicate his partner in exchange for immunity from prosecution.

Jesse and James both swore to each other several days before that they'd never rat on each other, but what happens when push comes to shove?

Comparing the payoffs of confessing or remaining silent

Following standard procedure, the police separate Jesse and James, questioning them in separate interrogation rooms. The police offer each of them the chance to give evidence against the other in exchange for immunity.



REMEMBER

The problem for each man is that what happens to him depends not only on what he does but on what his partner does as well. Each man can trade a confession for immunity, but he gets the deal only if his partner doesn't confess at the same time in the other interrogation room. Four outcomes are possible:

- » **Both stay silent.** If both men keep their pact not to talk and neither confesses to robbing banks, each man gets only a year in jail for the minor offense.
- » **Only Jesse confesses and gives evidence.** If Jesse confesses and agrees to give evidence against James while James remains silent, Jesse goes free because he cooperated with police, but James gets ten years for bank robbery.
- » **Only James confesses and gives evidence.** If James gives evidence while Jesse remains silent, James goes free while Jesse goes to prison for ten years.
- » **Both confess and give evidence.** If both men admit to the crime, both get five years in prison. Why five years each? If both confess, the police don't need to make such a generous deal; they don't need to give either man immunity to get evidence against the other. On the other hand, the police want to give each criminal an incentive to confess, so they send each man to prison for only five years instead of the ten years each would get by remaining silent while his partner offered evidence.

Figure 9-1 contains a payoff matrix. It illustrates the outcomes that each bank robber receives in terms of prison time, depending on the decision that each man makes about whether to remain silent or confess.

The columns give Jesse's options, the rows give James's options, and the four rectangles in the grid show the prison times that result from each of the four possible combinations of their individual decisions about whether to confess. For instance, the upper-left rectangle represents what happens if both confess. It's divided diagonally in half, with Jesse's payoff of five years in prison given in the upper shaded triangle and James's payoff of five years given in the lower unshaded triangle. Similarly, the upper-right rectangle gives each of their payoffs if Jesse remains silent while James confesses: Jesse gets ten years in prison while James gets zero years because he gave evidence against Jesse.

FIGURE 9-1:
The payoff matrix
of prison times
facing Jesse and
James.

		Jesse's Choices	
		Confess	Remain Silent
James's Choices	Confess	Jesse gets 5 years James gets 5 years	Jesse gets 10 years James gets 0 years
	Remain Silent	Jesse gets 0 years James gets 10 years	Jesse gets 1 year James gets 1 year

© John Wiley & Sons, Inc.

Determining the dominant strategy for each prisoner

The prisoner's dilemma is famous because it illustrates that with the way the police have set up the potential payoffs, each criminal has an incentive to confess — no matter what the other criminal does. For instance, concentrate on James. Should he confess or remain silent? Well, first examine which is the better option for him if his partner in the other interrogation room confesses. Looking at the left column of payoffs, you can see from the upper-left rectangle that if James confesses while Jesse confesses, James gets five years. On the other hand, the bottom-left rectangle tells you that if James remains silent while Jesse confesses, James gets ten years.

Clearly, the best thing for James to do if Jesse confesses is to also confess. But consider whether it's better for James to confess or remain silent when Jesse remains silent in the other interrogation room. Begin with the upper-right rectangle, which shows that if James confesses while Jesse remains silent, James gets zero years in jail. By contrast, the bottom-right rectangle tells you that if James remains silent while Jesse is also silent, James gets one year in jail. Clearly, if Jesse remains silent, the best thing for James to do is confess and get zero years in jail rather than one year.

In other words, it's *always* better for James to confess. If James confesses when Jesse confesses, James gets five years rather than ten. And if James confesses when Jesse remains silent, James gets zero years rather than one. So James should *always* confess no matter what Jesse is saying or not saying to the police in the other room. Because the payoff for confessing is always better for him than the

payoff for not confessing, game theorists refer to confessing as James's *dominant strategy*, by which they mean that the strategy is superior for James no matter what the other player does.

If you go through the payoffs from Jesse's perspective, you find that confessing is also Jesse's dominant strategy because no matter what James is doing, the payoffs to Jesse if he confesses are always better than those from remaining silent.

Confessing is thus a dominant strategy for both players — so you should expect both of them to separately confess. If they do so, they end up in the upper-left box of the payoff matrix, where they both get five years in prison.

Realizing that the dominant strategy leads to a lousy outcome for both players

Typically, both men go with their dominant strategy and confess. But because both separately decide to confess, they each end up getting five years in prison — a much worse outcome than if they'd both kept their promise to each other to remain silent. If they'd both kept their promise, they each would've gone to jail for only one year. The logic of the dominant strategy is so compelling, though, that they each break the agreement and end up going to prison for five years rather than one.



REMEMBER

Cartel members also face a prisoner's dilemma because they must decide whether to obey the cartel agreement (to reduce output to the monopoly level) or cheat and overproduce. The temptation for cartel members to overproduce and break their cartel output agreement is just as strong as the temptation for prisoners to confess and break their agreement not to talk to the police.

Enforcing the agreement: Resolving the dilemma with credible threats



REMEMBER

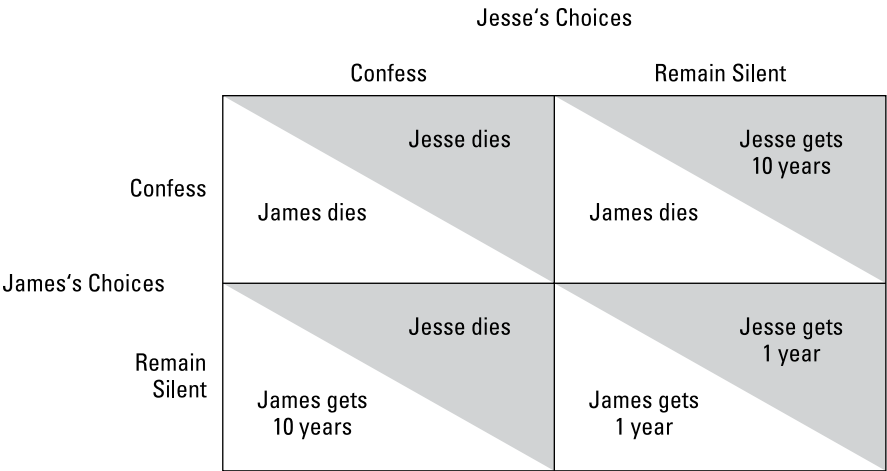
One way to get a pair of criminals to stay quiet — or to get a group of firms to stick to a cartel agreement — is to set up a strong threat against cheaters.

Mafia movies illustrate a bloodthirsty but effective system that mobsters developed to prevent people from confessing. The system is called *omerta*, which is Sicilian for “silence.” Basically, what the mafia does is change the payoffs to the prisoner's dilemma so that the dominant strategy switches from confessing to remaining silent. The mob does this by explaining to their criminal members that if they talk to the police and confess to anything or implicate anyone else, they're going to die.

The mafia also makes sure that everyone knows that this threat is credible and not just talk. Otherwise, the mafia’s members wouldn’t alter their behavior.

The mafia’s credible death threat rearranges the payoffs to the prisoner’s dilemma. Instead of just comparing prison times, as in Figure 9-1, prisoners now have to factor in death, as in Figure 9-2. In Figure 9-2, you find that the dominant strategy for both players is now to remain silent because if either talks, the mafia will hunt him down and kill him no matter what the other guy does. The result is that both Jesse and James will go to prison for only one year each because they’ll both keep their mouths shut.

FIGURE 9-2:
The dominant strategy in the prisoner’s dilemma changes when death becomes the payoff to confessing.



© John Wiley & Sons, Inc.

Paradoxically, the death threat benefits the two criminals. Even though the threat of death is scary, it actually serves their individual interests because it means that they’ll go to jail for only one year instead of five. If you’re going to be a criminal, you want to be part of a criminal organization that has enough power to bully you around and keep you from defecting on your agreements with your fellow criminals.

Seeing that OPEC is trapped in a prisoner’s dilemma

The basic version of the prisoner’s dilemma (discussed in the preceding section) is set up for just two people. But mathematicians have developed more advanced versions of the prisoner’s dilemma that you can use to analyze the behavior of larger numbers of participants. These models are invaluable for understanding

oligopoly industries with several firms and the incentives that firms in such industries face when they try to form cartels. The basic conclusion of these multifirm models is that the dominant strategy is usually to cheat on cartel agreements.

This result goes a long way toward explaining why the OPEC oil cartel has a hard time when it tries to achieve its goal of raising oil prices by reducing oil production. Quite simply, cheating on OPEC cartel agreements is a dominant strategy for OPEC member countries.

OPEC stands for the Organization of Petroleum Exporting Countries. Although OPEC is a dull name, it represents a lively group that includes Saudi Arabia, Iraq, Venezuela, Nigeria, Kuwait, Angola, and several other key oil-exporting nations. Together, these nations control the majority of the world's oil reserves, meaning that they occupy an oligopoly industry with only a few participants. Because there are only a few participants, they have a chance to form a cartel and try to produce the monopoly output and make monopoly profits. Do they succeed?

On the whole, no. I say “on the whole” because although the nations do negotiate agreements about oil production, these agreements are constantly broken. To see why that happens, you have to understand that OPEC (and its informal extension, OPEC+, which includes a dozen or so additional countries) has meetings where it decides how much total oil should be produced and what fraction of that overall amount each country should produce. At the meetings, each country is given a *quota* — a maximum amount that it's supposed to produce. For instance, Saudi Arabia might have a quota of 10 million barrels per day, whereas Venezuela's quota might be 1 million barrels per day.

The problems start after the meetings when all the oil ministers go home. Each country realizes that producing more than its quota is the best strategy. For instance, Venezuela is better off producing more than its 1 million barrel a day quota no matter what the other countries do:

- » If the other countries obey their quotas, Venezuela is better off producing more than its quota because it can sell lots of oil at a high price. (The high price is caused by the fact that the other countries are obeying their quotas and restricting supply.)
- » If the other countries break their quotas and overproduce, the price of oil will be low, meaning that Venezuela should also overproduce its quota. There's no reason to obey the quota if prices are low due to everyone else's cheating.

Because each country faces the same temptation to overproduce its quota, the OPEC cartel doesn't typically work very well. Overproducing is a dominant strategy and is simply too tempting to resist given the payoffs.

Using an enforcer to help OPEC members stick to quotas



REMEMBER

OPEC could benefit if it had some way of threatening its members for violating their quotas. Because the member countries are sovereign nations, death threats aren't an option as they are in the mafia. Rather, Saudi Arabia has sometimes tried to provide an economic threat against quota violators.

The economic threat comes in the form of super low oil prices. Saudi Arabia is in the best position to make such a threat for two reasons:

- » **It's long been the world's first or second largest oil producer.** In 2022, for instance, Saudi Arabia produced about 12 percent of the world's oil, behind the United States, which produced about 20 percent of the world's oil that year.
- » **It's the world's lowest-cost oil producer.** Saudi Arabia can produce profitably even if the price of oil falls down to \$9 per barrel. (Oil is typically priced at between \$40 and \$100 per barrel, and most other countries, including the United States, need a price of at least \$20 per barrel to break even.)

These two facts mean that if other countries cheat on their quotas, Saudi Arabia could potentially increase its production so much that the price of oil would fall very low. For instance, suppose the price fell to \$10 per barrel. Saudi Arabia would be the only OPEC member making a profit at that price; everyone else would be losing money.

As a result, Saudi Arabia appears to be in a position to threaten other OPEC members with bankruptcy if they violate their quotas. Unfortunately, the threat doesn't work that well in practice. The problem is that Saudi Arabia has limited pumping capacity. Although Saudi Arabia may be able to produce an extra 10 or 20 percent more oil per day than it normally does, that much of an increase wouldn't be enough to drive the price down to \$9 per barrel and bankrupt the other OPEC nations. Therefore, the Saudi Arabian threat isn't nearly strong enough to switch the dominant strategy from cheating on the quota to obeying it. And because OPEC has never figured out a way to effectively threaten quota violators, the cartel doesn't work very well.

Regulating Oligopolies

In some industries, cartels *are* effective at reducing output and raising prices. Often, these are industries in which one firm is large enough and powerful enough to threaten the other firms in its industry with bankruptcy.

In such instances, government regulators may decide to intervene to break up the cartel. One possible strategy is to break up an oligopoly industry with a relatively small number of firms into an even larger number of firms to promote competition. In other situations, prices or quantities may be regulated. The specific policy often depends on the specific circumstances of the industry and what policymakers think will best promote the general welfare.

Breaking up dominant firms

One important strategy for regulating an oligopoly is for the government to break it up into many smaller companies that will then compete with each other. In the 19th century, cartels were called *trusts* — for example, the Sugar Trust, the Steel Trust, the Railroad Trust, and so on. Therefore, laws that broke up monopolies and cartels were called *antitrust laws*. The most famous of these in the United States was the Sherman Anti-Trust Act. Most countries have now passed similar legislation to break up monopolies and cartels.

In U.S. history, the Standard Oil Company run by John D. Rockefeller during the 19th century dominated an oligopoly industry. It controlled something like 90 percent of the oil sold in the United States, and if a competitor didn't do what Rockefeller wanted, he would simply bankrupt the other firm by offering oil at a ridiculously low price that the competitor couldn't match.

Rockefeller would lose money temporarily while taking this action, but by bankrupting the competitors who disobeyed him, he was able to convince the remaining firms to help him restrain output and make huge profits. Indeed, because Standard Oil exerted so much control, its industry was much more like a monopoly than an oligopoly.

Rockefeller's effectiveness, however, soon brought a governmental response. Standard Oil was broken up into dozens of smaller, independent oil companies, none of which was large enough or powerful enough to dominate the industry and enforce collusion the way Standard Oil had.

Attempting to apply antitrust laws

A big problem with antitrust laws is deciding when to regulate oligopolies or break them up to promote competition. The first sign of a potential cartel is when you see only a few firms in an industry. But because of the prisoner's dilemma, in some cases even a two-firm industry (referred to as a *duopoly*) can't form an effective cartel. Consequently, prosecutors typically have to do a lot more than demonstrate that an industry has only a few firms.



REMEMBER

Typically, there has to be concrete proof of collusion. In other words, if one day every firm in an oligopoly industry decides without coordination to cut its output in half and thereby raise prices, that may not be illegal. But if even a single text message from a manager of one firm to a manager of another firm is found suggesting that the firms should enter into a cartel, that's illegal and enough for a prosecutor to hang a case on (even if the firms never actually act on the manager's suggestion).

Studying a Hybrid: Monopolistic Competition

An interesting form of competition that's found in some industries has the odd name of *monopolistic competition*. In such industries, you find characteristics of both monopolies (see Chapter 8) and competitive firms (see Chapter 7). This section provides an overview of monopolistic competition.

Benefiting from product differentiation

Like competitive firms operating in free markets, industries featuring monopolistic competition have lots of firms competing against each other. But unlike competitive free markets where all the firms in an industry sell an identical product, in monopolistic competition each firm's product is at least slightly different.

Think of the market for gasoline. Any large city has dozens, if not hundreds, of gas stations — all selling gasoline that's pretty much the same. But if you look at each gas station with a little wider scope, you notice that each one sells a product that's at least slightly different from the product its competitors sell. For instance, some stations have mini-marts, and others have car washes or provide fuel with special additives designed to improve engine performance. And crucially, each gas station is clearly differentiated from all the others because it has a unique location — something that's important to people who live nearby or are desperate because their car is running on its last fumes.



REMEMBER

Economists use the term *product differentiation* to describe the things that make each firm's product different from its competitors' products. The overall result of these differences is that they decrease the intensity of competition at least a little bit. Your local gas station, for instance, may be able to get away with charging you a few cents more per gallon than its competitors if it has nice facilities and the next-closest competitor is several miles away.

On the other hand, there's still a lot of competitive pressure in the industry. Although your local station may be able to use its unique characteristics to get away with charging you a little more, it couldn't charge you a lot more — if it tried to do that, you'd take your business to one of its competitors.

Similarly, all the restaurants in your neighborhood have to worry about what the other restaurants are charging even if the others specialize in completely different cuisines. Although you may be willing to pay 20 percent more for something exotic, you wouldn't likely be willing to pay 90 percent more. Product differentiation lessens but doesn't eliminate price competition.

Facing profit limits

You may think that because monopolistically competitive firms can use their unique characteristics to raise prices, they're guaranteed nice profit margins. After all, in pure competition where firms sell the same product and have no way of differentiating themselves from their competitors, prices fall so low that firms end up earning zero economic profits. (See Chapter 7.) If monopolistically competitive firms can raise prices above the competitive price, it seems like a no-brainer that they should be guaranteed to make profits. Unfortunately for them, this isn't the case.



REMEMBER

As pointed out by Cambridge economist Joan Robinson during the 1930s, monopolistically competitive firms still face competition. In particular, they face the prospect that if they're making tidy profits, those profits attract new entrants to their industry. When the new entrants begin producing, they take business away from the established firms and ruin their previously tidy profits. In fact, new entrants continue to arrive until profits have been driven all the way back to zero.

Dealing with downward sloping demand

Joan Robinson was able to show how this process works by slightly modifying the monopoly model that I introduce in Chapter 8. To see what she did, look at Figure 9-3, which shows a single monopolistically competitive firm initially making a profit. The figure shows the firm's marginal cost curve, MC , and average total cost curve, ATC , along with its demand curve, D , and the associated marginal revenue curve, MR .

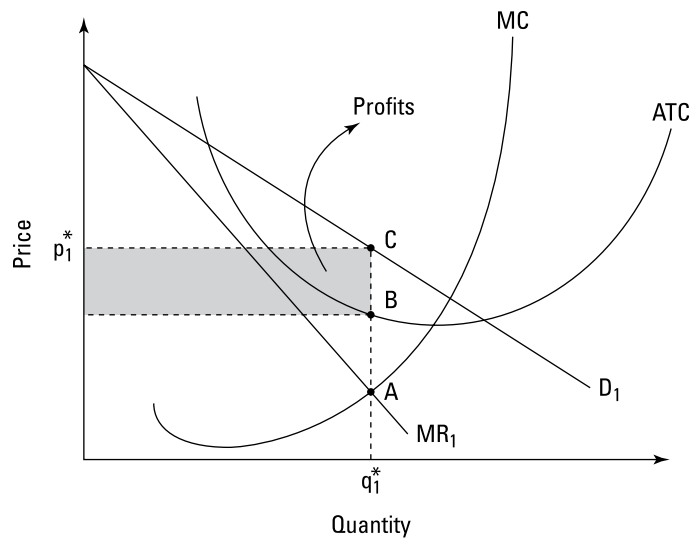


FIGURE 9-3:
A monopolistically competitive firm making a profit equal to the area of the shaded rectangle.

© John Wiley & Sons, Inc.

Because of product differentiation, the firm in Figure 9-3 faces the downward sloping demand curve, D_1 . Its demand curve is downward sloping because, like a monopoly, it has some control over its price. Product differentiation means that the firm can choose whether to set a higher or a lower price. At a higher price, the quantity demanded of its product falls because some customers won't think that the unique characteristics of the firm's product are worth the extra money. At a lower price, the quantity demanded increases because the lower price steals customers away from the firm's competitors.

By contrast, competitive firms that sell identical products have no control over the prices they set. Because competitive firms sell identical products, the only thing that matters to consumers when choosing among them is who offers the lowest price. The result is that all firms have to sell at the same price, the *market price*, which is determined by where the overall industry supply curve crosses the overall industry demand curve. The demand curve for an individual competitive firm's product is a horizontal line at the market price. (See Chapter 7.) This line stands in sharp contrast to the downward sloping demand curve facing the monopolistically competitive firm in Figure 9-3.

An important consequence of the downward sloping demand curve, D_1 , is that the marginal revenue curve, MR_1 , associated with demand curve D_1 is also downward sloping. Why is this so? The additional, or marginal, revenue that the firm can get from selling an additional unit of output is less than the marginal revenue it gets from selling the previous unit.

Declining marginal revenue is a natural consequence of a downward sloping demand curve. Because the only way to get consumers to buy more of your product is to entice them with a lower price, the marginal revenue you get has to fall with every additional unit you sell.

The monopolistically competitive firm optimizes profits by choosing to produce at Point A, where the downward sloping marginal revenue curve, MR_1 , crosses the upward sloping marginal cost curve, MC . Producing the associated quantity, q_1^* , will either maximize the firm's profit (if it's possible to make a profit) or minimize its loss. Whether making a profit is possible depends on the position of the firm's demand curve — on how much demand there is for the firm's product.

In Figure 9-3, demand is strong enough that the firm makes a profit. You can see this by comparing the firm's average total cost per unit at output level q_1^* with its selling price per unit at that output level. The average total cost per unit is found by going up vertically from the horizontal axis at output level q_1^* until you hit the ATC curve at Point B. The price per unit that the firm can charge at output level q_1^* is found by going up vertically until you hit the demand curve at Point C.

Because the distance up to Point C exceeds the distance up to Point B, you can immediately determine that the firm's selling price per unit exceeds the total cost of production per unit — meaning that the firm must be making a profit on each unit sold. The size of the firm's total profit on all units is the profit per unit times the total number of units sold, so its total profit is equivalent to the area of the shaded rectangle in Figure 9-3. The area of the shaded rectangle is the width of q_1^* units times the height of the profit per unit — the vertical distance between Points B and C.

Finding equilibrium: Firm entry and exit

What economist Joan Robinson realized was that profit attracts new entrants to the monopolistically competitive industry. Each new entrant steals some business away from existing firms. Graphically, this means that the demand curve for any existing firm, like that of Figure 9-3, shifts to the left. At each possible price that the firm may charge, it sells fewer units than before because new entrants have stolen some of its old business.

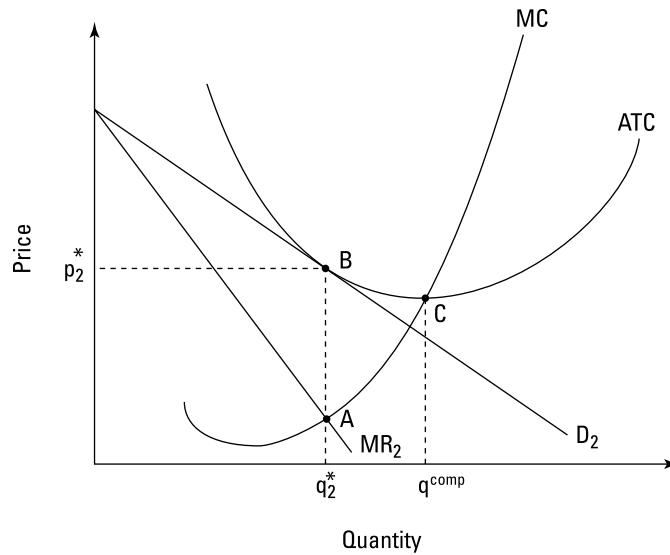


REMEMBER

Furthermore, new entrants continue to enter the industry and shift demand curves to the left until profits are driven all the way down to zero. Only then does the entry of new firms come to a halt.

You can see this sort of equilibrium in Figure 9-4. There, the demand curve has shifted left all the way to D_2 , where it's just tangent to the ATC curve at Point B. As the demand curve moves left, so does the marginal revenue curve, which now lies at MR_2 . Consequently, when the firm optimizes its production level by producing where the MC curve crosses MR_2 , it will produce at output level q_2^* .

FIGURE 9-4:
A monopolistically competitive firm makes zero profits after entry (or exit) has shifted its demand curve until it's tangent with its ATC curve.



© John Wiley & Sons, Inc.

At this output level, profits are zero. You can see this by using the graph to show that the average total cost per unit of producing output level q_2^* is equal to the price per unit that the firm can get selling those units. Go up vertically from the horizontal axis at Point q_2^* to Point B. Because Point B lies both on the demand curve, D_2 , and the average total cost curve, ATC , the vertical distance from the horizontal axis at Point q_2^* to Point B represents both the average total cost per unit as well as the price per unit that the firm can charge. They're equal, so the firm is making zero profits.

If the firms in a monopolistically competitive industry are losing money, some firms will exit the industry. As each of them exits, the remaining firms gain more business, and the demand curves for firms still in the industry shift to the right. Exit continues until the firms reach an equilibrium like that of Figure 9-4 and make zero profits.

Producing inefficiently

The equilibrium in Figure 9-4 implies that each firm produces less efficiently than would firms in a competitive industry. The best way to see this is to compare the monopolistically competitive firm's output level when the industry is in equilibrium, q_2^* , with the output level that would be produced by a firm with the same cost curves that was operating in a fully competitive industry in which all firms sold an identical product. I've labeled this output level as q^{comp} in Figure 9-4.

In Chapter 7, I explain how market forces push competitive firms to produce at q^{comp} and why it ends up being exactly the level of output at which the ATC curve hits its minimum — that is, q^{comp} is the output level at the bottom of the U-shaped

ATC curve. The socially significant implication of this fact is that competitive firms produce at the lowest possible average total cost per unit. That makes them as efficient as possible in terms of minimizing per-unit production costs.

By contrast, a monopolistically competitive firm operating in an industry where product differentiation allows it to have some control over the prices it charges will end up producing at a higher average total cost per unit. This is clearly the case in Figure 9-4 because the vertical distance from the horizontal axis up to Point B is longer than the vertical distance from the horizontal axis up to Point C. That fact implies that the firms in a monopolistically competitive industry are not as efficient as they would be if they were in a perfectly competitive industry.

Some look at this result and conclude that society would be better off if it could transform monopolistically competitive industries into competitive industries. But the cost savings might not be worth the loss of product differentiation. You know what they say about variety being the spice of life. Would you really want every restaurant to serve the same food in the same type of room, under the same lights, with identical furniture? Probably not.

But you have to decide for yourself whether you think the higher costs associated with variety are worthwhile — and in what situations. Although they may be worth it to you for restaurants, you may have a different feeling about the product differentiation found among gas stations.

COMMUNISM, WENDY'S, AND PRODUCT DIFFERENTIATION

One of the funniest TV commercials ever produced was a Wendy's commercial that made its debut in 1987, at the height of Soviet communist power. It depicted a communist fashion show. A woman walked down the runway in a drab gray factory worker's uniform, and the announcer shouted, "Day wear!" Then she marched down the runway again in the same outfit but holding a flashlight. The announcer shouted out, "Evening wear!" Next, she marched down the runway again — still in the same uniform — holding an inflatable beach ball. "Swimwear!"

The commercial made fun of the fact that the central planners who ran communist countries didn't care much about product differentiation. They typically made only one design of any given product in order to be able to mass produce that product at the lowest possible cost. The commercial helped to hammer home to U.S. consumers the idea that they should embrace the fact that the food produced by Wendy's was different from that produced by its main rivals, McDonald's and Burger King. Unlike the rigidly planned Soviet economy, free market U.S. capitalism allows for huge amounts of product differentiation.

3

Applying the Theories of Microeconomics

IN THIS PART . . .

Explore how externalities create socially inefficient outcomes and how societies reckon with the Tragedy of the Commons.

Find out why markets depend on full disclosure, how secrets can ruin a market, and why markets can't always provide public goods.

Understand why so many internet services get delivered for free despite being offered by firms that have massive monopoly power.

Probe the fascinating area of behavioral economics, including evolutionary irrationality, heuristics and cognitive biases, prospect theory, myopia, framing effects, and altruism.

- » Defining an ideal market
- » Seeing how externalities create socially inefficient outcomes
- » Taking steps to fix problems caused by externalities
- » Exploiting and exhausting commonly owned resources

Chapter 10

Property Rights and Wrongs

Economist Adam Smith came up with the *invisible hand*, the idea that even though individuals pursue their own interests, the common good is achieved if you allow markets to allocate resources. Smith was quite aware, however, that you achieve this nice result only if society's property rights are set up correctly before people start to trade goods and services in markets. In fact, he spent a good deal of his most famous book, *The Wealth of Nations*, talking about how governments must properly define property rights if they want markets to yield to socially beneficial outcomes. (Chapter 7 discusses this idea in more depth.)

The gist of the problem is that if property rights aren't set up correctly, a person won't fully take into account how their actions affect other people. For instance, consider two pieces of land. One is privately owned, and the other is land that nobody owns and everyone is free to use as they please. If you want to dump your trash on the privately owned land, you have to pay the owner for the right to do so. (In other words, the owner is running a trash dump.) But, like everyone else, you can dump trash for free on the unowned land because nobody has a legal right to stop you.

Naturally, the difference in property rights with respect to the two pieces of land leads people to dump a lot more on the unowned land because it's less costly personally to do so. But the problem is that although it's less costly personally, lots of costs are imposed on others. For example, what could've been a beloved park is now a heap of rotting garbage. Bad property rights lead to bad outcomes.

In this chapter, I talk about positive and negative *externalities* — situations where one person's behavior results in either benefits or costs to other people, but where the property rights are so badly defined that the costs and benefits aren't properly accounted for. I also show you how most cases of endangered or extinct species are the result of nonexistent property rights and how redefining property rights can save species from oblivion. (See, the dinosaurs were just lacking property rights!)

Allowing Markets to Reach Socially Optimal Outcomes

For markets to achieve socially optimal outcomes, they must plan for all the costs and benefits involved, regardless of who feels the effects of those costs and benefits. If markets do this, the demand curve captures all benefits, the supply curve captures all costs, and the market equilibrium quantity ensures that only units of output for which benefits exceed costs are produced.

Look at Figure 10-1, which shows a demand curve and a supply curve. The market equilibrium quantity is q^* , and the market equilibrium price is P^* . The important thing to realize is that it's socially beneficial to produce every unit up to and including q^* . You can see the reason for this by examining unit q_o . Looking at the demand curve, you observe that buyers are willing to pay price P_o for unit q_o , but it costs suppliers only C_o to produce unit q_o .

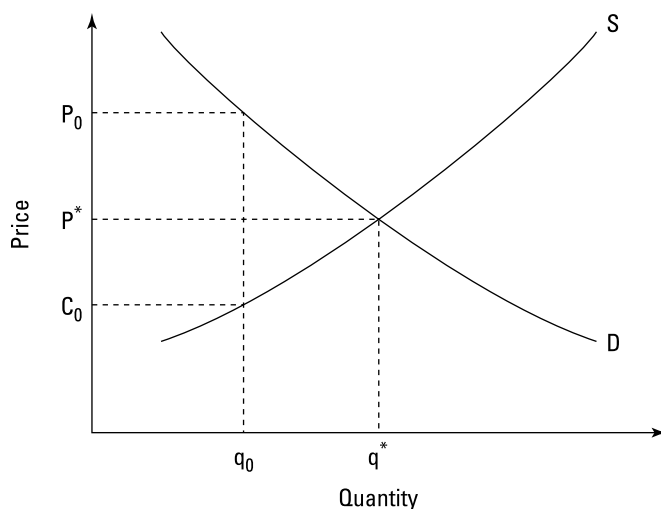
What does this mean? The overall happiness of society is improved by making unit q_o because it's clearly worth more to people to have it than the cost of the resources used in making it. Because the demand curve is above the supply curve for all units up to and including q^* , all those units are socially beneficial to produce.



REMEMBER

The wonderful thing about markets is that supply and demand just happen to cause the socially optimal level of output, q^* , to be produced. The fact that this happens entirely as the result of people pursuing their own selfish interests is, of course, why markets are so amazing. It's as though the invisible hand of some kindly deity magically turns the pursuit of individual goals into a socially optimal outcome.

FIGURE 10-1:
With full and complete property rights, the market quantity produced, q^* , is socially optimal.



© John Wiley & Sons, Inc.

What I show you next is that this nice result happens only if property rights are *full and complete*, meaning that the demand curve captures all the benefits that people are willing to pay for and the supply curve captures all the costs associated with production. As you'll see, if property rights aren't full and complete, markets won't generate socially optimal output levels like q^* . In such cases, the invisible hand turns out to be invisible for the worst possible reason — because it's not there!

Examining Externalities: The Costs and Benefits Others Feel from Your Actions

Property rights give owners control over their property. For instance, I can paint my car any color I want. I can modify the engine or the tailpipe. I can even install big, shiny 20-inch racing wheels to try to disguise the fact that, like most economists, I'm not very hip.

On the other hand, property rights aren't unlimited. Society does restrict what I can do with my car. For instance, I can't be a source of noise pollution by removing the muffler. I can't drive it 90 miles per hour past an elementary school. And it's illegal for me to play my 2,000-watt stereo at full volume late at night. I'm not legally allowed to do these things because I don't live on an island by myself. Rather, I live in a community with many other people, and making lots of noise or driving really fast affects their quality of life. Economists describe this phenomenon as my actions causing externalities.



REMEMBER

An *externality* is a cost or a benefit that falls not on the person(s) directly involved in an activity but on others. Externalities can be positive or negative:

- » **Positive externality:** A positive externality is a benefit that falls on a person not directly involved in an activity. Think of a beekeeper. This person raises bees to sell the honey, but the bees also happen to fly around pollinating flowers for local farmers, thereby increasing their crop yields and providing them with a positive externality.
- » **Negative externality:** A negative externality is a cost that falls on a person not directly involved in an activity. Think of a steel mill that, as a byproduct of producing steel, puts out lots of soot and smoke. The pollution is a negative externality that causes smog and pollutes the air that everyone living near the factory breathes.

This section takes a closer look at externalities.

Noting the effects of negative externalities

Goods and services that impose negative externalities — that is, costs on third parties — end up being overproduced. This happens because people don't consider negative externalities and the costs they impose on others when they decide how much of something to produce.

For example, in the case of a polluting steel mill, the mill's managers consider only their private costs of raw materials and running the plant.

If someone owned the atmosphere, the mill's managers would have to pay for the right to emit pollution. And if the atmosphere were owned by the people who would have to inhale the mill's pollution, the firm would be forced to pay those people for the right to pollute and would be forced to reckon with the harm that the pollution causes them. But because nobody owns the atmosphere and firms don't have to pay to pollute into it, there's no mechanism for making the mill's managers consider the costs of pollution that fall onto members of the broader community. The result is that the firm overproduces steel.

The following subsections explain how negative externalities such as pollution cause overproduction by shifting supply curves. When these shifted supply curves interact with demand curves, you can determine exactly how much overproduction will occur and how much harm will result from each unit that's overproduced.

Looking at differences in private and social costs

In Chapter 6, I explain that a competitive firm's supply curve is equal to its marginal cost curve. If a steel mill doesn't account for the marginal costs that its

production of steel imposes on others, its marginal cost curve (its supply curve) is too low and leads to an overproduction of steel.

You can see this situation in Figure 10-2, where I've drawn in two supply curves. The lower one is labeled *Private MC* because the firm's supply curve is its private marginal cost curve, which factors in only the firm's own costs of producing steel. The higher curve considers not only the firm's private costs but also the external pollution costs, which I've labeled *XC* (for external costs). This higher curve is called the *Social MC* curve and is useful because it captures all costs associated with producing steel — both the firm's costs of making it and the costs imposed on others as the result of negative externalities.

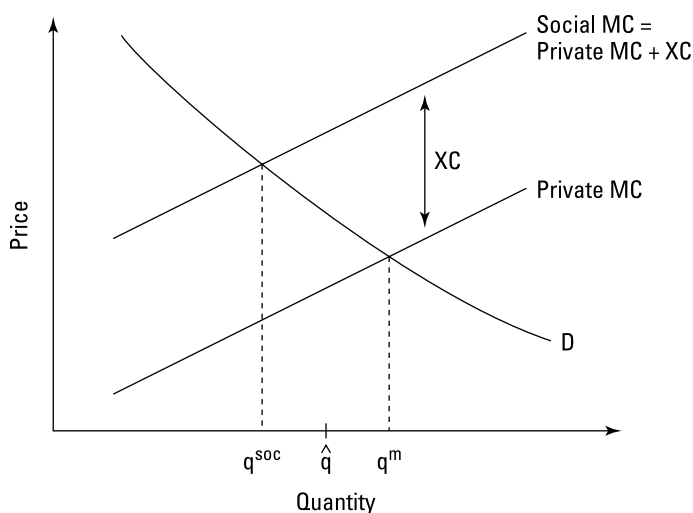


FIGURE 10-2:
The market overproduces a good that has a negative externality.

© John Wiley & Sons, Inc.

Producing units for which costs exceed benefits

Overproduction is the result of the *Private MC*'s curve failure to take account of external pollution costs. In particular, the market equilibrium features overproduction because it occurs where the demand curve intersects the *Private MC* curve rather than the *Social MC* curve.

To see why this leads to overproduction, first examine the market equilibrium in Figure 10-2, where the demand curve, *D*, crosses the *Private MC* curve. That equilibrium results in a quantity q^m of steel being produced, where *m* stands for market.

On the other hand, the socially optimal amount of steel to produce is q^{soc} , where soc stands for social. The socially optimal quantity is determined by where the Social MC curve crosses the demand curve. You can tell that q^{soc} is socially optimal because for every unit up to and including q^{soc} , the demand curve is above the Social MC curve, meaning that the benefits of producing these units exceed the costs of producing them. That's true when considering not only private costs but the external costs that fall on third parties as well.



REMEMBER

The problem with producing all the units from q^{soc} up to q^m is that although the benefits do exceed the firm's private production costs, they don't exceed the total costs when you account for XC, the cost of the negative externality.

For instance, look at output level q^A , which lies between q^{soc} and q^m . You can see by going up from q^A to the demand curve that the price that people are willing to pay for that level of output does exceed the private marginal cost of producing it. (That is, the demand curve is above the Private MC curve at output level q^A .) But if you go up even farther, you see that what people are willing to pay for that level of output is actually less than the total, social cost of producing that much output. (That is, the Social MC curve is higher than the demand curve at output level q^A .)

Output level q^A shouldn't be produced because the total cost of producing it exceeds what people are willing to pay for it. That's why it's unfortunate that output level q^m is, in fact, produced in a market economy. Every unit of output produced in excess of output level q^{soc} is a unit for which the total costs exceed benefits.

Accepting positive amounts of negative externalities



REMEMBER

An important thing to realize is that the common reaction to negative externalities — outlaw them! — is almost never socially optimal. The goal isn't to eliminate negative externalities. Rather, the goal is to ensure that when you weigh *all* costs and *all* benefits, the benefits from the units of output that are produced outweigh the costs of producing them — including the costs of the negative externalities.

In Figure 10-2, for all units of output up to and including q^{soc} , the total benefits are at least as great as the total costs, meaning that society as a whole benefits if these units are produced. Note in the figure that the socially optimal output level q^{soc} is a positive number. That is, it's socially optimal to produce steel even though some pollution will be produced along with it.

To understand the intuition behind this fact, think about internal-combustion automobiles. Gas and diesel cars pollute the air, and the only way to eliminate

their pollution fully and completely would be for society to immediately ban all fossil fuel vehicles.

But would you really want to do that?

It would take a decade, at least, to replace the lost fossil-fuel vehicles with electric or hydrogen vehicles. And until that happened, you wouldn't have enough transportation to move food from farms to cities. You'd also lack enough transportation to provide sufficient ambulances and fire trucks for emergencies. And billions of people wouldn't be able to get to work anymore due to a shortage of transportation.

So while you may be cheering for zero-emissions vehicles to replace gas and diesel vehicles as soon as possible, it has to be noted that the pollution costs currently being imposed on society by millions of those vehicles are more than compensated for by their social benefits. Thus, while we *would* want to immediately take off the road all the vehicles for which total costs exceed total benefits, there will still be tens of millions of vehicles that we would want to “keep on truckin” since their total benefits exceed their total costs.

The same holds true for the pollution the steel factory is producing at output level q^{soc} . The only way to eliminate the pollution from the steel factory is to shut it down. But that means removing from society all the benefits that steel can provide, such as earthquake-proof buildings and crash-resistant safety cages in automobiles.

The next thing to consider is how to make sure that only q^{soc} units are produced when the market wants to overproduce goods with negative externalities.

Dealing with negative externalities

You can deal with negative externalities in three basic ways:

- » **Pass laws banning or restricting activities that generate negative externalities.** For instance, most cities now forbid you to dispose of your trash by burning it.
- » **Pass laws that directly target the negative externality itself (rather than the underlying activity that leads to the externality).** For instance, steel mills are now required to install smokestack scrubbers that filter out most of the pollution before it goes into the atmosphere.
- » **Impose costs, such as pollution taxes, on people or firms generating negative externalities.** For instance, governments can charge companies for each ton of pollution they emit. This solution is appealing to economists

because it's the one that is most likely to lead to the production of the socially optimal output level.

Looking back at Figure 10-2, you can see why economists like pollution taxes. Recall that XC is the external cost of the steel mill's pollution on others. If the government imposes a tax of XC dollars on every unit of steel the firm produces, the tax raises the firm's cost curve up from Private MC to Social MC.

Setting the pollution tax at exactly XC dollars causes the firm's marginal cost curve to lie exactly where the Social MC curve does. Because a firm's marginal cost curve is its supply curve, the result is that when demand and supply now interact, the socially optimal output level q^{soc} is produced. That is, by imposing exactly the right tax on steel, the government can sit back and let the market do the rest. That makes this sort of pollution-reducing policy more attractive than other potential solutions.

Compare this solution with a system where firms are ordered to install smoke-stack scrubbers to reduce pollution. In such a system, you need to hire inspectors to constantly monitor factories to make sure they aren't cheating. This sort of system is much more costly to implement than simply imposing a tax on the mill's easily measured steel output and then letting supply and demand set the socially optimal output level. On the other hand, it may be difficult to figure out exactly how big the tax XC should be, so the pollution tax solution isn't without problems either.

Calculating the consequences of positive externalities



REMEMBER

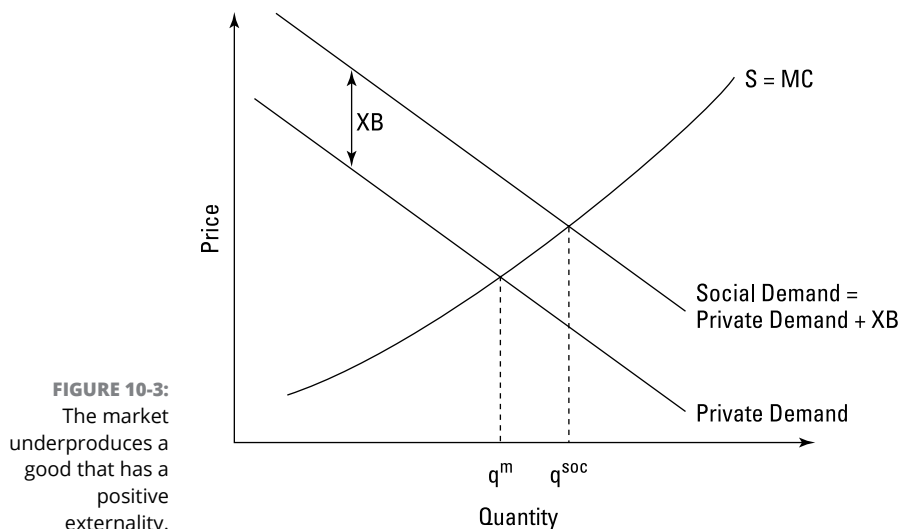
Externalities can be positive as well as negative. The key thing to understand about positive externalities is that goods and services that provide positive externalities to third parties end up being underproduced.

Underproduction is typical for goods that generate positive externalities. Because property rights are set up in such a way that the recipients of the positive externalities don't have to pay for them, the producers of goods that generate positive externalities have no incentive to provide extra units of output for the benefit of those receiving the positive externalities.

To see why goods with positive externalities are underproduced, consider a beekeeper named Kate. Kate raises bees so she can sell the honey and make some money. The people who buy her honey do so because the honey brings them utility when they eat it. But because Kate's bees go around pollinating the flowers of local farmers, these farmers also benefit from her beekeeping activities.

But — and here's the crucial point — the farmers don't pay Kate for the benefits that her bees bring them; the bees just fly in and out of their fields, and there's no way to keep track of them. The result is that Kate is going to raise fewer hives of bees than she would if the farmers were paying her for the benefits that her bees bring them.

You can see how this situation works in Figure 10-3. Kate's supply curve is her marginal cost curve, and I've labeled it $S = MC$. I've labeled the demand for her honey by the customers who pay for it as *Private Demand*. Where the supply curve and the Private Demand curve intersect gives the market equilibrium quantity of honey, q^m .



© John Wiley & Sons, Inc.

But this output level factor in the benefits that bees bring to farmers. Suppose that these benefits have a dollar value of XB , which stands for external benefits. Then the total social demand for Kate's honey is given by the *Social Demand* curve, which is the *Private Demand* curve shifted up by XB dollars because honey production benefits the farmers as well as Kate's honey-loving customers.

The socially optimal output level, q^{soc} , would be where the Social Demand curve crosses Kate's supply curve because for each unit of output up to and including q^{soc} , the total social benefit is at least as great as Kate's cost of production.

As you can see, the market equilibrium quantity produced, q^m , is less than the socially optimal output level, q^{soc} . In other words, because the market mechanism has no way of considering the positive externality, Kate produces less honey than is socially optimal.

Subsidizing things that provide positive externalities

Because markets tend to underproduce goods and services that have positive externalities, people have come up with ways to encourage higher levels of production.



REMEMBER

The most common way to encourage higher production of goods that generate positive externalities is with a *subsidy*, a government payment to an individual or firm to encourage it to produce more of a given type of output. Crucially, the government doesn't receive anything back from the recipient of the subsidy in return for the payment; the beneficiaries are the consumers who receive the additional output.

In the case of Kate's beekeeping business, the government may pay her a honey subsidy of, say, 20 cents per pound to encourage her to keep more hives. The result is more bees pollinating more flowers, leading to higher output levels for the farmers. In fact, the government may even tax the farmers to get the money to subsidize Kate's honey. Doing so would make the program pay for itself.

Similarly, governments also often subsidize the planting of trees in and around cities. They must do so because the markets don't factor in many of the benefits of trees — shade, cooling, cleaner air, less soil erosion, and so. Without the subsidy, fewer trees would be planted than is socially optimal.

Tragedy of the Commons: Overexploiting Commonly Owned Resources

An important economic problem that results from poorly defined property rights that don't consider negative externalities is called the *tragedy of the commons*. This section examines this problem in detail.

Overgrazing on a commonly owned field

The *tragedy of the commons* refers to a resource being overexploited due to the perverse incentives that common ownership create. Under common ownership, the resource is open for all to use as they please. These circumstances make rapid use and overexploitation likely because each person has an incentive to use up as much of the resources as possible before others do.

To understand the tragedy of the commons, think of a farming town in which most of the land is privately owned. However, there's one large field of *common land* where anyone can graze their cattle. In a private field, the owner has an incentive to limit the number of cattle put out to graze. That's because if the owner puts too many beasts in the field, they quickly eat up all the grass and ruin the field for later grazers. Putting out only a few cattle to graze reduces the owner's short-run profits (because the owner restricts the current number of cows) but maximizes the long-run profits (because the field stays in good shape, and the owner can keep grazing cattle well into the future).

With the commonly held field, however, everyone is going to put some cattle out there because the personal cost of doing so is nothing. Nobody has a personal incentive to preserve the field's future usability. The incentives are actually horribly perverse because if the common field is currently lush with grass, your incentive is to put as many of your cattle out there as quickly as possible to eat all the grass before the field is ruined. Everyone else sees things the same way, so there's a mad rush to put as many cattle out to graze as quickly as possible. The result, of course, is that the field is rapidly ruined for everyone. So although there's no personal cost to putting a cow out to graze on a common field, there's a social cost. Each additional cow causes damage to the field that reduces the future productivity of the field.



REMEMBER

The difference between what happens to the private field versus the common field is entirely the result of the different property rights governing the two types of land. In the case of privately owned fields, farmers have an incentive to weigh the costs as well as the benefits of putting more cattle out to graze. In particular, they calculate how much future profits will be reduced if current overgrazing ruins the future usability of the field.

Extinctions and poor property rights

Many environmental problems are caused by tragedy of the commons situations in which nobody owns the property rights to a given resource. Notably, most animal extinctions are the result of an absence of property rights.

For instance, think of tuna swimming in the open ocean. By international treaty, nobody owns the open ocean, which means the tuna swimming in the open ocean are no one's property either. On the other hand, if you catch a tuna and pull it up onto your boat, you have a property right over it and can sell it for money. The only way to economically benefit from tuna is to kill it.

The result is that tuna and many other fish species are hugely overfished, with many near extinction. That's because each angler has the incentive to harvest as many fish as quickly as possible before anyone else. Anglers are aware the

problem that results. But because of the way property rights are set up in this case, no individual angler can do anything to prevent the calamity. If one angler decides to hold back and take fewer fish so the species will survive, someone else will just come in and catch the fish that were spared. The species will go extinct anyway, so no one has an incentive to hold back.

Avoiding the tragedy



REMEMBER

When an economist sees a tragedy of the commons situation, their first instinct is to change the property rights system governing the resource in question. Instead of commonly held property rights in which each person has an incentive to take as much of the resource as possible before anyone else does, economists suggest private ownership so there will be an incentive to preserve the resource. Here are a few solutions:

- » **Area-based property rights:** In the case of overfishing, one solution has been to give anglers private property rights to an entire fishing ground — to all the fish in an area while they're still alive. That gives the new owners the proper incentive to manage the stock on a sustainable basis. Furthermore, because only one person has the right to fish in a given area, there's no longer a mad rush between competing anglers to harvest as many fish as possible before anyone else can get to them.
- » **Permits:** For fish species that migrate freely between different areas, a different solution has been developed. In such cases, biologists first determine the maximum number of fish that can be sustainably harvested each year. The government then auctions off fishing permits for exactly that amount of fish. This method prevents the tragedy of the commons by creating a new sort of property right — the fishing permit. It also creates a self-sufficient government program. The money raised from auctioning off the fishing permits can be used to hire game wardens to prevent unlicensed fishing, as well as for conservation and wildlife management programs.
- » **Local collective management:** Nobel Prize winner Elinor Ostrom studied instances in which the resource users in particular areas were able to prevent the tragedy of the commons by developing local collective management systems that restrained overuse.

Herders sharing pastures in Africa and farmers sharing irrigation water in Nepal have been able to avoid overexploitation by setting up systems in which outsiders can be excluded from exploiting the resource, insiders can be monitored against overuse, mechanisms exist for punishing those who take more than their allotted shares, and collective-choice arrangements allow most insiders to participate in the decision-making that regulates the system.

IN THIS CHAPTER

- » Realizing how markets depend on full disclosure
- » Seeing how secrets can ruin a market
- » Examining the used car and insurance markets
- » Understanding that markets can't always provide public goods
- » Getting public goods from governments and philanthropists

Chapter **11**

Asymmetric Information and Public Goods

Markets provide nearly everything that people consume. But sometimes they fail to provide certain products that people want, or they provide too much or too little of a particular good or service. Economists refer to such situations as cases of *market failure*. In this chapter, I discuss two of the most interesting and common causes of market failure:

- » **Asymmetric information:** This situation occurs when either the buyer or the seller knows more about the thing they're bargaining over than the other party knows. The classic example is a high-quality used car: The owner who's trying to sell the vehicle knows all about the car's high quality and reliability, but the potential buyer can only take the owner's word for it. Because the potential buyer has no reason to trust the seller's assertions that the car is really good, they assume the worst and offers a low price just in case the car turns out to be a low-quality "lemon." But because the owner knows that the car is of high quality, they reject the low offer, and the car goes unsold — all because there's no cheap and easy way to prove the car's high quality to the potential buyer.

» **Public goods:** These kill off markets in a different way. That's because the very nature of a public good makes it extremely difficult for private sellers to charge users. The classic example is a lighthouse. After it's up and running, it benefits all nearby ships, regardless of whether they pay for the service. That being the case, each ship tries to avoid paying for the service in the hope that somebody else will pay for it. But with no one paying, the lighthouse soon goes bankrupt, and society is denied a valuable service.

In this chapter, I discuss these two causes of market failure further, show you how pervasive they are, and describe some of the clever solutions that people have come up with to remedy them. But don't expect any asymmetric information here. I'll make sure that you end up knowing everything I do.

Please note that asymmetric information and public goods aren't the only things that can cause markets to fail. Market failures also result from monopoly (Chapter 8), oligopoly (Chapter 9), and externalities (Chapter 10).

Facing Up to Asymmetric Information

In many real-life situations, buyers and sellers don't share the same information. Regardless of which party is better informed, economists refer to these situations as cases of *asymmetric information* because one side has more information than the other.

For instance, sellers of used cars are much more knowledgeable about the true quality of the vehicles than the buyers are. On the other hand, the buyers of auto insurance policies are much better informed than insurance companies about whether or not the buyers are bad drivers.

In this section, I provide more in-depth information about asymmetric information.

Realizing that asymmetric information limits trade

Asymmetric information is very important in the real world because it limits what goods and services can be successfully traded in markets. If you know that the other person is better informed, you might be afraid that they'll use their information to take advantage of you.

In the case of used cars, buyers are afraid that sellers who know that their cars are bad will keep that fact to themselves and try to negotiate high prices, as if their cars were good. In the case of insurance, insurance companies are afraid that people who know that they're high insurance risks will pretend to be low insurance risks so they can get lower rates.

Depending on how bad the asymmetric information gap is, markets may even collapse completely. That is, if you're worried that the seller of the used car is exaggerating the value of the vehicle that he's trying to sell you, you probably won't buy. That sounds like a reasonable thing to do, but that fear prevents the sale of good cars because everybody's worried about bad cars. Similarly, if insurance companies can't figure out a way to tell the good insurance risks from the bad ones, they may charge high rates to everybody as though everyone were a high risk. And that, typically, causes the low-risk people not to buy insurance at all because they know they're being overcharged.



In extreme cases, asymmetric information can lead to what economists call *missing markets* — situations in which there's no market for a good or service. In these situations, markets are absent because many potential participants have been scared off by the fact that better-informed market participants may use their informational advantage to exploit less informed market participants.

In less extreme cases, asymmetric information only hobbles — but doesn't completely eliminate — a market. I explore how this happens in the next section.

Souring on the lemons problem: The used car market

George Akerlof was awarded the Nobel Prize in Economics in 2002 for a famous paper he wrote called "The Market for Lemons." The paper is all about asymmetric information and market failure, and it was especially memorable because Akerlof used the market for used cars as his primary example. What Akerlof correctly explained was that poor-quality vehicles, or lemons, dominate that market because asymmetric information drives away sellers who want to part with high-quality used cars.

To make the intuition behind that result clear, imagine that there are only three kinds of used cars for sale: good, okay, and bad. They all look the same on the outside and even test-drive pretty much the same, but they have major differences in terms of how much longer they're going to last before the engine gives out. Because of the difference in engine quality and how long the cars are likely to last, the good cars are worth \$30,000, the okay cars are worth \$20,000, and the bad ones are worth \$10,000.



The problem that leads to market failure is the asymmetric information that exists between buyers and sellers. In particular, although each seller knows how good their own car's engine is, the buyers have no way of knowing.

Buyers could, of course, ask sellers to be truthful about the quality of their cars, and no doubt many sellers — probably most — would tell the truth. But there's no way to *know* whether they're telling the truth. Consequently, when a particular seller tells you that their car is good, you're still going to be nervous about being cheated.

As I show you in the following subsections, this reasonable fear causes nearly all good and okay cars to be withdrawn from the market. The result is a used car market that's dominated by bad cars; as Akerlof put it, the used car market ends up becoming “a market for lemons.”

Seeing how quality used cars are driven from the market

Imagine that you want to buy a used car, but you don't want to overpay for it. You know that there are only three types of cars: good, okay, and bad. In addition, you happen to know that one-third of all used cars on the road are good, one-third are okay, and one-third are bad. How much would you be willing to pay for a used car?

Well, given the fact that good cars are worth \$30,000, okay cars are worth \$20,000, and bad cars are worth \$10,000, and also given the fact that you don't know which cars are which, imagine that you'd be willing to pay no more than \$20,000. Why \$20,000? Because that's what the okay used car — the car of average quality — is worth.

Because sellers have no way of proving to you how good their cars are, a sensible thing to do when presented with a used car is to assume that it's of average quality and, therefore, worth \$20,000. So you offer \$20,000. And so do all the other buyers presented with used cars because they, like you, can't tell the quality of used cars apart. Now look at how different sellers react to the \$20,000 offer depending on the true quality of their vehicles:

- » If a seller knows their car is bad and worth only \$10,000, they happily accept your offer.
- » If a seller knows their car is okay, they also accept because you're offering what the car's actually worth.
- » If a seller has a good car, they won't accept unless they're in some sort of dire circumstance. Because they know the car is worth \$30,000, they won't accept your \$20,000 offer unless they're desperate to raise cash in a hurry (perhaps to pay off some gambling debts).

The result is that nearly all the good cars on the market are withdrawn, leaving only bad and okay cars. Now consider how that situation changes what buyers are willing to offer. If all the good cars are withdrawn from the market, there's now a 50/50 chance that a car is okay or bad. In such a case, how much would you offer to pay if you were a buyer? Well, with a 50/50 chance of the car being worth either \$20,000 or \$10,000, you'll probably offer the average of these two values: \$15,000. But when you do, the market becomes even more dysfunctional. After all, how are the sellers of okay cars going to react to being offered \$15,000? They're going to reject the offer and withdraw their vehicles from the market.

The sad result is that with the good cars and then the okay cars withdrawn from the market, the only cars left are the bad ones, the lemons. Because of the asymmetric information problem, the used car market ends up being a market for lemons. Buyers know this, so they offer only \$10,000 for any car on the market. And because only bad cars are offered, sellers accept the \$10,000. So although it's true that bad cars end up being priced correctly in the used car market, no market exists for good or even okay used cars.

That's a problem because both buyers and sellers want to trade good and okay cars, and they would be much happier if they could. But unless some solution can be found to the asymmetric information problem, they're left out in the cold.

Making lemonade: Solutions to the lemons problem

The fundamental issue driving the lemons problem is that the sellers of good and okay cars have no way of convincing buyers that their cars are as good as the sellers know them to be. The whole problem could be resolved if sellers had some way to convince buyers that a good car is, in fact, a good car, and that an okay car is, in fact, an okay car.

In the next four subsections, I discuss several techniques that people have invented over the years to achieve this goal. These methods don't work perfectly, but because they offer a reasonably large amount of reassurance to buyers, buyers are willing to offer enough money to get sellers to part with higher quality automobiles.

BUYING INFORMATION FROM TRUSTED THIRD PARTIES

Several companies, like Carfax, specialize in selling vehicle history reports to people looking to purchase used cars. Potential buyers can enter the vehicle identification number (VIN) of the car they're interested in and obtain a report that notes mileage, reported accidents, service and repair visits, the number of owners the car has had, whether the car has ever been leased, and so on.

Although these vehicle history reports aren't guaranteed to be exhaustive or perfectly accurate, they're an inexpensive method by which potential buyers can help themselves identify problem vehicles. The reports reduce the likelihood of market failure by helping make information more symmetric between buyers and sellers.

OFFERING A WARRANTY

One way a seller can convince a buyer that she really has a good car is to offer the buyer a warranty. The seller knows that her good car won't break down after she sells it, meaning that she'll never have to pay for any repairs.

A warranty is convincing because only the seller of a good car would be willing to offer a warranty. By contrast, the seller of a bad car would never offer a warranty because they know that their car will probably break down and that they'll have to pay for the repairs. Consequently, if someone is willing to offer you a warranty, they almost certainly have a good car. That's why you see so many used-car dealers offering warranties on the vehicles they sell. If they didn't offer warranties, the lemons problem would quickly take over, and prices would fall so low that only bad cars would be bought and sold on the used car market.

BUILDING A REPUTATION

Another way to solve the lemons problem is to reassure buyers by setting up a market in such a way that sellers can build a reputation for honesty and fair dealing. That's why most good used cars are sold through used car dealers rather than directly between individuals.

Compare a used car dealership with an individual selling their used car online. Who has more of an incentive to tell the truth about car quality? The used car dealer makes their living selling used cars, so if they overcharge one customer by pretending a bad car is good, they get in trouble. When that car starts to break down, the buyer becomes angry and uses social media to tell all their friends that the dealer cheated them. And that loss of reputation cuts into the dealer's future sales. In fact, the dealer will quickly go bankrupt if they develop a reputation for lying.

By contrast, an individual selling their used car doesn't have to worry about developing a reputation for lying because their main source of income isn't selling cars. If they cheat you and you get mad and tell all your friends, it won't affect them much at all. An individual has much more of an incentive to lie than does the used car dealer, who has to worry about their reputation. As a result, many good used cars are sold through used car dealers. (But even at used car dealers, people still need some reassurance, which is why most used car dealers also offer warranties.)

GETTING AN EXPERT OPINION

Because the heart of the lemons problem is asymmetric information, another method of resolving the problem is for skeptical buyers to hire an expert who can give them the information they need to distinguish good, okay, and bad cars. Many car buyers employ this strategy when they doubt a seller's honesty.

For a relatively small fee, a buyer can hire a disinterested third-party expert — for instance, a professional mechanic — to inspect the vehicle and make a list of repairs that will most likely be needed in the near future. In this way, the buyer can get a better picture of the car's quality and what a fair price would be.

This method doesn't fully resolve the asymmetric information problem because the expert probably can't discover *everything* that may be wrong with the car. To the extent that this is true, buyers may still be suspicious, and there may still be some potential for market failure. That's why you often see buyer-initiated inspections used in conjunction with other methods of resolving asymmetric information, such as warranties and sales by dealers who have a reputation to protect.

Issuing insurance when you can't tell individuals apart

An insurance company faces an asymmetric information problem of its own. Its problem is that the people buying insurance know more than the company does about the risks they face.

Consider automobile insurance. Who needs it more: good drivers who hardly ever get into accidents, or bad drivers who get into lots of accidents? Now, clearly, even good drivers want insurance because they're sometimes involved in accidents for which they're not to blame. But bad drivers want insurance even more to help pay for all the accidents they're likely to cause from poor driving.



REMEMBER

An asymmetric information problem faces the insurance company because although individual drivers know whether they're good or bad, the insurance company can't easily tell them apart. If the insurance company *could* tell them apart, it would simply charge the good drivers a low rate for insurance and the bad drivers a high rate.

But because it can't tell the good and bad drivers apart, the insurance company runs a serious risk of going bankrupt. To see why, imagine that the insurance company offered the same low rate to every driver, as though they were *all* good drivers. This would soon lead to bankruptcy because the insurance company wouldn't be collecting enough in premiums to pay off all the damage the bad drivers cause.

To avoid bankruptcy, the insurance company could go to the other extreme: charging everyone as though they were bad drivers. But then the good drivers wouldn't bother buying insurance because it would be overpriced for them. Only bad drivers would sign up for insurance.

This is a poor result socially because you want everyone to be able to buy insurance at a rate that fairly reflects their driving ability. Good drivers should be able to get insurance at a fair rate. And because good drivers make up most of the drivers in the real world, insurance companies will lose out on lots of potential profits unless they can figure out a way to separate the good drivers from the bad.

Grouping individuals to help tell them apart

Insurance companies have come up with a paradoxical way of dealing with the fact that they can't tell whether an individual is a good or a bad driver. Instead of focusing on the individual, they look for clues about the individual based on the groups they belong to. Doing so often gives the insurance companies a pretty good idea about whether the individual is a good or a bad driver.

For instance, government data indicates that males under 25 get into many more serious accidents than females under 25 do. So if two people walk into an insurance company and one of them is a 23-year-old male and the other is a 22-year-old female, chances are that the male is a worse driver than the female. Consequently, you charge the male a higher insurance rate.

This situation ensures that everybody can get insurance at what is *likely* to be a fair price given the fact that, on average, males under 25 get into many more accidents than females under 25.



TIP

In reality, this nice little result isn't the compelling reason behind insurance companies' decisions to infer as much as possible about their customers by looking at which groups they belong to. These companies really have no choice; competition *forces* them to do so.

Why is this true? Consider two insurance companies, only one of which uses group membership information to set rates. The company that doesn't use group information has to set high rates because it's afraid that all its customers may be bad drivers. Doing so drives away the good drivers who don't want to pay bad-driver rates for their insurance.

But the company that uses group information can offer multiple rates, such as high ones to young men and lower ones to young women. Doing so allows it to capture the business of many good drivers who don't want to deal with the first

insurance company that sets only one high rate for everyone. The result is that insurance companies are always looking for ways to estimate an individual's unknown risk profile based on the well-known risk profiles of the groups they belong to. (For a discussion of whether this is fair or not, please see the section titled, "Group discrimination, individual identification" later in this chapter.)

This process can lead to some rather unfair conclusions. The oddest is that good-driving young males end up paying higher rates than good-driving young females because the only thing insurance companies have to go on is gender. But such a system is still better than the even more unfair alternative in which *all* good drivers would have to pay bad-driver rates, which is what would happen if insurance companies were banned from using group membership information to try to distinguish their customers. The closer insurance companies can get to fully distinguishing good and bad drivers using group-membership information, the fairer rates will be.

Keep in mind that the drivers for whom insurance companies have the greatest need to use group-membership information are new drivers. Because insurance companies don't have accident or violation records for new drivers, there's a pressing need to try to separate the good from the bad drivers using group-membership information. As drivers gain more experience, the insurance companies can get increasingly accurate accident and violation information that distinguishes the good from the bad.

Avoiding adverse selection

Using the groups to which an individual belongs to estimate their insurance risk goes only part of the way to resolving the asymmetric information problem that exists between insurance companies and their customers. Obviously, there's still a lot of individual variation within any group. For instance, even if young women are, on average, better drivers than young men, some young women are bad drivers. This leads to a difficult problem known as *adverse selection*.

If an insurance company sets a premium for young women on the basis of how often young women *on average* get into accidents, insurance will be more attractive to young women who are really bad drivers than to young women who are really good drivers.



REMEMBER

As a result, young female bad drivers will be more prone to sign up for insurance than young female good drivers. This tendency is known as *adverse selection* because it's as though the bad, or adverse, insurance risks self-select into buying insurance policies. The result is a customer pool that contains a disproportionately high number of bad drivers.

Adverse selection is a difficult problem because it feeds on itself. The insurance company has to raise rates to account for bad drivers being more likely to sign up than good drivers. But when it raises rates, the problem just gets worse because the higher rates make insurance even less attractive to good drivers, meaning that the pool of applicants is going to be even more disproportionately dominated by bad drivers.

Mitigating moral hazard

A big problem facing insurance companies is called *moral hazard*. Moral hazard arises because buying insurance tends to change people's behavior. For instance, if I *didn't* have car insurance, I'd drive much more carefully, knowing that I'd have to use my own money to pay for any damage I cause. But because I *do* have car insurance, I drive faster and more recklessly knowing that if something goes wrong, the insurance company is going to be stuck with the bill. (Please know that when I say "I" in examples like this one, I don't actually mean me. You see, I am above moral hazard.)

Car insurance companies deal with moral hazard by offering discounts in exchange for high deductibles. For instance, if I get into an accident, the \$1,000 deductible that I've chosen means that of any bills that ensue from an accident, I have to pay the first \$1,000.

GROUP DISCRIMINATION, INDIVIDUAL IDENTIFICATION

The idea of grouping individuals to help sort them extends beyond insurance. For instance, companies want hard-working employees but can't tell when you walk in for an interview if you are, in fact, hard-working. So they try to estimate the chances that you're hard working by seeing which groups you belong to.

For instance, nearly all straight-A students are hard-working. Therefore, if you're a straight-A student, a company is going to be much more likely to hire you. You may actually be lazy, but the firm improves its odds by hiring those who have good grades.

The practice of using information about the groups to which an individual belongs to try to figure out personal characteristics is referred to as *statistical discrimination*. Although this practice typically improves economic outcomes, you have to decide for yourself whether — and in what cases — you think it's fair or unfair.

The deductible serves as a strong inducement for me not to give in to moral hazard and drive recklessly. And because the insurance company knows that my high deductible gets rid of most of my moral hazard problem, it's willing to offer me insurance at a lower rate than if I'd opted for only a \$100 deductible. Deductibles are a clever way of reducing moral hazard problems and helping to make insurance more affordable for responsible drivers.

Providing Public Goods

Public goods are products that can't be profitably produced by private firms because there's no way to exclude nonpayers from using them. The inability of private firms to profitably produce public goods derives from the fact that they have two special characteristics. Public goods are both nonrival and nonexcludable:

- » **Nonrival:** Nonrival means that one person's using the good doesn't diminish another's ability to enjoy the good. Think of an outdoor fireworks display, a statue in a park, or a television show broadcast over the airwaves. Your consumption doesn't in any way diminish that of others. This stands in stark contrast to most goods, where if you consume more, less remains for others. (Think of cookies.)
- » **Nonexcludable:** Nonexcludable means that it's hard to prevent nonpayers from consuming a good or service. For instance, when you produce an outdoor fireworks display, everyone in the vicinity gets to see it for free no matter how much you'd like to charge them for it. A more serious example is an army: When it's in place to provide national defense, it provides national defense for everybody, including those who don't want to contribute to the cost of the army's maintenance. A more futuristic example is asteroid defense. If any single person or national government pays for a system to redirect or destroy asteroids that are on a collision course with the Earth, there would be no way to prevent nonpayers from being just as saved as those who were willing to contribute to the costs of setting up such a system.



REMEMBER

The nonrival and nonexcludable characteristics of public goods make it hard for private firms to make any money producing them. Think about trying to get people to buy tickets to an outdoor fireworks display. Because people know that they're going to be able to see it for free, they won't buy tickets. Because they won't buy tickets, there's no way to raise the money needed to put on the display.

This sort of chicken-and-egg problem is frustrating because although people don't want to pay for something they can get for free, they actually do like fireworks displays — meaning that they're fundamentally willing to pay something to see them. The problem is figuring out how to get them to pay.

Taxing to provide public goods

The most common solution to the problem of how to provide public goods has been for governments to step in and use tax money to pay for them. In the case of fireworks, because nearly everybody likes them, there's no problem getting enough political support for spending tax money on fireworks displays. And after the taxpayer has funded them, everyone can enjoy them.

The government has historically provided national defense because it, too, is very much a nonexcludable, nonrival public good. For instance, because protection from foreign invaders is nonexcludable, there's always a temptation not to help pay for it because you know that if someone else does pay, you get to enjoy safety from foreign invaders for free. And because national defense is nonrival, you know that the safety you enjoy is of just as high a quality as the safety everyone else enjoys. This fact, too, lessens your incentive to pay. As a result, governments force people to share the expense for national defense by levying taxes.

Taxes and a good portion of government spending are often derided as wasteful (and often are wasteful), but keep in mind that in many cases, taxes are the only way to fund the wide variety of public goods you enjoy. Although nobody likes taxes, you probably wouldn't want to trade a reduced tax burden for no public parks, no national army, no public roads, no public sewers, and so on. Without the government's ability to force people to pay for these things, you'd likely not have them — at least not in the quantity and variety that you currently enjoy.

Enlisting philanthropy to provide public goods

Although most public goods are paid for through government taxation, some are paid for privately. So don't think that governments are absolutely necessary to provide public goods. They aren't. However, governments are a much more reliable way to provide public goods because you don't have to rely upon the philanthropic largesse of the rich, who are under no obligation to spend their wealth on public rather than private goods.



WARNING

Please don't make the common mistake of thinking that public goods are called *public* because they're typically provided by the government rather than the private sector. Economists call them public goods because private firms can't *profitably* produce them, not because the government *has* to produce them. Private philanthropy can produce public goods without government help.

For example, in the middle of Los Angeles, where I grew up, there's a huge and beautiful mountain park called Griffith Park. A millionaire named Griffith

J. Griffith donated the land for it. He provided a public good at his personal expense.

The desire for fame or recognition may help encourage the philanthropic provision of public goods. Rich people often trade million-dollar donations to hospitals, universities, and charities for having a building named after them. And corporations often sponsor cultural public goods such as fireworks displays, arts festivals, and parades so they can obtain name recognition and goodwill.

SUPPORTING FREE INFORMATION WITH ADVERTISING

Google is currently the world's largest and most popular internet search engine. But when it first came on the scene, it lost money hand over fist. Its problem was that it was providing internet search results for free, as a public good. People went to google.com, typed in a search term, and got ranked results without having to pay anything. This service was helpful to users, but Google would've soon gone bankrupt if it hadn't stumbled onto a way to generate income from its services.

The solution was the same one that broadcast TV stations and newspapers had long used. Google would charge advertisers for the right to have their ads appear next to Google's free search results. Thus, for instance, if a person were using Google to search for the word *dog*, the free search results would be accompanied by ads paid for by companies selling products related to that word — for instance, hot dogs, dog food, and concert tickets for Snoop Dog.

Google's search ads had an unintended casualty: They began to kill off newspapers. This happened because Google's targeted search ads were much more cost-effective for advertisers than newspaper print ads. For example, if a company that wanted to sell dog collars took out a full-page ad in the *New York Times*, only a small percentage of that paper's readers would've had any interest in that ad because the majority of them would've had no current need to buy a dog collar. But if that same company placed its ads only next to Google search results that were generated for people who had been searching for "dog collars," then nearly every person who saw the ad might be in the market for a dog collar.

In the decade after Google introduced search ads, newspaper advertising revenues fell by more than half, and hundreds of newspapers went bankrupt or were forced to merge with other newspapers to survive. Newspapers have attempted to adjust to the internet by publishing online and trying to capture online and mobile ad revenues. But because the news agencies' ads aren't as well targeted as those accompanying Google search results, they don't command very high rates.

Providing a public good by selling a related private good

Broadcast radio is a public good. After a radio signal has been sent out over the airways, it's nonrival: My listening to a radio program doesn't reduce anyone else's ability to tune in. It's also nonexcludable: There's no way to stop anyone with a radio from tuning in. So given that radio programming is a public good, how is it that so much radio programming is produced and broadcast every day by privately owned and operated radio stations?

The answer is that the broadcast industry figured out that although radio itself is a public good, the commercials that they play on the radio are private goods that radio stations can charge a lot of money for. That is, if a car maker or beer maker (or the publisher of a revolutionary new economics book by yours truly) wants an ad to be broadcast to the millions of listeners who tune in for free to the public good known as radio, the company has to pay for commercial airtime.

The trick behind broadcast radio is that the privately sold good called advertising pays for the freely provided public good called radio. To a more limited extent, many newspapers and magazines work the same way. Although they raise some money from subscription fees or the newsstand price, a substantial chunk of their revenue comes from the advertising they sell.

Ranking new technology as a public good

You live in an age of rapidly rising living standards. Why is this the case? Because institutions are fostering the creation of new and better technologies that allow people to produce more goods and services from the same resources or to produce entirely new goods and services that were previously impossible to produce.

Technological progress is a public good. And because it is, society has had to come up with ways to make sure that technological progress happens given the fact that private individuals and firms have little incentive to invent new technology.

Patenting to turn public goods into private goods

To understand how new technologies are public goods, consider the invention of the movable-type printing press by Johannes Gutenberg in 1435. Before Gutenberg, books were copied by hand. But after he invented the printing press, it was much cheaper to make new copies by printing them. Furthermore, think about how simple the technology really is. The printing press is basically just a big version of the rubber stamps that kids like to play with. The invention was immediately understandable to anyone and everyone who heard about it, which meant that they could make their own printing presses as soon as they heard about it.

So how does this invention satisfy the characteristics of a public good?

- » It's nonrival because my building and using a printing press doesn't in any way lessen your ability to build and use a printing press.
- » It's basically nonexcludable because the cost of communicating the new idea to another person is so low — just a short conversation does the trick.

The result is that unless society creates some sort of an institution to reward the creation of new ideas, there's not going to be much of a profit incentive to go into the invention business. In fact, what happened to Gutenberg was that everyone copied his idea and didn't pay him for it. So unless you can come up with a way to financially reward the creation of new inventions, you're not likely to get many of them.



REMEMBER

The solution has been the creation of *patents*. By giving inventors the exclusive right to market and sell their inventions for 20 years, patents provide a financial incentive to get people to invest the time and energy necessary to come up with new technologies that benefit everyone. It's no coincidence that economic growth took off after government-enforced patents became widely available in the United States, and then Western Europe, starting in the late 18th century.

Subsidizing research into technologies that can't be patented

Not every new innovation can be patented. That's because you can patent only something you invent, not something you discover. For instance, if you think up a chemical that's never existed before and then synthesize it, you can patent it. But if you merely discover an existing chemical that's been floating around the sea or lying in the soil, you can't patent it.

This is a big problem for things like cancer research because many potential cures are chemicals derived from plants and animals — chemicals that have existed in nature for eons. These chemicals have huge potential benefits, but because they can't be patented, nobody has a strong financial incentive to try to discover them.

As a result, the government and many private philanthropic groups fund research into areas of science where the public goods problem would otherwise limit research. This solution is important to an economist because providing public goods is an economic problem that markets and the invisible hand can't fix. Other types of market failure, like asymmetric information, have pretty decent private sector solutions (as I discuss in the section "Making lemonade: Solutions to the lemons problem," earlier in the chapter).

- » Understanding how digital versions of economies of scale drive internet oligopolies
- » Explaining why so many digital products are given away for free
- » Noting the power of network effects
- » Explaining how digital platforms decide whom — and how much — to charge

Chapter **12**

Internet Economics: Networks and Platforms

The digital economy is dominated by large oligopoly firms like Apple and Google that display an economic behavior that was entirely unknown before the advent of the internet: that of a major firm giving away its main product for free.

This phenomenon has been a boon for consumers, who can access and enjoy an almost incomprehensibly wide variety of business and entertainment services at no cost. But it has, simultaneously, been quite confounding for government anti-trust (antimonopoly) regulators who are normally very concerned whenever major companies dominate their respective industries.

If U.S. Steel, for instance, controlled 93 percent of the entire market for steel production in the United States, regulators would aggressively investigate whether it was using its monopoly power to restrict output so it could drive up the price of steel and grant itself a large, undeserved profit (as monopolies often do, as I discuss in Chapter 8).

By contrast, the fact that Google processes 93 percent of all internet searches outside of China generates much less consternation because it gives away those search services for free. There's also no evidence that Google restricts output because anyone with an internet connection can run as many Google searches as they desire at zero cost to themselves.

The dominance of leading software, search, and social media companies like Microsoft, Google, and Facebook in their respective sectors of the internet also raises questions about why so many internet industries are dominated by leading products and how those products maintain their dominance despite well-funded, technologically advanced rivals.

Economists believe that the answers to those questions can be found by examining (1) the *economies of scale* available to many internet firms and (2) the *network effects* present in the cost structures of many internet businesses. Working in tandem, economies of scale and network effects can drive digital products to dominant positions that can be difficult for rivals to contest. They can also explain why so many internet services are free to users.

Examining Economies of Scale among Internet Businesses

Many prominent internet companies, such as Apple, Google, and Microsoft, have a tendency toward *natural monopoly*, the situation in which *economies of scale* are so strong that a single firm can produce the industry's output at a lower cost per unit than would be possible if more than one firm produced the industry's output.

This inclination toward a natural monopoly in the realm of internet companies can be largely ascribed to the substantial fixed costs related to product development. This is combined with the minuscule marginal costs associated with expanding services to accommodate an ever-growing user base. As an unavoidable outcome, these digital corporations are privy to significant economies of scale. The average total cost per user diminishes as the number of individuals using their products escalates. This characteristic economic landscape is becoming synonymous with the sphere of digital technology.

High Fixed Costs

The products produced by digital firms involve large costs that must be incurred *before* those products are offered to the public. Nobody wants to play a half-finished video game or shop on a half-built website that crashes every three minutes. Nor do

they want to stream an incomplete music track or an unfinished movie. Digital products must be complete and functional before someone uses them. Thus, the costs of developing digital products are fixed costs with respect to the number of eventual users because they have to be incurred whether a digital product ends up having no users, fifty users, or a billion users.

High *fixed costs* also apply to software maintenance and upgrades, and even to advertising costs. For example, Snapchat may spend \$150 million developing a new feature or \$60 million on a massive online ad campaign, but those activities might not attract new users. They may even repel some current users. Consequently, Snapchat's business model is dominated by fixed costs that are independent of the number of people using its service.

Low Marginal Costs

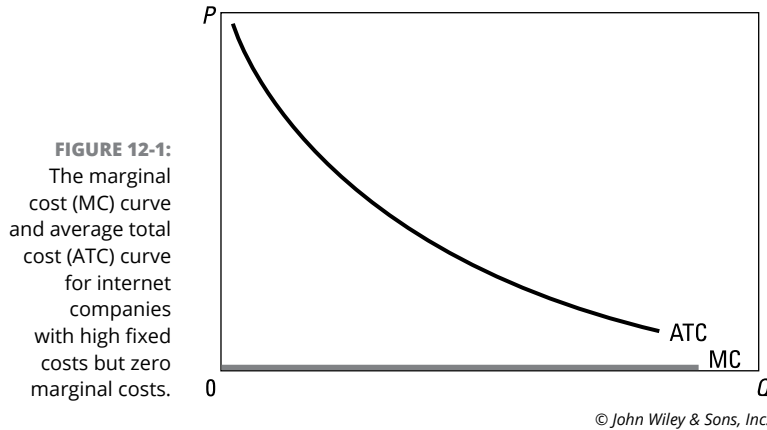
An internet firm's marginal cost of onboarding and providing service to an additional user is extremely low for three reasons:

1. Internet businesses are built on software, and software doesn't degrade or depreciate once it is built out. You can run the same code over and over again — forever! — without hurting it or having to pay for maintenance.
2. The only marginal cost associated with running a piece of computer code one additional time is a small amount of electricity that costs practically nothing thanks to the low cost of electricity in the modern world.
3. The marginal cost of transferring data across the internet is practically zero thanks to massive previous investments in data transfer capacity (bandwidth) in the form of fiber-optic cable, satellite relays, cell phone towers, and so on.

By contrast, the marginal cost of providing goods or services in a traditional, non-digital business is typically quite high. Building an additional Tesla automobile, for example, requires plenty of tangible physical inputs, such as steel, computer chips, and plastic. It also requires substantial amounts of human labor despite the mechanization of factories. And then the finished car still needs to be transported from factory to showroom on trucks, trains, and transport ships. The marginal cost of traditional goods and services is thus quite high relative to the zero or near-zero marginal cost of digital products, all of which are services.

As an example of the extremely low marginal cost of providing a digital service, consider Twitter. Because it already owns massive server capacity and has paid for the current version of its software, its marginal cost (MC) of servicing one more user on that current software “build” is practically zero, or $MC = 0$. But Twitter *did* have to pay for all of that server capacity and for building out the current version of its software. Those high fixed costs in conjunction with its $MC = 0$ for servicing

additional users imply that the average total cost (ATC) and marginal cost (MC) curves for Twitter and similar internet firms will tend to look like those in Figure 12-1.



Graphically, the key consequence of high fixed costs combined with zero or near-zero marginal costs is that the ATC curve starts at a high level when the quantity of users is low and declines rapidly as the quantity of users increases. Thus, ATC starts high on the left side of Figure 12-1, where the number of users is low, before falling rapidly as the quantity of users increases.

We can generalize by saying that if all or nearly all of an internet product's costs are fixed costs, then ATC will decline continuously as the number of users increases and those fixed costs are spread over more and more users. To get a grip on this dynamic, a throwback to Chapter 6's cost curve examination is in order. The fundamental takeaway is that as long as MC stays below ATC, the latter has no choice but to diminish, as argued in that chapter. This inherently implies that if MC is zero, as represented in Figure 12-1, ATC will relentlessly reduce toward zero as user numbers grow. Consequently, economies of scale will persist, regardless of the magnitude of the user base.

Pricing, profits, and elasticity

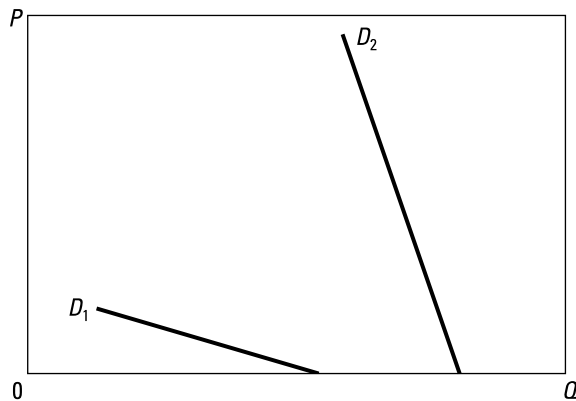
You might think that having ATC sliding steadily toward zero would be a golden ticket for any internet mogul. After all, the sweet smell of success (read profit) for these tech wizards is nothing more than a game of numbers — the gap between what they charge per user and the average total cost per user. Because ATC keeps dropping, shouldn't there be a guaranteed "pot of gold at the end of the rainbow" for these firms? Shouldn't it be the case that if they could get enough users to hop

on their platform, they could sit back, relax, and watch ATC slide below the price tag they put on their services? Surely, that's a recipe for raking in the profits, isn't it?

Sadly, things aren't nearly that easy. A profit will be achieved only if firms can manage to do two things simultaneously. First, they have to gain enough users so that ATC per user falls to a low level. Second, they have to get all those users to pay a price that exceeds ATC per user. For instance, an online newspaper may need to have 15,000 subscribers before the ATC of publishing the newspaper falls below \$3.99 per user per month. But if the newspaper tries to charge \$4.00 per user per month to make a small profit, will all of its subscribers stick around?

For many internet businesses, the answer is no. The problem is that although the number of users may be high when the product is given away at a price of zero ($P = 0$), the demand curve for most internet products is quite flat, implying that the quantity demanded is highly elastic (price-sensitive) at any positive price, as with demand curve D_1 in Figure 12-2. Consequently, any attempt to raise the selling price much higher than zero will cause the producer to lose many or even most of its users — and thus any hope of spreading its large fixed costs over a large number of users. (For a quick overview of elasticity, see Chapter 4.)

FIGURE 12-2:
If a firm's demand curve is flat (highly elastic), as with D_1 , then raising the selling price above $P = 0$ won't reduce the quantity (Q) of users very much. But if the demand curve is steep (and thus rather inelastic), like D_2 , then the selling price can be raised above zero without losing many users.



© John Wiley & Sons, Inc.

However, not every internet business faces this problem to the same extent. Internet firms that enjoy *network effects* (which we discuss later in this chapter) may have steeper demand curves that imply that the quantity demanded is relatively *inelastic* (not-price sensitive) at any positive price, as with demand curve D_2 in Figure 12-2. Thus, if they raise their prices above zero, fewer users will quit.

But even network effects are no guarantee that a digital firm will have substantial pricing power (that is, a strong ability to raise price without losing many customers). In fact, many popular internet services have highly elastic demand. Examples include internet search services (Google, Bing, Yahoo!), information reference sites (Wikipedia, Dictionary.com), and news and current event sites (*New York Times*, *Wall Street Journal*).

The question then arises: If internet firms with highly elastic demand can't rely on being able to raise their prices above zero to cover their costs and at least break even, how can they survive? They've developed several strategies, sometimes using two or more of these strategies simultaneously.

Selling Advertising

One strategy for covering the high fixed costs of product development without charging prices that might chase away users is to sell advertising within an internet product. Google, for example, sells targeted advertisements that are tailored to a user's search requests. Advertisers pay Google handsomely for targeted ads, and the revenue that Google makes from selling targeted ads more than covers the cost of providing search results for free to the billions of people who use Google's search engine.

Selling Data

Many internet businesses make money by selling data about their users to other internet businesses. For example, if you're on a website looking at engagement rings, you're probably planning to get married relatively soon. The jeweler can sell that information to companies offering wedding-related services so that they can target you with their ads.

Employing Price Discrimination

Many internet firms use various forms of price discrimination to cover their costs without chasing away a lot of users. *Price discrimination* occurs when a producer charges different prices to different groups of users based on differences in their average willingness to pay. The key to successful price discrimination is getting users to voluntarily self-select into different groups whose members are willing to pay varying amounts for the same or similar products:

» **Freemium pricing** involves giving away a basic product for free and then charging users to unlock a set of additional features. This strategy is useful for maintaining a large network of users (those who are satisfied with the free version) while bringing in revenue from those who are willing to pay for the additional features.

- » **In-app purchases** are a variation on freemium pricing that's popular with video game developers, among others. Under this pricing model, video games are given away for free, but revenue is obtained from the sale of digital items that enhance or extend game play. *Fortnite* is a good example. Playing *Fortnite* is free, but players can pay for items such as special equipment and weapons to outfit the digital characters that they control in the game. By making the basic game free, *Fortnite* has built a massive user base, retaining even those players who have an extremely high price elasticity of demand.
- » **Subtractive versioning** occurs when users have the option to pay to have something removed from the free version of a product. The most common example is the removal of advertising. Consider regular YouTube versus YouTube Premium. Regular YouTube is free, but it comes with ads. In contrast, users who are willing to pay a monthly fee for YouTube Premium can enjoy all of YouTube's audio and video content without ads. This tiered-pricing model allows YouTube to retain users who have a high price elasticity while generating revenue from those who dislike advertising strongly and are willing to pay to avoid it.

Pricing and competition at zero MC

A key consequence of (1) economies of scale, (2) zero marginal costs, and (3) highly elastic demand is that internet firms face substantial pressure to keep prices low or zero whenever they face competition. This is especially true for internet firms that offer similar products. You can find hundreds of smartphone apps for playing chess, fighting zombies, or doing yoga. Each firm tries to deliver a slightly different experience, but, in the eyes of many consumers, their products are largely interchangeable. As a result, the competitive equilibrium involves each firm keeping the price of its basic product low or zero so that it won't lose users to competitors. The firms then attempt to generate revenue in other ways, such as by selling ads.

However, not all internet firms have a lot of competitors. As noted earlier in this chapter, internet behemoths like Google and Facebook are oligopolists that dominate their respective sectors of the internet. That dominance might seem to provide them with enough market power to charge a price well above zero. As we'll see, however, the large majority of internet oligopolists don't use their market power to charge high prices for their main product. Instead, they usually opt to provide their main product for free and then rely on advertising or other alternative revenue streams to generate income. The reason they do that has to do with network effects, as we explain in the next section.

Noticing Network Effects

In economics, a *network* is a group of people or objects connected to each other by a service or by a piece of infrastructure that facilitates flows of goods, services, or information. Examples of the connectors that stitch various networks together include roads, water pipes, waterways, 5G cellular networks, Amazon.com, WhatsApp, computer operating systems, wi-fi, social media apps, video game platforms, and the internet itself.

A network effect is said to occur when the value of a network depends on the number of people connected to it. Network effects can be either positive or negative.

- » **Positive network effects**, or positive network externalities, are present when a network's value to each user *increases* as more people join the network. Consider dating apps. The more people who join a particular dating site, the more chance each user has to find someone compatible. The larger the number of users, the more valuable the site becomes to each user. The same is true for review sites like Yelp. The more people who write restaurant reviews on Yelp, the more valuable Yelp becomes to each user.
- » **Negative network effects**, or negative network externalities, occur when a network's value to each user *decreases* as more people join the network. One source of negative network effects is network congestion, which happens when too many people attempt to use a network at the same time. As one example, think of how the download speeds of wi-fi networks slow dramatically if too many users try to stream data simultaneously. *Network pollution* is another type of negative network effect. Most commonly observed on social networks, network pollution occurs when the presence of too many users degrades the average quality of the information that users obtain from the network. If, for instance, too many people are connected to you on Facebook, your Facebook news feed may end up becoming highly polluted by irrelevant or undesired content from people you barely know. Other things equal, the larger the number of people who are connected to you on Facebook, the more likely your news feed is polluted.

Network effects are demand shifters, meaning that when they change, they'll cause demand curves to either shift left (a decrease in demand) or shift right (an increase in demand).

- » Positive network effects make a good or service more desirable, thereby increasing demand and shifting the demand curve to the right, all other things equal.

» Negative network effects make a product less desirable, thereby decreasing demand and shifting the demand curve to the left, all other things equal.

Also note that positive and negative network effects may exist at the same time.

Understanding how network effects enhance market power

Positive network effects are important because they can generate monopoly power for any firm that develops a popular network that has many users.

Consider Facebook, which is the world's largest social networking site, with around 3 billion regular users. Facebook has many rivals, but if one of your main goals is to connect with as many people as possible, you have an incentive to join Facebook over any other social network because those rival social networks have fewer users and thus generate smaller positive network effects.

The advantage that Facebook's 3 billion users gives it in terms of positive network effects implies that Facebook is somewhat insulated from competition in the social networking industry.

Although definitely not a single-firm monopoly, Facebook is the dominant firm in an oligopoly industry in which its only substantial rivals are firms that have targeted specific social networking niches — such as business networking in the case of LinkedIn — or have developed substantially different product experiences — like Snapchat, with its automatically erasing posts, or Instagram and TikTok, which focus on allowing users to quickly share and rapidly comment on (Like!) single photos or short videos.

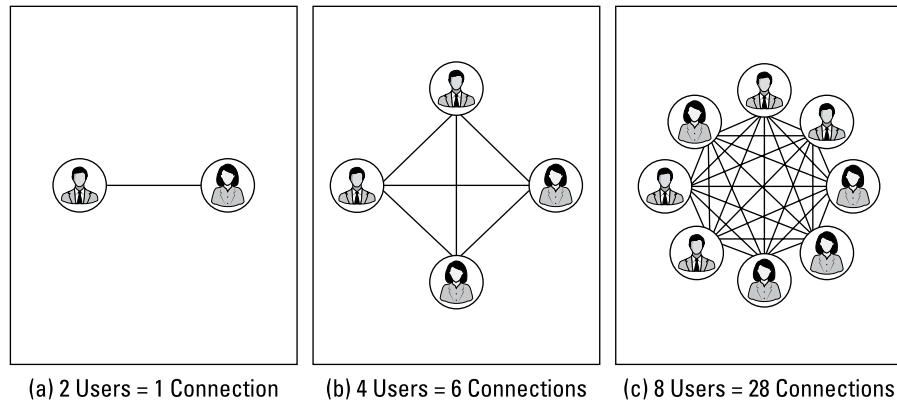
Thirteen online social networks had 500 million or more monthly active users in 2021. Several, including Facebook Messenger, WhatsApp, and WeChat, can be categorized as more or less “pure” messaging services. Others, such as TikTok, YouTube, and Snapchat, are built more around the sharing of video or picture content. This diversity in terms of user experience reflects these platforms' ability to compete for users based on *product differentiation* rather than merely on network size. If network size were the only thing that mattered, all the various networks' users would migrate to Facebook because it has the largest network and thus offers the largest number of possible connections. But billions of people have accounts on multiple social networks, largely because they appreciate the differences between the various networks in terms of interface design, featured content, page layout, and the demographics and interests of other users.

In addition, many social networking users have accounts on multiple networks, a behavior known as multihoming or multitenanting. Their willingness to use multiple networks is another competitive restriction facing social networking firms, which can't rely on positive network effects alone to attract new users or keep existing users actively engaged.

Visualizing the exponential scaling of network effects

The ability of positive network effects to shield internet firms from rivals is supercharged by the fact that the strength of a network effect grows exponentially as additional users join a network. Figure 12-3 illustrates this point. When a network contains only two users, only one possible connection can generate benefits, as shown in Figure 12-3a. But if a network has four users, as in Figure 12-3b, six connections are possible. And with eight users, as in Figure 12-3c, 28 connections are possible.

FIGURE 12-3: The number of network connections for two, four, and eight users. The number of connections in a network grows exponentially faster than the number of people or objects connected to the network.



© JOHN WILEY & SONS, INC.

A simple mathematical formula captures how quickly the number of connections in a network grows as the number of users increases. For any number of users n , there will be exactly $n(n - 1)/2$ connections. If you plug in some additional values for n , you'll notice that the number of potential connections grows much faster than the number of users. As a result, the size of the network effect grows much faster than the number of users even if not every connection is actually used. Thus, if Facebook has twice as many users as a rival social network, the relative value of Facebook's network won't simply be twice as large. Instead, it'll be many times larger, so long as Facebook can limit or prevent negative network effects such as network congestion and network pollution.

THROWING SAND IN THE GEARS

Tech giants don't take it lightly when their patrons play the field, splitting their loyalty between rivals or, heaven forbid, flirting with several products at once. These cyber juggernauts pull all sorts of tricks to keep their users on a tight leash, dishing out digital aggravations galore to sour the taste of any competing service. This is the world of fabricated network-switching costs, designed to make customers think twice before defecting to a competitor's turf.

Take a gander at your favorite music streaming titans — Apple Music, Spotify, Pandora. They're having none of it when it comes to carting your playlists over to the enemy lines. And those email powerhouses like Microsoft Outlook? They have their emails locked down tighter than Fort Knox, using proprietary archiving systems that keep your old correspondence from seeing the light of day on any other service. Even social media czars like Facebook play the game, putting a wrench in your plans to ship your photos and files to the competition.

These intentional frictions are a form of competition, but not all competition favors consumers, and these examples of “throwing sand in the gears” are prime examples of competitive behavior that hurts, rather than helps, consumers.

Facebook has worked hard to avoid negative network effects by rapidly increasing its computing power to avoid network congestion and by using algorithms and privacy settings to mitigate the network pollution that would otherwise cause people's news feeds to be dominated by junk that they don't want to see. By taking these steps, Facebook has ensured that its users enjoy the positive network effects of being connected to the world's largest social network without having to endure very much in the way of network pollution or network congestion. The result is a situation in which Facebook's large size and attention to user experience give it a substantial competitive advantage over its smaller rivals.

Deconstructing Digital Platforms

Amazon.com, Monster.com, and Xbox gaming consoles may not seem to have much in common, but they all serve as *digital platforms* that facilitate interactions between two or more distinct but interdependent groups of users.

- » Amazon.com has two distinct sets of users: buyers and sellers.
- » Monster.com links three distinct sets of users: people looking for jobs, employers wishing to hire, and advertisers looking to reach members of both groups.

» Xbox has two distinct sets of users: the gamers who play video games on the Xbox and the game developers whose video games are played on the Xbox.

Other examples of digital platforms include Uber and Lyft, which allow ride-providers and ride-seekers to interact; Facebook, Twitter, and Instagram, which allow individuals to interact with each other and with advertisers; the Visa and Mastercard networks, which allow consumers and merchants to interact to complete payments; YouTube, Spotify, and Vimeo, which allow media consumers to interact with media creators and advertisers; and search engines Google, Bing, and Baidu, which facilitate interactions between advertisers and people seeking search results.

Balancing user engagement against the need to charge fees

The economic potency of digital platforms lies predominantly in their ability to catalyze interactions between diverse cohorts of individuals. Consider the video game industry, for example, where Sony's PlayStation console is a marquee offering. Its success is deeply intertwined with the presence of game developers who produce content for its specific platform.

These game developers are primarily motivated by a vibrant community of gamers who demonstrate a keen desire to utilize PlayStation consoles for their gaming needs. Similarly, gamers are naturally drawn to platforms that offer them a rich tapestry of games to engage with, a factor heavily dependent on the participation of developers.

Consequently, Sony's fiscal strategy hinges on maintaining a robust cycle that caters to both these groups. It's in their best interest to strive for a teeming ecosystem of PlayStation users, complemented by an equally robust ensemble of developers creating games for the console.

For Sony, the ideal situation is a vast, balanced network of both parties. Gamers are enticed by PlayStation's variety of offerings, leading to a fervent affinity for the platform. Concurrently, developers are captivated by the prospect of millions of potential customers for their wares, fostering a preference for Sony's platform as their development base.

The general takeaway is that a digital platform will have an incentive to maximize the size of each of the groups using its platform, so that the sheer size of each group will increase the value of being on the platform to the members of the other groups using the platform. However, it's often difficult or impossible to maximize the size of all the groups at once. In particular, a platform has to figure out a way

to make money to cover its costs and, hopefully, make a profit. That necessity implies having to charge fees or prices to one or more of the groups using the platform, which means that the platform must act prudently because any group that's charged a fee will tend to use the platform less, which in turn will make the platform less attractive for the other groups using it.

As a concrete example, suppose that Sony decides to charge gamers a fee. That fee will discourage at least some gamers from buying PlayStation consoles, which in turn will make game developers less eager to invest the money necessary to make games for PlayStation because there will be fewer PlayStation users to sell their products to. Thus, Sony has an interesting economic problem. It has to charge fees or prices to a least one of the groups, but it also has to consider not only the reaction of the group(s) being charged but also the reactions of all the other groups using the platform.

Please note that this economic problem isn't specific to Sony, PlayStation, or video game platforms. Rather, it's a challenge that faces *all* digital platforms. Thus, in the next section, we use a different platform — Facebook — to demonstrate how platforms think about pricing and profit maximization when they know that they must consider how the fees or prices charged to one group of users may affect the behavior of other groups of users.

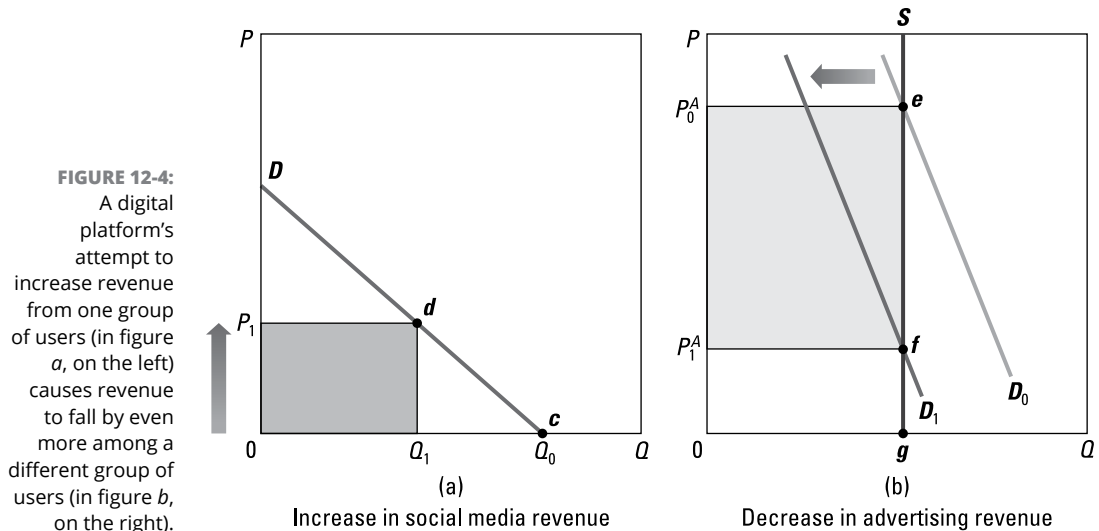
Parsing platform profit maximization

Facebook offers its social networking platform free to users; to compensate, it makes its money by charging fees to advertisers. Indeed, “free for users but costly for advertisers” has been Facebook's pricing strategy since it got started in 2016. If it wanted to, though, Facebook could change things up and start charging its social networking users a monthly fee.

Raising the Price for One Group of Users

Figure 12-4a illustrates the consequences that result from Facebook's decision to begin charging its social networking users a positive amount of money. Those consequences derive from Facebook's initial decision to increase the price of using Facebook from 0 to P_1 on the vertical axis. This price increase reduces the number of social networking users from Q_0 to Q_1 , which we can think of as an upward movement along demand curve D from point c to point d . Associated with this upward movement is the fact that the revenue (= price per user times number of users) that Facebook collects from its users will increase from zero before the price increase to an amount equal to the area of the rectangle OP_1dQ_1 after the price increase.

The increased revenue that Facebook gets from its users might look like pure gain, but Facebook needs to consider the fact that the decline in the number of Facebook users from Q_0 to Q_1 in Figure 12-4a will affect the behavior of another group that uses Facebook's platform: advertisers. Specifically, advertisers will be less willing to advertise on Facebook once they find out that the number of people who will be looking at their ads will be falling from Q_0 to Q_1 . That decrease in advertisers' willingness to pay shows up in Figure 12-4b as a leftward shift in the demand for Facebook ads. Because Facebook gets to decide on how many ads are displayed to viewers each day regardless of the price that it charges for ads, the supply curve for ads, S , is vertical in Figure 12-4b. Thus, when the demand for ads shifts left from D_0 to D_1 , the equilibrium price of Facebook ads falls from P_0^A to P_1^A .



© John Wiley & Sons, Inc.

Note that when the demand for ads shifts left, the amount of Facebook's advertising revenue also declines, from the area of rectangle OP_0^Aeg to the area of rectangle OP_1^Afg . The dollar amount of the decline in advertising revenue is therefore equal to the difference between the areas of those two rectangles, or the area of the rectangle, $P_1^AP_0^Aef$.

We can determine whether Facebook's decision to raise the price that it charges its users from 0 to P_1 increased its overall profit by comparing the size of the revenue gain rectangle in Figure 12-4a with the size of the revenue loss rectangle in Figure 12-4b. Because the area of the revenue gain rectangle OP_1dQ_1 in Figure 12-4a is less than the area of the revenue loss rectangle $P_1^AP_0^Aef$ in Figure 12-4b, it's immediately obvious that Facebook reduced its overall revenue by raising the price of social networking from 0 to P_1 .

This decline in overall revenue also implies a decline in Facebook's overall profit. Why? Because Facebook is operating a digital business that has extremely low marginal costs for adding or subtracting users. Thus, we can think of Facebook's costs as not changing at all while it experiences the overall revenue decline illustrated in Figures 12-4a and 12-4b. The result is a lower profit for Facebook because overall revenue declines while costs stay the same.

To summarize: Facebook's new policy increased the amount of revenue that it received from its social networking users but prompted an even larger decline in advertising revenue. Thus, Facebook would have been better off sticking to its former policy of not charging its social networking users anything.

Two general points derive from our analysis of Figure 12-4.

» ***Understanding that a platform can't look at each group of users in isolation.***

Instead, the platform has to consider the complicated reactions and interactions that exist across the distinct groups using its platform. Only then can it create a pricing policy that maximizes its overall profit.

» ***Realizing that a platform may have a strong financial incentive to set a price of zero for one or more of the groups using its platform.***

This explains why consumers enjoy free social media apps (Instagram, Snapchat, Facebook), free news and information sites (Wikipedia, the Weather Channel), free business productivity apps (Google Workspace), and free email, messaging, and video chat applications (WeChat, WhatsApp, Hotmail, Zoom). In each case, a platform has decided that the way to maximize profit is by offering consumers a price of zero for its main product.

Comprehending How the Sizes of Elasticities and Other Responses Matter

When looking at Figure 12-4, we should keep one other thing in mind. The *sizes* of both groups' responses to Facebook's decision to start charging its users a fee matter quite a bit. To see why, suppose that in Figure 12-4a the demand curve D had been steeper and thus more inelastic at every possible price. In that case, the decline in the number of users would have been smaller when Facebook raised the price from 0 to P_1 . That smaller decline in the number of Facebook users would then have resulted in a smaller leftward shift in the demand for ads in Figure 12-4b, which in turn would have resulted in a smaller decline in Facebook's ad revenue.

If the responses of both groups to the price increase had been smaller (due to a lower price elasticity of demand on the part of Facebook's social networking users and a smaller leftward shift in demand by advertisers), then Facebook might have been able to increase, rather than decrease, its overall profit by raising the price from 0 to P_1 .

What you should gather from that possibility is that the profit maximization problem faced by digital platforms is both complex and opaque. In actual practice, firms won't know ahead of time how strongly each group on the platform may react to any given change in prices or fees. As a result, accurate predictions probably won't be possible. The upshot is that pricing strategy is, to a significant extent, a matter of trial and error for many digital platforms, which, naturally, may be reluctant to deviate from what they're already doing if what they're already doing has been profitable so far.

Regulating the Digital Economy

Many internet industries are dominated by just a handful of companies, which suggests that governments should be concerned about monopoly power and should regulate those industries to promote competition and reduce efficiency losses.

INTERNET CONSUMER SURPLUS

In a remarkable testament to the value of the internet's everyday tools, economists have calculated the immense consumer surplus that digital platforms generate. These statistics underscore the impressive figures.

- The digital encyclopedia, Wikipedia, is perceived by Americans to be worth an average of \$150 annually per person. Extrapolating this to the national scale, it suggests a collective consumer surplus of an astonishing \$50 billion annually.
- Facebook, the dominant social networking platform, fares even better. It's estimated to be worth around \$576 annually per user in the U.S. With 260 million regular domestic users, this translates into a total consumer surplus of approximately \$149 billion each year.
- Internet search, however, tops the list. The average American asserts a value of up to \$17,000 annually for these indispensable services, leading to a staggering consumer surplus of \$5.6 trillion nationwide.

These astronomical figures shed light on the cautious approach by government regulators in dealing with internet behemoths. Breaking up a company like Facebook, for instance, could foster competition within the sector. However, it could also drastically cut the consumer surplus that millions enjoy as part of the world's most extensive social network.

On the other hand, the majority of the most popular internet products are priced at zero, which suggests not only that no monopoly power is being exercised but also that consumers are receiving a huge amount of *consumer surplus*, or the gain that people receive when they can buy a good or service at a price that's lower than what they were willing to pay. (For all the details on consumer surplus, see Chapter 7.)

In addition, many internet firms appear to be more than happy to offer their main products for free to as many users as care to sign up for them — a fact that argues strongly against any *efficiency losses* from underproduction because such losses occur only when a firm with monopoly power restricts output to drive up price. With no such restrictions occurring, we shouldn't have to worry about output being less than the socially optimal amount at which $MB = MC$.

These contradictory facts — of high market concentration on the one hand and unlimited output provided at $P = 0$ to consumers on the other — have left both economists and government regulators debating whether and in what ways internet businesses should be regulated. The overall tendency has been to concentrate on business-to-business issues rather than consumer issues. After all, with consumers getting so many products and services for free, there seems to be little reason for regulators to step in to protect them. By contrast, the leading digital products and platforms may be exploiting their market power to hurt competitors. It's that later possibility that has drawn the most attention from regulators.

Engaging with the Microsoft and Google cases

In 2000, Microsoft was found guilty of using the dominance of its Windows operating system in the personal computer market to give its own web browser, Internet Explorer, a significant market advantage over competing web browsers. Microsoft had “persuaded” computer manufacturers such as Dell and Compaq to install Internet Explorer as the default web browser on their computers. That put rival browsers like Netscape Navigator at a huge disadvantage because, to gain users, they had to persuade existing Internet Explorer users to quit using Internet Explorer after they'd gotten used to it.

More recently, European Union antitrust (antimonopoly) regulators declared Google guilty on three separate counts: using its dominance in internet search services to favor its own online marketplace; using its dominance in search advertising to coerce advertisers into not placing ads on rival ad networks; and using the dominant position of its smartphone operating system, Android, to favor various Google services and apps over similar products offered by rivals.

Today, government regulators continue to examine internet firms for various forms of anticompetitive behavior, most of it related to firms using their dominant positions in one area (such as operating systems or search) to coerce other business into favoring related products or services produced by the same firm. Other areas of concern include major internet firms using their deep pockets to buy up potential rivals (as Facebook did when it acquired Instagram in 2012), fixing prices, and coercing retailers into selling their products on only one market platform rather than on several.

Distinguishing between winner-take-all and contestable industries

When economists began examining the economics of the then-new internet in the 1990s, there was substantial concern that economies of scale and positive network effects would lead to winner-take-all industries in which a single monopoly firm would come to dominate its industry and enjoy an almost uncontested control over that industry such that any potential rivals would be destined to go bankrupt. Once established, economists suggested, these unassailable monopolies would start charging high monopoly prices even if they had previously given their products away for free.

With three decades of hindsight, those fears appear to have been largely unfounded. $P = 0$ pricing continues to be the norm for the most popular internet products despite the existence of dominant firms. And while economies of scale and network effects do create substantial market power, the typical outcome is multifirm oligopoly rather than single-firm monopoly.

In addition, even highly concentrated internet industries such as social networking appear to be contestable industries, meaning that upstart firms are able to grow their user bases and gain market share even when dominant incumbent firms already enjoy economies of scale and positive network effects. Indeed, even Facebook (founded 2004) was once an upstart against Myspace and Friendster, the leading social networks of the early 2000s. The more recent ability of Instagram (started 2010), Snapchat (started 2011), and TikTok (started 2016) to compete against Facebook also argues in favor of most internet industries being contestable industries in which the sheer size of the leading firm's user network is not an insuperable barrier to competition.

The contestability of social networking and other internet industries appears in large measure to be the result of two phenomena that we've already discussed:

- » **Product differentiation.** TikTok and Snapchat, for example, offer a different social networking experience than Facebook does. This product differentiation gives them the opportunity to contest the social networking market despite Facebook's substantially larger userbase.
- » **Multihoming.** The users of a particular digital product typically have the opportunity to use one or more competing products *at the same time*. People are free, for instance, to employ several different email or social networking apps simultaneously. That ability to "multihome" makes for substantial contestability even when a leading product has a large number of users.

Dealing with concentration

One problem that product differentiation and multihoming don't solve is that of takeovers and mergers. As already noted, Facebook purchased Instagram, a rival social network, in 2012. These types of acquisitions limit competition and increase industry concentration. But, as noted previously, regulators may not step in because if both services are still offered for free to consumers and if consumers can multihome between them, it's not clear that such mergers harm consumers at all.

Breaking Up Is Hard to Do

An important related issue is whether government interventions to break up dominant digital firms into multiple competing businesses would be able to reduce industry concentration in the long run. To understand the problem, note that if an industry has strong positive network effects, it will have a tendency to become highly concentrated, all other things equal. Thus, any attempt to break up the leading firm(s) into multiple smaller competitors would probably be futile in the long run because the positive network effects present in the industry would tend to reconcentrate the industry as time passed. The leading firm(s) might not be the same as before, but overall industry concentration might hardly budge in the long run.

Focusing on Specific Practices

Because breaking up digital firms that have positive network externalities may not promote competition in the long run, government regulators have fewer options than they have when regulating nondigital industries, in which such breakups have historically proven to be highly effective in permanently reducing industry concentration. Given this constraint, regulators have focused not on breaking up digital firms but, instead, on regulating specific anticompetitive practices, as noted earlier.

IN THIS CHAPTER

- » Defining irrationality as systematic error
- » Explaining how evolution delivered a brain prone to systematic error
- » Understanding heuristics and cognitive biases
- » Probing prospect theory
- » Focusing on myopia and time inconsistency
- » Assessing the evidence for fairness and non-self-interested behavior

Chapter **13**

Behavioral Economics: Investigating Irrationality

Remember that Victorian-era historian named Thomas Carlyle? He was the one who skewered economics as “the dismal science” in the late 19th century. Fortunately for everyone, behavioral economics popped onto the scene in the late 20th century. Within a few short years, the dismal began to dazzle.

Explaining the Need for Behavioral Economics

Conventional 20th-century neoclassical economics makes many accurate predictions about human choice behavior and how it responds to financial incentives and incrementally changing prices. But when the decisions involve uncertainty

and require the chooser to risk or commit or trust, neoclassical predictions often fail.

The key underlying problem is that real people are often irrational. That's problematic for neoclassical economics because neoclassical economics *assumes* that people are rational. Because rationality is at the heart of why neoclassical economics sometimes fails, let's begin our review of behavioral economics by precisely defining rationality.



REMEMBER

Economists define *rationality* as decision-making that avoids systematic errors.

A *systematic error* is an error that you make over and over, as if you can never learn from your mistakes. Rational decision-makers wouldn't be subject to systematic errors. They'd learn from their mistakes and figure out how to get what they want at the least cost and effort. Outside factors might still derail things, but anything that rational decision-makers could have done to maximize their chances of success would have been learned and applied.

If people were always rational, then standard, mid-20th century neoclassical economics would have always generated reliable predictions about human decision-making. But people regularly and repeatedly engage in behaviors that reduce their likelihood of achieving what they want. They engage in systematic errors. Behavioral economics attempts to explain these systematic errors by combining insights from economics, psychology, and biology. The goal is to develop theories that can deliver more accurate predictions about human choice behavior, including all the irrational stuff.

Decades of research have allowed behavioral economists to develop theories that can explain why our brains employ error-prone mental shortcuts, why we don't save enough for retirement, why we fall for marketing gimmicks, and why we're satisfied with smaller chocolate bars during an inflation. Armed with those insights, behavioral economists have been able to develop beneficial correctives in some cases. And, least dismally of all, behavioral economists have found extensive evidence that people are *not* purely self-interested.

Complementing Neo-Classical Economics with Behavioral Economics

People aren't *always* irrational. They're sometimes very rational, and as a result, neoclassical economic models sometimes make good predictions about human decision-making. That's why you should think of behavioral economics and

neoclassical economics as *complementary* methods for trying to explain why people make the choices that we observe.

At the same time, it's important to realize that the two sets of theories are often radically different both in content and in what they imply about how we should approach the world. In particular, they imply unique strategies for helping people improve their situations.

When people *are* being rational, a simple way to help them is to offer them more options. After all, if they always act rationally and select the best option from whatever is available, they can only be made better off by offering them more options. Thus, neoclassical economists often try to solve consumers' problems by presenting them with additional options.

But when people are making systematic errors, adding more options is unlikely to make things better. It may even worsen the situation because a systematic error would now have more options to play with. As just one example, if you have an impulse-eating problem, your life isn't improved if your mom sends you a package of cookies in addition to a package of brownies. That's why behavioral economists focus on helping people make better choices from already available options.

Behavioral economists also tend to assume that complex decision-making situations will tend to be best explained by a combination of neoclassical economics and behavioral economics.

Let's analyze how people shop at a supermarket:

- » Neoclassical economics tells us that *incentives matter*. Customers care about prices. When prices go up, they buy less. When prices go down, they buy more.
- » Behavioral economics gives us the additional understanding that people tend to buy what they happen to see. This behavior is called *impulse buying*, and it falls outside of the neoclassical notion that the consumer calculates marginal utilities and compares prices. It also explains why grocery stores put the milk cooler at the back of the store — to force you to walk past lots of other items and thus be more likely to see something tempting and make an impulse buy. And it's why they stock candy bars by the checkout lines — to tempt you with an impulse buy just as you're probably quite hungry from looking at everything else in the store.

But how did we end up with brains that need both rational, neoclassical models as well as irrational, behavioral models to explain? The answer is four billion years of evolution.

Examining Our Amazing, Efficient, and Error-Prone Brains

The most complex object in the known universe is the human brain, which has approximately 100,000,000,000 neurons sharing around 10,000 times as many connections. This incredibly dense web of computing power allows us to observe our environments, think creatively, and interact with people and objects.

But despite all that processing power, the brain is quite prone to glitches. This is the result of four billion years of evolution having to deal with the fact that the world was a tough place for our ancestors. Evolving bigger, smarter brains helped our ancestors survive, but bigger brains were costly in terms of calories. The modern brain, for instance, represents just 5 percent of your body weight but consumes about 20 percent of all your calories. So although there was evolutionary pressure to develop bigger brains, there was also a strong limiting factor in that finding enough calories to feed a bigger brain was difficult before agriculture, fire, and modern hunting tools.

In response to the caloric constraint, evolution did two things: To the extent possible, it evolved a bigger (and hungrier!) brain. But it also generated many low-energy mental shortcuts, or *heuristics*, to get the most decision-making power out of each ounce of brain tissue.

Deciphering heuristics



REMEMBER

Because they're shortcuts, heuristics aren't the most accurate mental-processing options. But in a world where calories were hard to come by, a low-energy "good enough" heuristic was better than a "perfect but costly" alternative. In economic terms, there were diminishing returns to deploying additional units of brain power. Heuristics developed because the opportunity cost of perfection was too high.

The brain employs heuristics for nearly every type of action and decision we make. Three examples will give you the general idea about how these mental shortcuts economize on calories and calculations:

» **The gaze heuristic:** You're a centerfielder, and the batter hits a baseball in your general direction. If you lock your eyes on the ball and run so that the ball remains at the same angle above the horizon that it was when you first locked your eyes on it, then, provided you're fast enough, you'll end up running wherever you need to be to catch the ball. This is the method people actually use when catching fly balls. They don't use physics equations or an

app to predict where to run. They simply lock their gaze and run to keep the angle constant. It's a simple solution to a difficult problem.

- » **The steering heuristic:** You're riding a bicycle and moving forward. If you start to fall and you reflexively steer in the direction you're falling, you'll generate enough centrifugal force to stay upright long enough to steady the bike. This is what novice bike riders subconsciously learn to do: If falling to the right, steer right; if falling to the left, steer left. Will this basic rule always work? Not on ice, it won't. And not if you're standing still. But if you're moving forward, it gets the job done nicely indeed.
- » **The recognition heuristic:** You're asked to guesstimate whether Munich or Stuttgart has the larger population. Most people don't know enough about Germany to give an informed answer to this question. But they'll often give the correct answer by choosing the city with the more recognizable name. The recognition heuristic often works because there's a positive relationship between population size and how often cities are mentioned in books, conversation, and social media.

Heuristics help us make decisions rapidly and at low cost. But the fact that they're hardwired means that expecting to avoid or unlearn them isn't practical. For the most part, we're stuck with them and prone to do whatever they suggest.



WARNING

This hardwiring can leave us vulnerable to being played by those who understand heuristics and are willing to use them against us. On the other hand, we can be helped by those who would like to place us in situations in which heuristics tend toward desirable outcomes. But if we cynically consider which of those two alternatives is more likely, we can see the wisdom of shoring up our defenses with more information about how our brains are structured and how their decision-making systems work.

Deconstructing brain modularity

The modern human brain has evolved with a modular structure, so specific areas deal with specific sensations, activities, and emotions such as vision, breathing, and anger.

Located in the back of the head, adjacent to the entry of the spinal cord into the skull, the older parts of the brain control subconscious activities like breathing and sweating as well as automatic emotional reactions such as fear and joy.

Located up front, near the forehead, are the newer parts of the brain that allow creative thought, imagination about the future, and keeping track of your social network. (Mark Zuckerwho?) These newer areas are largely under conscious control.

The differences between the older and newer brain modules allow us to think about decision-making as being controlled by two interrelated systems that were given really creative names: System 1 and System 2.

- » **System 1** is controlled by the older brain modules at the back of the head. System 1 produces quick, unconscious reactions and “gut instincts.”
- » **System 2** is controlled by the newer brain modules up front. System 2 works more slowly, “thinks things over” deliberately, and delivers conscious calculations about how to handle various situations.

Some decisions are the result of your brain combining suggestions from System 1 and System 2. But you need to keep in mind that a large body of evidence suggests that the majority of decisions are probably either fully or mostly the result of System 1. That’s important because System 1 decision-making is prone to a variety of systematic errors.

Cogitating on cognitive biases



REMEMBER

Cognitive biases are the misperceptions or misunderstandings that produce systematic errors. Cognitive biases fall into two categories:

- » Those resulting from faulty heuristics.
- » Mental-processing errors that are the result of attempting to use brains that evolved in much simpler times to solve modern problems. Our brains evolved before modern complexities such as calculus, computer science, and trying to figure out how to get more “likes” on social media. So our brains simply haven’t developed the tools necessary to solve these problems very well.

Imagine a vicious underworld populated by thinking mistakes. You’re looking at a line-up of “the usual suspects.” Do you see some familiar bad guys here? (And can you give personal examples of each of these mistakes and biases?)

- » **Confirmation bias** is the common human tendency to pay attention only to information that agrees with, or confirms, one’s preconceptions. This causes information that contradicts one’s preconceptions to be ignored completely or rationalized away so that the incorrect opinion persists.
- » **Self-serving bias** refers to people’s tendency to attribute their successes to personal effort or personal character traits but their failures to factors outside of their control. This bias makes it difficult for people to learn from their mistakes.

- » **The overconfidence effect** refers to people's tendency to be overly confident about the correctness of their opinions and judgments and to act without taking time to verify their initial hunches.
- » **Hindsight bias** occurs when people retroactively believe that they were able to predict past events. A faulty "I-knew-it-all-along" perspective causes many people to overestimate their predictive abilities.
- » **The availability heuristic** causes people to base their estimates about the likelihood of an event not on objective facts but on whether mental images of similar events are quickly and readily brought to mind. Events that carry vivid and emotionally charged mental images are presumed to be much more common than events that bring to mind dull or boring mental images. Shark attacks are, in fact, rare, but the mental image of being attacked by a shark is extremely vivid for most people. That vividness, or *salience*, makes people overestimate the likelihood of shark attacks.
- » **The planning fallacy** refers to the tendency people have to substantially underestimate the time needed to complete a task. When a student waits until 2 a.m. to start cramming for an 8 a.m. exam and when Congress waits until right before its August recess before trying to pass a long-awaited reform, both are up against the planning fallacy.
- » **Framing effects** have to do with providing context, or framing, for a decision. Framing effects are in play when a change in context (frame) causes people to change their decision-making. Changes in context can cause extraordinary changes in behavior. When experimenters tagged an area with graffiti and scattered trash around, ordinary people were twice as likely to litter, steal, and trespass. Upscale retail marketers play the opposite game, spending on expensive packaging, displays, and even architecture to increase the perceived value of their merchandise — and, you guessed it, how much customers are willing to pay.

Surveying Prospect Theory

Neoclassical economics focuses much of its attention on consumer-choice situations in which people have to choose between different "goods," as when selecting items at a supermarket. But in the real world, people have to make choices that involve the possibility of "bads" as well — they have to deal with uncertainty and the possibility of bad outcomes as well as good ones.



Prospect theory makes predictions about how people will make decisions when presented with both potential gains and potential losses. It's based on three key factors that influence how people choose in these situations:

- » People judge good things and bad things in relative terms, as gains or losses relative to their current situation, or *status quo*.
- » People experience both diminishing marginal utility for gains (each successive unit of gain feels good, but not as good as the previous unit) as well as diminishing marginal disutility for losses (each successive unit of loss hurts, but less painfully than the previous unit).
- » People experience *loss aversion*, meaning that for losses and gains near the status quo, losses are felt much more intensely than gains — in fact, about 2.5 times more intensely.

In addition to shedding important light on how consumers weigh their prospects, prospect theory provides insights into how advertisers select their tactics, how manufacturers decide on pricing, why people tend to save too little for retirement, and why people often can't stick to weight loss programs.

To better see why psychologist Daniel Kahneman was awarded the Nobel Prize in Economics for developing prospect theory, this section reviews instances of consumer behavior that would be hard to explain without it.

Shrinking packages and loss aversion

Because people see the world in terms of gains and losses relative to the status quo situations they're used to, businesses have to be careful about increasing the prices they charge for their products since consumers will tend to treat any price increase as a painful loss. This has led firms to develop a “workaround” that can help them raise prices without angering consumers.

About 100 years ago, when the Hershey's chocolate brand was closely identified with its 5-cent candy bar, the company would decrease the size of the nickel bar when the cost of raw materials rose and increase the size when the cost of raw materials fell. You might think of this as a way to shield customers from changes in raw materials prices. But if you think a little harder, you'll realize that the input prices were still passed along as changes in how much (or little!) chocolate customers got for a nickel. The price *per ounce* still went up and down.

But Mr. Hershey understood that customers don't pay much attention to the price per ounce. They're fixated on the price per unit. He understood that any increase in the price per unit would be categorized mentally as a loss (higher prices!)

relative to the 5-cent per unit status quo. Thus, he kept the price fixed at a nickel to avoid triggering people's sense of loss aversion.

This trick is often referred to as *shrinkflation*, since shrinking package sizes become commonplace whenever the overall economy is experiencing a high rate of price inflation and, consequently, rising input costs.

A similar trick involves keeping the exterior package size of a given product the same but reducing the amount of usable product inside each package. During the early 2020s inflation that followed the Covid-19 pandemic, for example, some people on TikTok were complaining about fewer raisins in each box of Raisin Bran cereal.

People also complained about many brands of toilet paper wrapping less and less paper around each central tube. That by itself was annoying, but to try to disguise the shrinkage, the toilet paper makers increased the diameter of the central tubes, so that even though there was less usable paper, the overall diameter of each paper roll (i.e., the width of the hollow tube plus the width of the wrapped-around paper) stayed the same.

Framing effects and advertising

Because people evaluate situations in terms of gains and losses, their decision-making can be sensitive to the mental *frame*, or context, that they use to evaluate whether a possible outcome should be viewed as a gain or a loss. In terms of prospect theory, the framing information defines the status quo and thus our sense of whether any possible outcome will be perceived as a gain or a loss.

Here are two examples of *framing effects* and how changing the frame can flip your perception of whether a possible outcome is to be viewed as a gain or a loss:

- » You're offered a salary of \$100,000 per year. Nice! But what if you're used to earning \$140,000 per year? Not nice.
- » You have a part-time job. Your boss walks in and says he's going to give you a 10 percent raise. Sweet! But then your boss says that all the other part-timers will be getting a 15 percent raise. Sour!

Advertisers know how important context is and go out of their way to frame their offers as gains rather than losses. Think about ads for ground beef. If they say "80 percent lean," we think of the meat as offering gains to health; but if they say "20 percent fat," the same meat feels like a potential coronary.

WANNAMAKER'S WARNING

In the late 19th century, American entrepreneur John Wannamaker invented the department store, the price tag, and the money-back guarantee. But he was famously frustrated by marketing and advertising. "Half the money I spend on advertising is wasted — the trouble is, I don't know which half."

His frustration echoes today. Most advertising campaigns fail. Most movies flop. And 80 percent of new product launches are abandoned within three months. So while advertisers and marketers are constantly doing their best to get consumers to buy this or that — including trying to utilize behavioral economics to get people to do their bidding — we can go about our lives with a little smirk because most of their efforts seem to be pretty ineffective in altering our behavior.



REMEMBER

Any frame that alters gain-or-loss perceptions affects purchasing decisions. So sellers go out of their way to frame potential purchases as gains rather than losses.

Anchoring and credit card bills

Before people can calculate their gains and losses, they must first define the status quo from which to measure those changes. Unfortunately, people's sense of the status quo can be distracted by irrelevant information, including irrelevant numbers.

Credit card companies know this trick, known as *anchoring*, because the irrelevant number can be used to anchor, or set, people's sense of what the status quo should be. Credit card companies do this by printing small minimum payment amounts on monthly credit card statements.

The small minimum payment amounts can unconsciously set people's status quo in terms of how much they should pay each month. If that happens, any higher amount would feel like a loss relative to the minimum payment amount suggested by the credit card company. And that would make it harder to select a higher amount.

Getting you to select the small minimum payment amount is in the better interest of the credit card company because it wants you paying off your credit card debt as slowly as possible so that it can charge you as much interest as possible. It tries to anchor you at the small amount it prints as the minimum payment.



Here's a pro tip: Decide how much you want to pay before you look at your monthly credit card statement. That way you can't be influenced by any information that might anchor you into paying too little.

Examining the endowment effect

If you show a person a new mug and ask them to put a price on it, they might say \$10. But if you gift that person the mug and then ask them how much you'd have to pay them to buy it back, they'll typically quote you a substantially higher price, like \$15.

What's interesting is that the higher buy-back price isn't the result of the person bluffing or attempting to drive a hard bargain. The way human brains are wired, people tend to put a higher valuation on anything that they currently possess, or are endowed with, than on identical items that they don't own but might purchase. This phenomenon is called the *endowment effect*. It shows up all over the place, including with home purchases.

It appears that loss aversion is the likely culprit behind the endowment effect. Once you possess something, having it becomes your new status quo. And that means the prospect of parting with it would be perceived as a potential loss relative to the status quo of retaining possession. Consequently, people demand more money to part with something they currently possess than they would to purchase the same item if they didn't own it yet.

That difference in perceived value can be a problem for negotiations between buyers and sellers. The sellers will on average be demanding more than the buyers are going to be willing to pay.

Stipulating status quo bias

Prospect theory also explains *status quo bias*, which is the tendency people have to favor any option that's presented to them as being the default, or status quo, option.

The effect of status quo bias can be massive. Consider organ donation programs tied to driver's license applications and renewals. In countries where the default option is donating your organs if you die in a car accident, nearly 100 percent of people are registered as donors. By contrast, in countries where the default option is *not* donating if you die in a car accident, only about 10 percent of people are registered as donors.

Both cases would require finding and filling out a form to switch the applicant away from the default option. And in both cases, hardly anyone bothers to find and fill out the form. They just go with whatever option was presented as the default, or status quo.

Prospect theory explains such examples of status quo bias as a combination of the endowment effect and loss aversion. Put into a novel situation, people will treat the default option as an endowment and will regard any other option as a prospect with the potential for loss. Loss aversion then kicks in and reinforces the appeal of the default option. The result is a bias toward the status quo and sticking with whatever looks like the default option.

Researchers who wanted a simple method for increasing retirement savings put status quo bias to good use. They made automatic enrollment into a savings program the default situation for thousands of workers. Those workers ended up saving far more than workers whose default option was to not be enrolled.



TIP

Such research leads to a more general point: Because people for the most part tend to stick with whatever the status quo is, why not make the status quo something helpful or useful?

Countering Myopia and Time Inconsistency

Our distant forbears had to be almost entirely focused on the present moment — on how to avoid diseases and predatory animals and starvation, for example, and on how to get through the next few weeks or months.

That's why part of our genetic inheritance is our difficulty with long-run planning and decisions that involve trade-offs between the present and the future. Two of our major stumbling blocks are myopia and time inconsistency.

Focusing on myopia

In medicine, *myopia* means nearsightedness. People with myopia can see nearby objects clearly, but everything farther away looks blurry. By analogy, economists use *myopia* to refer to the fact that our brains have a hard time conceptualizing the future: It seems fuzzy, out of focus, and hard to see. That's why we have great difficulty imagining, for example, the additional spending power that will be ours in 30 years if we begin saving money now.

A primary consequence of decision-making myopia is that when consumers are forced to choose between one thing that will generate benefits quickly and another thing that won't yield benefits for a long time, they have a strong tendency to favor the more immediate option even if the long-run option is well worth waiting for.

Consider having donuts today versus having an athletic figure down the line. Those donuts are close and warm and tempting. You can smell and see them now. But the benefits of being in great shape in three or four years (washboard abs!) will seem relatively vague and far less pressing. Your brain will tell you to eat the donut — or two donuts — and thus lower the likelihood of your having a great physique in three years, even if that's something you genuinely want.

Tattling on time inconsistency

We view the future myopically from the here and now. This leads to a tendency to make bad predictions about what we want to do at particular points in the future.

Imagine that it's currently 10 p.m., and you decide to set an alarm for 6 a.m. the following morning so you can wake up and go to the gym. Your 10 p.m. self assumes that your 6 a.m. self will be on board with going to the gym once the alarm goes off. After all, it's the same you. But when the alarm goes off at 6 a.m., your 6 a.m. self doesn't want to go — and even turns the alarm off, goes back to sleep, and never gets to the gym.

The underlying problem is that your 10 p.m. self myopically fails to see how painful getting up and going to the gym is going to be. But when that moment arrives, your 6 a.m. self feels the pain keenly. Thus, your 6 a.m. self makes a decision upon waking that your 10 p.m. self didn't correctly anticipate the night before.

Economists refer to situations like this as examples of *time inconsistency*. It happens whenever our current selves don't understand what our future selves will want once the future becomes the present.

Beating self-control problems with precommitments

Time inconsistency is a major cause of self-control problems. The misperceptions and misjudgments entailed by time inconsistency subtly encourage people to put themselves in temptation's way. You can say today that you're going to spend three hours tomorrow working on your tax return. But when tomorrow comes around, you binge watch your favorite show instead.

GETTING GREEN THE MYOPIC WAY

Myopia causes consumers to focus too strongly on the up-front costs of installing solar panels. They can't see far enough into the future to appreciate the massive long-run savings they could achieve if they paid for solar panels now in exchange for cheap electricity in the future.

Several solar-installation companies have figured out a way to work with, rather than against, people's myopia. They do so by offering leasing and financing options that eliminate the need for consumers to pay for the up-front costs of installing a solar system. Instead, the installation company pays for the upfront costs and then makes its money by splitting the resulting cost savings with the consumer.

This arrangement actually benefits from myopia because consumers get to focus on their share of the savings, which is immediately available. It's a great example of working with, rather than against, a behavioral bias.

Fortunately, behavioral economists have developed some techniques to help us take actions today to control our rascally future selves.



TIP

One of the leading strategies for dealing with time inconsistency is to make *precommitments* that your future self can't easily reverse. This strategy may include tactics such as setting multiple alarm clocks so that you can't easily get back to sleep when you're supposed to go to the gym; signing up for automatic payroll deductions so you don't have to decide every two weeks whether to save for retirement or blow all the money on in-game purchases for your Minecraft videogame character; and joining a weight loss competition so that there will be ongoing social pressure to keep you committed to reaching your weight loss goals.

Gauging Fairness and Self-Interest

Neoclassical economics assumes that economic behavior is driven primarily — or even exclusively — by self-interest. Adam Smith, the founder of modern economics, put this idea into words in *The Wealth of Nations* in 1776: "It's not from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest. We address ourselves not to their humanity but to their self-love, and never talk to them of our own necessities but of their advantage."

Whoa, lighten up, Mr. Smith! That sounds like it could have been written by Ebenezer Scrooge or the Grinch.

But don't judge Smith too quickly. He was a philosopher before he was an economist, and in his other famous book, *The Theory of Moral Sentiments*, Smith wrote, "How selfish soever man may be supposed, there are evidently some principles in his nature which interest him in the fortune of others and render their happiness necessary to him though he derives nothing from it except the pleasure of seeing it."



REMEMBER

Was Adam Smith an old softie from the get-go, despite his more famously quoted statement on self-interest? It appears so. And it appears that he was right to feel that way. Behavioral economists have amassed a pile of evidence that indicates that there's an ever-present human tendency toward altruism and caring about one's fellow man. Although self-interest is also always present, most people do care deeply about others and about the fairness of their own behavior in interactions with other people. For that reason, economic transactions are heavily influenced by moral and ethical factors.

Defining fairness



REMEMBER

In economics, *fairness* is a person's sense of whether a price, wage, or allocation is considered morally or ethically acceptable. Any tendency toward pure self-interest is modified by people's sense of fairness or propriety. In some cases, people go so far as to act precisely the opposite of how you would expect a purely self-interested person to behave. Examples include giving anonymously to charity, obeying the law even when there's zero chance of getting caught, and being willing to pay a premium price to purchase a "fair-trade" product.

Examining the experimental evidence for fairness

In recent decades, experimental games have been created to test people's feelings about fairness. One key feature is that the games are played for real money. Participants have a strong incentive to act in a self-interested manner so they can win as much money as possible.

Another key feature is that players interact anonymously via computer screens. Economists provide that anonymity to give each player the chance to be a jerk without having to fear retaliation either during the game or afterward. That freedom from retaliation liberates players to use whatever strategies are most likely to maximize their own winnings during the game. The question is whether players will use that freedom to act in a purely self-interested manner — or behave more nicely.

The dictator game and the ultimatum game are two of the most famous. Each game involves two players.

Dishing on the dictator game

The experimental researcher who runs this game puts up a fixed amount of money, say \$10, and designates one of the two players as the “dictator.” The dictator’s job is to determine the split — be it \$10/\$0 or \$8.67/\$1.33 or another division — between the dictator and the other player. The dictator chooses the division. The researcher executes it, paying each player their share. And the game ends.

The dictator game has been played hundreds of thousands of times all over the world. A review of the outcomes shows that roughly one-third of dictators keep all the money for themselves. The other two-thirds of dictators show generosity by allowing the other player some money — on average, \$4.20. Seventeen percent of all dictators split the money evenly, and 5 percent of all dictators give the entire amount to the other player.



REMEMBER

From such experimental results, behavioral economists have come to believe that some people are incredibly selfish, others are incredibly generous, and most of us are somewhere in between. But the most important thing to remember is that the most common behavior is to give a perfectly anonymous stranger a substantial share of the money even though there are zero consequences for being entirely selfish.

Unraveling the ultimatum game

This game is a bit more complex, represents a more realistic money-splitting situation, and is often played for higher stakes than \$10. In this real-money, two-person game, there’s no dictator who can unilaterally decide what the split will be. Instead, both players must jointly agree on any proposed division of the money for the split to take place.

In real-world markets where buyers and sellers interact voluntarily, a transaction takes place only when a proposal made by one party is accepted by the other. In the ultimatum game, one player is randomly assigned to function as *proposer* and the other as *responder* — and both face the possibility that if they can’t agree, both will forgo all possible benefits.

The proposer issues an offer (the ultimatum) suggesting a split — perhaps 70 percent for themselves and 30 percent for the other player — and the responder gets to either accept it or reject it. If the responder accepts the proposed split, the split is made, the game ends, and both players go anonymously on their way. But if the responder rejects the proposer’s suggested split, neither player gets anything, and the game ends without anybody getting any money.

Here's what typically happens: Almost no proposers suggest allocating the money perfectly unfairly, 100/0 percent in their own favor. The majority of proposers suggest either a perfectly equal 50/50 split or something close to it, like 55/45 percent.



REMEMBER

The key thing to notice is that proposers in the ultimatum game are even more considerate in their proposed splits than are dictators in the dictator game. That happens because the responders in the ultimatum game can be active and reject, whereas the second players in the dictator game must be passive and just take whatever the dictator proposes. The ability of responders in the ultimatum game to reject offers means proposers must be careful not to offend responders, especially responders' sense of fair treatment. That leads proposers toward suggesting equal splits.

Digesting the experimental evidence on fairness

A good way to interpret the behavior we observe in the dictator and ultimatum games is to notice that the two games support both the gentler instincts discussed in *The Theory of Moral Sentiments* and the power of self-interest that features so prominently in *The Wealth of Nations*:



REMEMBER

- » The fact that dictators who face no possibility of retaliation on average give quite a bit of the money to the other player in the dictator game is consistent with human beings having an intrinsic sense of fairness that causes them to help others even when there's nothing in it for themselves.
- » The fact that splits become more equal when there's a threat of rejection in the ultimatum game shows that incentives for good behavior can make people behave even more fairly than they do when left completely to their own devices without any threat of punishment or retaliation.

The increased equality of splits that we see when both parties must consent to them gives us another way to understand Adam Smith's metaphor of the *invisible hand*. When people are constrained by having to consider the needs and interests of others, they tend to behave in more mutually beneficial, socially optimal ways — as if guided to do the right thing by an invisible hand.

That's a powerful result from two simple games. And it's yet another illustration of the dazzling insights that behavioral economics has delivered since it burst onto the scene back in the 1980s.



WARNING

NUDGING PEOPLE TOWARD BETTER OUTCOMES

Behavioral economists have begun “nudging” people toward better decisions by using gentle reminders or mild peer pressure. Many nudges can be implemented at extremely low cost. Here are some examples:

- Weekly text message reminders increased attendance at adult literacy classes in England by 33 percent.
- Letters that told taxpayers in Guatemala that most of their neighbors had already paid their taxes tripled collections.
- Personal savings was increased by a factor of seven in the Philippines by offering people commitment savings accounts, which restricted people from withdrawing money either until a specific date or until the account reached a specific dollar amount.
- Households slashed their electricity consumption substantially after an electric company started printing sad-face icons like ☹ on the bills of high-consumption households.

But before applauding these efforts too loudly, consider for yourself whether nudges are ethical. Would you be okay with being subconsciously manipulated if the manipulation helped you get something you desired?

4

Macroeconomics: The Science of Economic Growth and Stability

IN THIS PART . . .

Discover gross domestic product — the total value of goods and services produced in a country.

Probe inflation, price indexes, and interest rates to see how inflation is caused by printing too much money.

Explore the business cycle and why recessions happen.

Find out what government should or shouldn't do to fight recessions using tools like stimulus spending and fiscal and monetary policy.

Understand that borrowing and debt drive price bubbles and financial crises and why fiscal and monetary policy don't work very well after a financial crisis.

- » Measuring gross domestic product:
The total value of goods and services
- » Decomposing GDP into $C + I + G + NX$
- » Understanding why free trade is good
for you

Chapter 14

How Economists Measure the Macroeconomy

Macroeconomics studies the economy as a whole. Seen from on high, either businesses or the government produces goods and services. Businesses produce the bulk of what people consume, but the government provides many goods and services, including public safety, national defense, and public goods such as roads and bridges. In addition, the government provides the legal structure in which businesses operate and intervenes in the economy to regulate pollution, mandate safety equipment, and redistribute income from the rich to the poor. (For more on the division of tasks between private businesses and the government, see Chapter 3.)

To study production, distribution, and consumption with any real understanding, economists must keep track of exactly how much is produced and where it all ends up. Consequently, economists have developed a huge accounting apparatus, called National Income and Product Accounts (NIPA), to measure economic activity. This system produces many useful statistics, including the famous gross domestic product (GDP), which measures the total quantity of goods and services produced in a country in a given period of time.

The system can seem arcane, but knowing the accounting is indispensable. In fact, it's the basis for all the mathematical models that economists use to understand and predict the business cycle, inflation, economic growth, and monetary and fiscal policy. (I present some of these models in Chapters 16 and 17.) So make sure to take, uh, "proper account" of what I'm about to show you.

Getting a Grip on the GDP and Its Parts



REMEMBER

Gross domestic product (GDP) is a statistic that calculates the value of all goods and services produced in a given country in a given period of time. In the United States, the Bureau of Economic Analysis at the Department of Commerce computes this statistic every three months, giving an idea of how much economic activity took place in the previous quarter. Add up the numbers for four consecutive quarters, and you get GDP for a year.

GDP is important because higher levels of output afford greater access not only to essentials like good food and quality healthcare, but to a variety of goods and services that bring people pleasure and enjoyment. I'm not saying that money is the only thing that matters, but economists evaluate economies by how successfully they maximize happiness. Although money can't buy you love, it can sure as heck buy you a lot of other things that make you happy, such as food, education, and vacations. So, a high and quickly growing GDP is preferable because it reflects lots of economic transactions that provide people with the goods and services they desire. (To examine some reasons GDP may *not* always reflect increased happiness, see the later section "Watching GDP rise with the good, the bad, and the ugly.")

Chapters 2 and 3 discuss how people's fundamental economic goal is to maximize happiness given limited resources. Because people like to consume goods and services, measuring GDP allows economists to quantify, in some sense, how well a country is doing at maximizing its citizens' happiness given the country's limited resources. Thus, a rising GDP indicates that a country is figuring out ways to increase overall happiness, all other things held equal.

In this section, I show how and why the economists who tabulate the NIPA divide GDP into its constituent parts. Breaking up GDP lets you analyze each part separately and get a good handle on the major factors that influence the production of goods and services. But first, I give you a short explanation of what GDP *doesn't* take account of.

Leaving some things out of GDP



REMEMBER

The GDP statistic counts only transactions that involve money, so if you do volunteer work for your parents or if a mother stays home to take care of an infant, that economic activity — though very productive and socially beneficial — isn't counted in GDP.

In economies like that of the United States, GDP is good at capturing nearly all output that's produced because almost everything that's produced within the U.S. is subsequently sold. But in a largely rural and agrarian society of small farmers, most production is for consumption within the household. Consequently, most of their output never makes it into the official GDP statistics.



WARNING

As countries transition from rural agrarian economies with lots of household production to market economies where nearly everything produced is sold for money, GDP appears to rise because a lot of output is being counted for the first time. But this apparent change may not be an increase in output. These limitations can make comparing the GDPs of various countries misleading.

Tallying up what counts in GDP

Counting sales where money trades hands can get a little tricky because both a buyer and a seller are involved in every such transaction. The money that the buyer spends has to equal the money that the seller receives. Translated into economist lingo, income has to equal expenditure. Consequently, you can measure GDP by totaling all the expenditures in the economy or by summing together all the incomes in the economy. If your calculations are correct, both methods give you the same value for GDP.



REMEMBER

When thinking about GDP, you also have to consider the goods and services that are being traded for money. Economists simplify life by saying that households directly or indirectly own all the resources or factors of production of a society — land, labor, capital, and entrepreneurial ability. (See Chapter 3.) *Households* can be made up of one person or several — individuals or families. *Firms* buy or rent the factors of production from the households and use them to produce goods and services, which are then sold back to the households. This process sets up a *circular flow* of resources moving from households to firms and of goods and services moving back the other way, as Figure 14-1 shows.

Moving opposite to the flow of resources and goods are payments in dollars. When firms buy factors of production from households, they have to pay money to the households. That money is *income* to the households. And when households buy goods and services from the firms, they pay for those goods and services with money, which in Figure 14-1 are *expenditures*.



REMEMBER

Households own firms; firms don't exist on their own. As a result, any money that a firm receives when it sells a good or a service flows on as income to some individual or group of individuals. Because of this, incomes in Figure 14-1 have to equal expenditures. Also, any resources that firms own — land, factories, mines, and so on — households also own because households own all the firms in the economy. Thus, all income generated by selling resources to produce goods and services flows directly or indirectly to households.

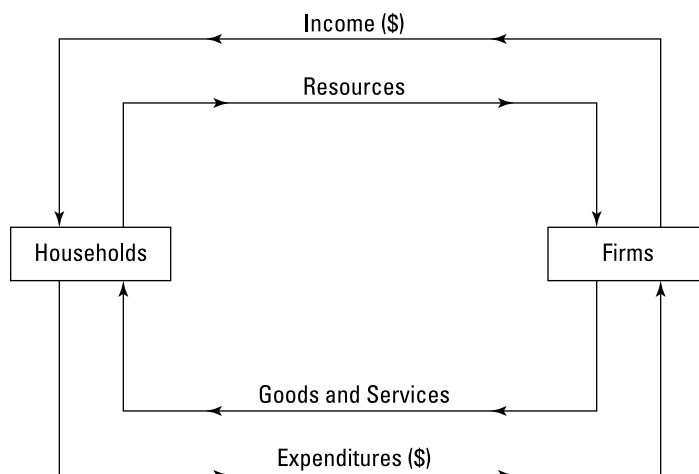


FIGURE 14-1:
The simple
circular flow
diagram.

© John Wiley & Sons, Inc.

Accounting for streams of incomes, and assets



REMEMBER

Although you can use either incomes or expenditures to measure GDP, economists prefer to use incomes because governments make individuals and businesses keep track of every penny of income they receive so that it can be taxed. This requirement provides extensive, accurate data about incomes.

Tracing the flow of income

All the income in the economy flows into one of four categories:

- » Labor receives wages.
- » Land receives rent.
- » Capital receives interest.
- » Entrepreneurial ability receives profits.

Because you need land, labor, and capital to make things, you have to pay for them. That's why some of the income in the economy flows their way. But in a dynamic, competitive economy, you also need individuals who are willing to bear business risks like going bankrupt and who are willing to invest money in risky new technologies that may never catch on. To get them to do so, you have to pay them, so some income must also flow as profits to risk-taking entrepreneurs to compensate them for providing entrepreneurial ability.

Each of the four payments is a flow of money that compensates for a flow of services needed in production:

- » **Wages:** Workers charge wages for the labor services they provide.
- » **Rent:** Owners of buildings and land charge rent to tenants for the services that real estate and physical structures provide.
- » **Interest:** Firms wanting to obtain the services of capital, such as machines and computers, must pay for them. This payment is considered interest because, for example, the cost of obtaining the services of a \$1,000 piece of capital equipment is the interest payments that a firm must make on a \$1,000 loan to buy that piece of equipment.
- » **Profits:** The firm's profits must flow to the entrepreneurs and owners of the firm, who bear the financial risk that the firm may do poorly or even lose money and go bankrupt.

Taking assets into consideration

What happens to the flow of income if a firm buys land and office space rather than rents it? Or if a firm owns its capital outright instead of borrowing money to buy it? If a firm owns these things, it no longer has to pay a flow of money to obtain a flow of services. Do expenditures still equal incomes? Yes. Incomes still equal expenditures. But you have to do some fancy accounting to see why. The key to this balancing act is understanding what an asset is.



REMEMBER

An *asset* is something durable that isn't directly consumed but that gives off a flow of services that you do consume. A house is an asset because it provides shelter services. You don't consume the house (just think of all the fiber!); you consume the services it provides. Similarly, a car is an asset because although you don't consume the car itself, it provides transportation services.

You often have a choice between buying an asset outright and thereby owning all the future services that the asset provides or letting someone else own the asset and sell you the services as they're produced. For example, you can buy a house and thereby get all future shelter services that the house provides, or you can rent the house and get those same services by paying for them each month. For this reason, an asset is considered to be a *stock*, and the services it provides are referred to as a *flow*.

Following the funds around and around

The simple circular flow diagram of Figure 14-1 captures the fact that an income exists for every expenditure. However, because the diagram divides the economy only into firms and households, it misses a lot of the action that goes on in the real world. In Figure 14-2, you can see a much more realistic and detailed circular flow diagram that divides the economy into firms, households, and the government, with these entities making transactions in the following three markets:

- » **Markets for factors of production:** Money is exchanged to purchase or rent the land, labor, capital, and entrepreneurship used in production.
- » **Financial markets:** People who want to lend money (savers) interact with those who want to borrow money (borrowers). In these markets, the supply and demand for loans determine the *interest rate*, which is the price you have to pay to get someone to lend you their money for a while. Because most governments run deficits (in other words, they're always in the hole) and have to borrow a lot of money, they're major players in the financial markets.
- » **Markets for goods and services:** People and the government buy the stuff that firms make.

In Figure 14-2, arrows show the flows of dollars throughout the economy. Firms make factor payments — rent, wages, interest, and profits — to households to obtain the factors of production — land, labor, capital, and entrepreneurial ability. (See the previous section “Accounting for streams of incomes, and assets” for details.) Households take the income they get from selling these factors and use it to pay for goods and services, to pay taxes, or to save. The government buys goods and services using either the tax revenues it takes in or the money it borrows in the financial markets. The financial markets also provide dollars for corporations to make investments. These dollars add to those that firms get from selling goods and services to households and the government.

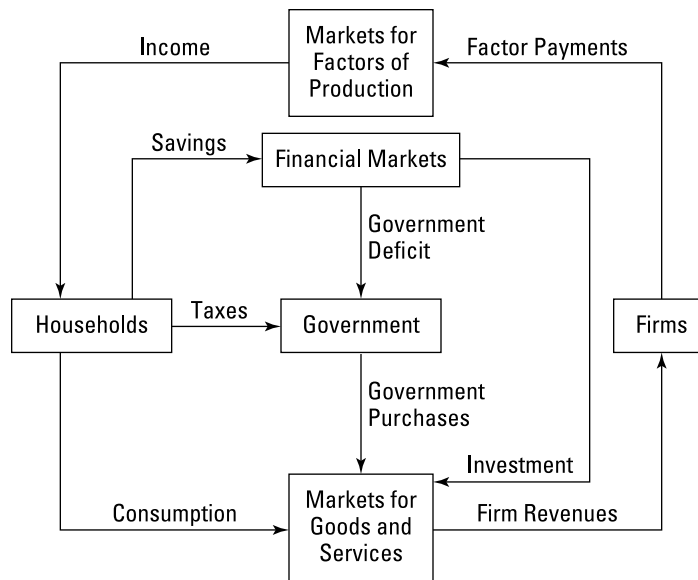


FIGURE 14-2:
The detailed
circular flow
diagram.

© John Wiley & Sons, Inc.

Counting stuff when it's made, not when it's sold

Newly produced output is counted as part of GDP as soon as it's produced, even before it gets sold. That makes keeping track of the money associated with new production a little tricky.

For example, as soon as construction on a new house is completed, its market value of \$400,000 is estimated and counted as part of GDP right then, even though the house may not be sold for months. Suppose construction is completed on December 29, 2026, adding \$400,000 to the year 2026's GDP. If the house is sold on February 21, 2027, it doesn't count in the year 2027's GDP because double counting isn't allowed.

When the house is sold, it's considered old property and not new production. Economists just say that the property right to this now-old house has changed hands from the builder to the new owner. Because trading old assets obviously involves no new production, it doesn't count in GDP. Similarly, items such as used cars and used books don't count in GDP, no matter how many times they're bought and sold in resale markets.

This accounting convention applies to firms producing any sort of output good whatsoever. If Samsung produces a TV on December 31, 2026, the value of that TV is counted in the year 2026's GDP even though it won't be sold to a customer until the next year.



TIP

A handy way to think about this is to imagine that Samsung builds the TV and then sells it to itself when putting the TV into inventory. This “sale” is what’s counted in GDP for the year 2026. When the TV is later sold out of inventory to a customer, the sale is just an exchange of assets (trading the TV for cash).



WARNING

The fact that output is counted when it’s produced rather than when it’s sold is a red flag when interpreting GDP statistics to gauge the health of the economy. High GDP means only that a lot of stuff is being produced and put into inventory. It doesn’t necessarily mean that firms are selling lots of stuff.

Economists who try to forecast where the economy is heading pay much more attention to inventory levels than they do to last quarter’s GDP. Why? Because it’s quite possible that although GDP is high, the economy is about to go into a recession because inventories are piling up and managers will soon cut back on production to get inventories back down to target levels.

Watching GDP rise with the good, the bad, and the ugly

Generally speaking, higher GDP is better than lower GDP because more output produced means higher potential living standards, including better healthcare for the sick and more money to aid the needy. But higher GDP doesn’t guarantee that happiness is increasing because GDP often goes up when bad things happen. For instance, if a hurricane destroys a big section of a city, GDP goes up as reconstruction kicks into gear and lots of new output is produced to replace what was destroyed. But wouldn’t it have been better not to have had the hurricane in the first place?

Similarly, higher GDP may be possible in certain situations only if you’re willing to tolerate more pollution or greater income inequality. Countries experiencing the early stages of industrialization and quickly rising living standards often also get dirtier environments and more social unrest because some people are getting richer much faster than others. The GDP number doesn’t reflect these negative outcomes.

GDP also doesn’t count the value of leisure. Many of my favorite memories are of times when I was neither producing nor consuming anything that would count in GDP. I was sitting on the beach, climbing a mountain, taking a walk, or working out with friends. Moreover, an increase in GDP often comes at the price of sacrificing these leisure activities — meaning that when you see an increase in GDP, overall well-being or happiness hasn’t necessarily improved. So although policies that raise GDP are generally beneficial for society, you must always examine the costs involved in creating the rising output.

Diving In to the GDP Equation

Now's the time for you and GDP to make friends so you can understand all of GDP's little secrets — in particular, its constituent parts and how they behave. The discussion in this section is interesting by itself, but it's doubly useful because it makes the standard Keynesian macroeconomic model (which I introduce in Chapter 16) much easier to understand and manipulate.

In the section “Tallying up what counts in GDP” earlier in this chapter, I explain that you can measure GDP either by adding up all the expenditures made on purchasing goods and services or by adding up all the incomes that are derived from producing goods and services. The two numbers have to be equal, so this switch to the expenditure method of counting up GDP is kosher. (It's also the perfect opportunity for you to understand the economy from the point of view of where money is spent, as opposed to who gets to keep what's earned.)



REMEMBER

The expenditure equation for totaling up GDP adds together the four traditional expenditure categories — consumption (C), investment (I), government spending (G), and net exports (NX). Their sum is abbreviated with the symbol Y and is equal to the value in dollars (or whatever currency a given country is using) of all the goods and services produced domestically in that country during a given period of time (typically a quarter or a year). In terms of algebra, the GDP equation looks like this:

$$Y = C + I + G + NX$$

Here's a quick look at the four expenditure variables that total up to GDP:

- » **Consumption:** C stands for consumption expenditures made by households on goods and services, whether produced domestically or abroad.
- » **Investment:** I stands for investment expenditures that firms make on new capital goods, including buildings, factories, and equipment. I also contains changes in inventories because any goods produced but not sold during a period have to go into firms' inventories and are counted as inventory investments.
- » **Government expenditures:** G stands for government expenditures on goods and services. (They've got to buy paperclips and staples for all those government documents.)
- » **Net exports:** NX stands for net exports, which are a country's exports (EX) minus its imports (IM), or $NX = EX - IM$. EX is the number of dollars of your nation's output that foreigners are buying. IM is the number of dollars of their output that your country is buying.

These four expenditures make up the GDP because, as a group, they buy up every last bit of output produced in your country in a given period of time. The following subsections offer more detail about each one.

“C” is for consumption (that’s good enough for me!)

Household consumption spending accounts for about 70 percent of GDP — far more than the other three GDP components combined. Many factors affect how much income households decide to spend on consumption and how much of it they decide to save for the future.

Microeconomists spend a lot of time studying the various factors that affect such decisions, including expectations about whether the future looks bright or dark and how high or low the rates of return are on savings. (See Part 3 of this book for info on microeconomics.) Macroeconomists, on the other hand, step back from these factors because, when studying the economy as a whole, what matters is the total amount of consumption across the entire economy rather than the consumption decisions of particular households or individuals.



TIP

Macroeconomists model consumption as a function of people’s after-tax, or *disposable*, incomes. You can derive disposable income algebraically using this handy three-step process:

1. Start with Y , the total income in the economy.

In Keynes’s equation, Y equals total expenditures, but because income equals expenditures, you can use it for income as well. Remember that any money you spend is income to someone else.

2. Figure out how much taxes people have to pay.

For simplicity’s sake, assume that the only tax is an income tax and that the income tax rate is given by t . For instance, $t = 0.25$ would mean a tax rate of 25 percent of people’s incomes. Consequently, the total taxes people pay, T , will be given by $T = tY$ — the tax rate times the total income in the economy.

3. Subtract people’s taxes, T , from their incomes, Y , to figure out their after-tax incomes.

Economists refer to this value as *disposable income* and write it algebraically as Y_D . Subtracting taxes from income looks like this:

$$Y_D = Y - T = Y - tY = Y(1 - t)$$

After you derive disposable income, you use a simple model of the consumption expenditures that households make. The model says that consumption, C , is a function of disposable income and a couple other variables, C_o and c .

$$C = C_o + cY_D$$

Lowercase c is called the marginal propensity to consume (MPC). It's always a number between 0 and 1 that indicates the rate at which people choose to consume income rather than save it. For instance, if $c = 0.9$, they consume 90 cents of every dollar of disposable income that they have after paying taxes. (They save the other ten cents.) The value of the marginal propensity to consume, c , is determined by the individual and varies from person to person, depending on how much of their disposable incomes they like to save.

You can think of C_o as how much people consume even if they have zero disposable income this year. (If you assume that $Y_D = 0$ in the equation $C = C_o + cY_D$, then that equation reduces to $C = C_o$.) But where does the money come from to pay for C_o if a person has zero disposable income? It comes from their personal savings, which they've piled up over the years, or from borrowing from other people who've saved over the years.



TIP

What the overall equation $C = C_o + cY_D$ says is that the total consumption expenditure in an economy is the amount people consume even if they have no income (C_o), plus a part of their disposable income (cY_D).

For the rest of this book, I assume that the equation $C = C_o + cY_D$ is a good-enough model of how consumption expenditures are determined in the economy. It's not perfectly realistic, but it does show that higher tax rates reduce consumption and that people make decisions about how much of their disposable incomes to save or consume. The equation allows us to analyze the effects of policies that change tax rates as well as the effects of other policies that encourage people to spend higher or lower fractions of their incomes.

"I" is for investment in capital stock



REMEMBER

Investment is vital because the economy's capacity to produce depends on how much capital is available to make output.

As discussed in Chapter 3, a nation's stock of capital includes both physical capital (such as buildings, infrastructure, computers, and factories) and intellectual capital (the accumulated stock of knowledge about how to make things and do things.)

The stock of physical capital increases when firms purchase new tools, buildings, machines, computers, and so on to help produce consumption goods. The stock of

intellectual capital increases when individuals, firms, universities, and the government spend money on research and development. Either way, investment spending is a flow that increases the capital stock of the economy.

Of course, physical capital wears out as it's used. Some of it rusts. Some of it breaks down. Some of it is thrown away when it becomes obsolete. Economists refer to these flows that decrease the physical capital stock as *depreciation*.

Naturally, firms must make some investments just to replace the physical capital that has depreciated. But any physical capital investment in excess of depreciation causes the overall size of the physical capital stock to increase, creating more potential output for people to consume.

By contrast, because intellectual capital does *not* depreciate, any money spent on R&D goes entirely toward creating more intellectual capital. (None of that spending has to go toward depreciation because old ideas or techniques are forever and don't depreciate in the way that an old engine will rust or break down.)



REMEMBER

The flow of investment spending over any period of time depends on the comparisons that firms make between the potential benefits and costs of buying pieces of capital or investing in R&D. The potential benefits are measured in terms of potential profits, and the costs of buying are measured by the interest rate, regardless of whether a firm takes out a loan to buy a given piece of physical capital or to fund the R&D spending necessary to create additional intellectual capital.

To build your intuition about why the interest rate matters so much for firms when they think about spending money on physical or intellectual capital, let's focus on just physical capital since that spending is more immediate and more tangible. (We also for simplicity ignore any investment activities undertaken by universities, individuals, non-profits, or the government.)

If a firm needs to take out a loan to buy physical capital, higher interest rates make the firm less likely to borrow money because the loan repayment costs will be high. However, even if a firm has enough cash on hand to buy a given piece of equipment, higher interest rates force the firm to decide between using the cash to buy the equipment and lending it to someone else. The higher the interest rates, the more attractive loaning out the equipment becomes. Consequently, higher interest rates discourage investment regardless of whether firms have to borrow to fund investment.

Economists model the amount of investment expenditure that firms desire to make, I , as a function of the interest rate, r , which is given as a percentage. The equation that I use here is standard in introductory books on macroeconomics (although notation does vary from book to book):

$$I_r = I_o - I_r r$$

This equation is similar in spirit to the consumption equation in the preceding section except for the minus sign, which indicates that when the interest rate rises, I falls.

The parameter I_r tells you how much I falls in an entire economy for any given increase in interest rates. For instance, suppose that r rises by 1 percentage point. If I_r is, say, 10 billion, you know that each percentage point increase in interest rates will decrease investment by \$10 billion.

The parameter I_o tells you how much investment would occur if interest rates were zero. In truth, interest rates almost never fall all the way to zero, but suppose that they did. Then the second term in the equation would be equal to zero, leaving you with $I = I_o$.



TIP

The equation as a whole says that if interest rates were zero, investment expenditures would max out at I_o . But as interest rates rise above zero and keep on rising, investment falls more and more. In fact, rates could potentially rise so high that investment spending would fall to zero.

The big “G” (government, that is)

In most countries, the government consumes a huge portion of GDP. In the United States, government at the local, state, and federal levels consumes about 35 to 40 percent of the GDP. In many other countries, the proportion is even higher. In most of Europe, for instance, it’s closer to 50 percent.



REMEMBER

The government gets the money to buy all that output from taxation and borrowing. If a government’s tax revenues are exactly equal to its expenditures, it has a *balanced budget*. If tax revenues are greater than expenditures, it’s running a *budget surplus*. But if the government’s expenditures exceed its tax revenues (which the government can arrange for by borrowing the difference on the financial markets), the government will be running a *budget deficit*.

Governments borrow by selling bonds. A typical bond says that in exchange for \$10,000 right now, the government will give you back \$10,000 in ten years and, in the meantime, pay you \$1,000 per year for each of the intervening years. If you accept the deal and buy the bond, you’re in effect lending the government \$10,000 right now and getting a 10 percent per-year return until the government gives you back your \$10,000 in ten years.

A huge amount of political maneuvering goes into determining how much a government is going to spend in a given year. Many groups lobby for special programs to benefit their hometown or their industry, and no matter what, governments have to provide for essential governmental functions such as national defense and law enforcement.

However, economists largely ignore the political machinations that go into determining government expenditures because the economic effects of government expenditure, G , depend on how big the expenditure turns out to be — not on how it got to be that size. So, for the rest of this book, I make the simplifying assumption that government expenditures can be denoted as

$$G = G_o$$

That extremely simple equation says that G is equal to some number G_o that the political process determines. G_o may be high or low depending on politics, but in the end, you care only about how big or small it turns out to be and can ignore where it came from.



TIP

G includes only government expenditures on newly produced goods and services. It doesn't include government expenditures that merely transfer money from one person to another. For instance, when the government taxes me and gives the money to a person in need, that transaction has nothing to do with currently produced goods and services and consequently doesn't count as part of G . So remember that when I talk about G , I'm talking about only the government's purchases of currently produced goods and services.

Measuring foreign trade with “NX”

When your country sells domestically made goods and services to someone or some firm in another country, such sales are called *exports*, or EX . When someone in your country buys something produced abroad, such purchases are called *imports*, or IM . *Net exports*, or NX , is simply the total value of all exports minus the total value of all imports during a given period of time (mathematically, $NX = EX - IM$). When using the expenditure method for totaling up GDP, you add in net exports, NX . But why add in only *net* exports?

The whole point of totaling up expenditures to get GDP is to figure out how many total dollars were spent on products made in your own country. Most of that expenditure is made by locals, but foreigners can also buy your products. That's exactly what happens when they pay you for the goods that you export to them. Consequently, you have to add in EX if you want to get a correct measure of expenditures made on stuff you produce domestically.

You have to subtract imports of foreign goods to differentiate the total expenditures that domestic residents make on *all* goods and services from their expenditures on *domestically made* goods and services. Total expenditures on all goods and services, both domestic and foreign, are C (as covered earlier). To get just the part that's spent on domestically made stuff, you subtract the value of imports, IM , because all money spent on imports is money that's *not* spent on domestically made goods and services. So $C - IM$ gives the amount of money that domestic residents spend on domestically produced output.

The result is that you can write your GDP expenditures equation that totals together all expenditures made on domestically produced output as follows:

$$Y = C - IM + I + G + EX$$

But the equation normally puts the exports and imports next to each other:

$$Y = C + I + G + EX - IM$$



TIP

Rearranging the GDP expenditures equation to include $EX - IM$ quickly reveals your country's *trade balance*. When $EX - IM$ is positive, you're exporting more than you're importing; when it's negative, you're importing more than you're exporting. Economists like to present the math in a way that tells a little story. International trade is hugely important, and you should understand not only why trade balances can be positive or negative but also why you shouldn't necessarily worry if it's negative (covered next).

Making Sense of International Trade and Its Effect on the Economy

Modern countries do a huge amount of trading with other countries — so huge, in fact, that for many countries, imports and exports equal more than 50 percent of their GDPs. So now's as good a time as any to focus a little more deeply on the NX part of the GDP expenditure equation, $Y = C + I + G + NX$.

This section explains why trade deficits (negative values of NX) aren't necessarily bad and why engaging in international trade — even when it means sustaining trade deficits — is typically hugely beneficial.

“Trade deficit” ain't fightin' words

If your exports exceed your imports, you have a *trade surplus*, and if your imports exceed your exports, you have a *trade deficit*. Unfortunately, the words *surplus* and *deficit* carry strong connotations that make surpluses sound like they're necessarily better than deficits. That's just not true, but you wouldn't know it from the rhetoric that politicians throw around. They make trade deficits sound as though they only lead to calamity.



WARNING

To focus on whether a trade deficit or surplus exists is to miss the point that international trade is a rearrangement of assets that makes everyone happier — even the country running the trade deficit. As long as international trade is voluntary, all trades enhance happiness, even if politicians say otherwise.

Consider two individuals who want to trade. Each person starts with \$100 cash, and each produces a product for sale. The first guy grows and sells apples for \$1 each. The second guy grows and sells oranges, also for \$1 each. Each of them produces 50 pieces of fruit.

Next, suppose that the guy who grows apples really likes oranges and wants to buy 30 of them for \$30, and suppose that the guy who grows oranges wants to buy 20 apples for \$20. Each guy is happy to satisfy the other guy's desires, so the apple grower spends \$30 buying oranges from the orange grower, and the orange grower spends \$20 buying apples from the apple grower.

Their trades shouldn't cause alarm bells to ring, but when people start looking at their trades using the terms *trade surplus* and *trade deficit*, they often come to the false conclusion that only one of the guys benefits from the trades that, in reality, they both were quite eager to make. To see where the confusion arises, notice that in the vocabulary of international trade, the apple guy exports only \$20 worth of apples but imports \$30 worth of oranges. The orange guy exports \$30 worth of oranges but imports only \$20 worth of apples. So, you have a situation in which the apple guy is running a \$10 trade deficit and the orange guy is running a \$10 trade surplus.

Does this mean that the apple guy is worse off than the orange guy? No. Each person started with \$150 worth of stuff: their respective \$100 cash piles plus \$50 each worth of fruit. When they finish trading, they each still have \$150 worth of stuff. The apple guy has \$90 of cash plus \$30 worth of apples and \$30 worth of oranges. The orange guy has \$110 of cash plus \$20 worth of oranges and \$20 worth of apples.

Saying that their trading has made either one poorer is way off the mark. In fact, both are happier with their arrangements of wealth after trading than they were before because their trades were voluntary. If the apple guy would've been happier keeping his initial \$100 cash and 50 apples, he wouldn't have traded for oranges — and the same goes for the orange guy.

Considering assets — not just cash

So why do some people get upset about all of those voluntary trades? Because instead of seeing the overall picture, they obsess over the fact that countries that run trade deficits end up with smaller cash piles. To people who hate trade deficits, changes in cash holdings seem alarming. But when gauging whether trade is beneficial, you need to consider the value of the traded assets, not just how much cash each party ends up with.

For instance, the fact that the apple guy's cash pile (see the preceding section) falls from \$100 before the trade to only \$90 after the trade looks spooky to deficit-haters because they're focused on the fact that the apple guy is \$10 poorer in terms of cash after trading with the orange guy. And they're even more peeved because that \$10 ends up with the orange guy, giving him a commanding \$110-to-\$90 advantage in terms of cash piles.

This perspective misses the fact that the apple guy's overall wealth is still \$150 and he now has a distribution of assets that's more pleasing to him than what he had before. Deficit-haters respond to this by asking you what happens after the apple guy eats his 30 apples and 30 oranges and after the orange guy eats his 20 apples and 20 oranges. In the end, all that the two guys have left are their cash piles. Because the apple guy has \$20 less cash than the orange guy, he must be worse off by running a trade deficit.

Again, this reasoning misses the point that the apple guy was happier trading and ending up with \$90 of cash than he would've been not trading and having \$100 in cash. If not for trade, he would've had a boring diet of only apples.

Opponents of trade deficits really make things seem scary when they start talking about land trading hands due to international trade. ("Oh no!" they say. "Foreigners are taking over our country!") To see their point, imagine that instead of starting with \$100 each of cash, the fruit farmers each start with 100 acres of land worth \$1 per acre. The only way for the apple guy to come up with \$10 of cash to pay for his trade deficit is by selling 10 acres of land to the orange guy. That is, the overall exchange that they engage in is 20 apples plus 10 acres of land worth a combined \$30 in exchange for 30 oranges worth \$30. Because 10 of the apple guy's acres of land now belong to the orange guy, deficit haters think the apple guy sold out his country — literally.

Such transfers of property do happen in real life. During the 1980s, the United States ran huge trade deficits with Japan. The result was that Japanese corporations and individuals ended up owning many famous U.S. buildings and companies. This really spooked many U.S. politicians, but they missed the point that all trading in life — be it with foreigners or fellow citizens — is designed to make both parties happy. After all, what good is keeping all of your 100 acres of land if you're happier trading 10 of them for foreign-made goods? Or in the case of the United States during the 1980s, what good is continuing to own Times Square or Columbia Pictures if you'd rather trade them for Honda Accords and Sony televisions? (The anti-Japanese hysteria at the time was even sillier given that the largest group of foreign owners of U.S. property was, and still is, the British!)

Much to the chagrin of economists, the argument that the point of trade is to make you happier doesn't always fly well. Many people view trade as an antagonistic contest to dominate other countries by running trade surpluses so that you eventually own all the other countries' assets. They argue for restrictions on trade designed to rig trade relations so that their own countries always run surpluses. But such policies inevitably fail because anytime you put up a tariff barrier or an import tax to discourage imports and improve your trade balance, other countries can do the same. The result of such trade wars is that all the barriers, restrictions, and taxes imposed by both sides reduce international trade to a trickle. No one comes out ahead, and no one's happy.

Consequently, for the past 70 years, national governments have increasingly pushed for fewer and fewer restrictions on international trade. This *free trade* movement has resulted in hundreds of millions of new jobs and a vast improvement in living standards and happiness because people all over the world are free to trade and buy whatever it is that makes them happiest — even if that means buying from another country.

Wielding a comparative advantage

The argument that even countries running trade deficits are better off because they get to consume a mix of goods and services they couldn't get otherwise rests solely on the benefits of trading things that have already been produced. But an even better argument for international trade is that the trade actually increases the total amount of output produced in the world — meaning that there's more output per person, so overall living standards rise.

English economist David Ricardo developed this argument, known as *comparative advantage*, in 1817 as a forceful rebuttal against the British import tariffs known as the *Corn Laws*, which heavily taxed imports of foreign-grown grain flowing into the United Kingdom. These laws kept the price of grain high, so the nobility who owned the majority of farmland in the British Isles favored keeping them. Naturally, the poor were opposed because the laws drove up the price of their basic food supply: bread.

Ricardo pointed out that abolishing restrictions on international trade would, in addition to helping England's poor, actually make England and all the countries it traded with richer by encouraging them to specialize in the production of goods and services that each of them could produce at the lowest opportunity cost in terms of other goods and services forgone. He demonstrated that this process of specialization would increase total worldwide output and thereby raise living standards.

You can most easily understand the logic behind the comparative advantage argument by thinking in terms of people rather than countries. Consider a patent lawyer named Heather and her brother Adam, a bike mechanic. Heather is very good at filing patents for new discoveries. She's also very good at repairing bicycles — in fact, she's faster at that than her brother. Adam is a smart guy and can file patents, too, although not as quickly as Heather can. Table 14-1 lists how many bike repairs and patent filings each of them could do in one day if each of them put all of his or her effort into only one of the activities.

TABLE 14-1

Productivity for Heather and Adam per Day

Person	Patent Productivity	Bike Repair Productivity
Heather	6	12
Adam	2	10

In one day's work, Heather can produce 6 patents or repair 12 bikes; Adam can file 2 patents or repair 10 bikes. She is more efficient at producing both patents and bike repairs because she can convert one day's labor into more of either good than Adam can. Heather has an *absolute advantage* over Adam at producing both goods. She's the more efficient producer of both; with the same amount of labor input (one workday), she can produce more than her brother. Before David Ricardo came up with the idea of comparative advantage, the only thing anyone knew to look at was absolute advantage. And when they saw situations like that of Heather and Adam, they concluded (incorrectly) that because Heather is more efficient at both tasks, she has no need to trade with him.

In other words, people used to believe that because Heather is better than Adam at repairing bikes, not only should she work hard as a patent attorney filing lots of patents, but she should also fix her own bike whenever it breaks down. Ricardo pointed out that this argument based on absolute advantage is bogus and that Heather should, in fact, *never* fix bikes, despite the fact that she's the most efficient bike repairperson around. The nifty thing Ricardo realized is that the world is better off if each person (and country) specializes.



WARNING

Comparative advantage's insight is that the proper measure of cost when considering whether you should produce one good or another isn't how many hours of labor input it takes you to produce each one (which is the logic behind absolute advantage). The true cost is the opportunity cost of how much of the production of one good you have to give up to produce a unit of the other good.

To produce one patent, Heather must give up the chance to repair two bikes. To make one patent, Adam would have to give up the chance to repair five bikes. Heather is the lower-cost producer of patents and therefore should specialize in filing patents. And Adam should specialize in bike repairs because he's the lower-cost producer of bike repairs. In economic terminology, we say that Heather has a *comparative advantage* in the production of patents, whereas Adam has a *comparative advantage* in the production of bike repairs.

On a larger scale, countries should specialize in the production of the goods and services that they can deliver at lower opportunity costs than other countries. If countries are free to do this, everything that's produced comes from the lowest-cost producer. Because this arrangement leads to the most efficient possible production, total output increases, thereby raising living standards.

Politicians often argue that countries shouldn't be "dependent" on other countries for various goods and services. Any policy that takes this warning seriously by impeding trade and specialization increases costs and makes total output fall. Rather, by letting comparative advantage guide who makes what, free trade increases total world output and thereby raises living standards. In free trade, each country specializes in its area(s) of comparative advantage and trades with other countries for goods and services it desires to consume.

- » Risking inflation by printing too much money
- » Measuring inflation with price indexes
- » Adjusting interest rates to account for inflation

Chapter **15**

Inflation Frustration: Why More Money Isn't Always Good

nflation is the word economists use to describe a situation in which the general level of prices in the economy is rising. Although some prices may stay the same and a few may fall, the majority of prices rise.

Inflation is typically mild, with the overall level of prices rising only a small percentage each year. But people dislike even mild inflation because — face it — who likes paying higher prices? Mild inflation also causes problems such as making retirement planning difficult. After all, if you don't know how expensive things will be when you retire, calculating how much money you need to be saving right now is hard.

Things can go from bad to worse if inflation really gets out of control and prices begin rising 20 or 30 percent per month — something that has happened in more than a few countries in the past century. Such situations of *hyperinflation* usually accompany a major economic collapse featuring high unemployment and a major decrease in the production of goods and services. (For more on prices and how they affect the economy, see Chapter 16.)

The good news is that economists know exactly what causes inflation and how to stop it. The culprit is a money supply that grows too quickly, and the solution is to slow or halt the growth of the money supply. Unfortunately, some political pressure is always exerted in favor of inflation so that simply knowing how to prevent inflation doesn't necessarily mean it won't develop.

In this chapter, I tell you about money and inflation, including why governments are often tempted to print a lot of money to pay for budget deficits, why doing so is actually a form of taxation, and why some constituency is always encouraging the government to go ahead and print a ton of money. I also show you why printing lots of money causes inflation, how to measure inflation, and how to measure the effect that inflation has on interest rates. One thing I don't tell you is how to print your own money — this ain't *Counterfeiting For Dummies*.

Buying an Inflation: When Too Much Money Is a Bad Thing

I don't think I can overstate how important money is to the proper functioning of the economy. Without it, you'd waste most of your time *bartering*, or arranging trades of one good for another — kind of like in kindergarten. ("I'll trade you my cookie for your brownie.") Bartering works well only in the rare circumstance that you run into somebody who has what you want and who wants what you have. In contrast, money provides a medium of exchange that allows you to trade for the brownie from the kid next to you, even if you don't have a cookie.

Money can be any good, object, or thing — including electronic inscriptions on "blockchain ledgers," as is the case with cryptocurrencies. Money's defining characteristic is that it's accepted as payment for other goods and services. In today's economy, people pay for things using an assortment of monies, including government-issued coins and cash, checks drawn on private bank deposits, and electronic payments facilitated by debit cards and cryptocurrencies like Bitcoin and Ether. Because money affects nearly every economic transaction that takes place, it's at the heart of *macroeconomics*, the study of the economy as a whole.

As with everything in life, balance is essential. If a government prints too much money, prices go up, and inflation results. If a government prints too little, prices go down, and deflation results. But how much money is the right amount? And why does printing too much or too little cause inflation or deflation? In this section, I go into more detail about these money matters.

Balancing money supply and demand

Basically, money's value is determined by supply and demand, which I discuss in detail in Chapter 4:

- » The *supply* of money is under government control, and the government can easily print more money anytime it wants to.
- » The *demand* for money derives from its usefulness as a means of paying for things and from the fact that having money means not having to engage in barter.

For any given supply of money, supply and demand interact to set a value for each unit of money. If money is in short supply, each piece of money is valuable; fewer pieces of money translate into fewer chances to avoid having to engage in barter. But if the government greatly increases the supply of money, each unit of money loses value because getting enough money together to avoid barter is easy.



TIP

Prices and the value of money are *inversely related*, meaning that when the value of money goes up, prices go down (and vice versa). To see how this relationship works, suppose that money is in short supply and is consequently very valuable. Because it's so valuable, it buys a lot of stuff. For instance, 1 dollar may buy 10 pounds of coffee (that's 10 cents per pound). But if money's common, each unit isn't very valuable. In this case, 1 dollar may buy only 1 pound of coffee (that's 1 dollar per pound). So the greater the supply of money, the higher the prices.

The demand for money tends to grow slowly over time. Growing economies produce more stuff, and consumers demand more money to buy the available stuff. Depending on how a government reacts to consumer demand for more money, three scenarios are possible:

- » **The government increases the supply of money at the same rate as the growing demand for money.** In this case, prices don't change. In other words, if supply and demand for money grow at equal rates, the relative value of money doesn't change.
- » **The government increases the supply of money faster than the demand for money grows.** Here, inflation results as money becomes relatively more plentiful and each piece of money becomes relatively less valuable. With each piece of money possessing less value, you need more pieces of money to buy any particular item — which is the same as saying that the item's price will rise.
- » **The government increases the supply of money slower than the demand for money grows.** In this case, deflation results because each piece of money grows relatively more valuable. Buying any given good or service requires less money.

CROESUS AND KUBLAI: THE KINGS OF MONEY

King Croesus of Lydia is usually given credit for solving the problem of bogus metal money. In the sixth century BC, Croesus issued the first government-certified coins that guaranteed purity and weight.

Lydia was located in what is now western Turkey, and soon all the major trading nations of the Mediterranean were using the new Lydian coins because they were by far the most trustworthy medium of exchange available. The new coinage gave Lydian traders a major advantage, and the kingdom soon became very wealthy, so much so that Croesus was considered the richest man in the world — even richer than King Midas (of *Midas touch* fame), whose gold Croesus minted into coins.

But coins are hard to carry around in large amounts, and it was up to the Chinese emperor Kublai Khan to create the first paper money in the 13th century. This paper money was actually a kind of precious metal certificate; people holding one of these certificates could go to a government vault and redeem it for gold. Consequently, the pieces of paper were as good as gold, but a stack of paper was a whole lot easier to carry than a heavy bag of coins.

Paper money was such a radical innovation that when Marco Polo came back from China and told Europeans about it, they laughed, unable to conceive of anything other than gold or silver coins serving as money. Their incredulity was hard to overcome, and after paper money fell out of favor in China, it would be centuries before another government issued any again.



REMEMBER

You may be wondering whether there's any way to know exactly how much inflation you can expect from printing any given amount of extra money. You're in luck! The *quantity theory of money* states that the overall level of prices in the economy is proportional to the quantity of money circulating in the economy. *Proportional* just means that things go up by equal ratios or proportions, so the quantity theory can also be stated this way: If you double the money supply, you double the prices.

But *why* would any government want to cause inflation or deflation of any size whatsoever? For the answer to that question, read on.

Giving in to the inflation temptation

Inflation is caused primarily by governments printing more paper money or producing a large amount of cheap-metal coins, which vastly increases the supply of money and makes each piece of money less precious. Because sellers demand higher prices to make up for the fact that each piece of money is worth less, inflation results.



REMEMBER

So why would a government ever opt to increase the money supply so fast that inflation might occur? Good question. Historically, governments circulate more money in three circumstances:

- » When governments can't raise enough tax revenue to pay their obligations
- » When governments feel pressure from debtors who want inflation so they can repay their debts using less valuable money
- » When governments want to try to stimulate the economy during a recession or depression

As you find out more about these three reasons for increasing the money supply, keep the following in mind: If the supply of money increases faster than the demand for money, inflation results. (For more on this, see the earlier section “Balancing money supply and demand.”) Consequently, no matter what reason a government has for increasing the supply of money, it runs the risk of inflation. And that’s true both for good reasons, such as wanting to help the economy out of a recession, and for bad ones, such as helping debtors repay their loans using less valuable money.

Paying bills by printing bills: Hyperinflation

Governments almost always have debts, and printing extra money can be a tempting way to pay them. Quite often, a government may want to spend more money than it’s collecting in tax revenue. One solution is to borrow the shortfall, but another is to simply print new bills to cover the difference.

The trouble with allowing a government to pay its debts and obligations by printing money is that as soon as the newly printed money begins to circulate, people spend it, drive up prices, and cause inflation. And if the government prints successively more money, you end up with people offering shopkeepers and producers successively more cash for the same amount of goods and services. It’s like a giant auction where everybody bidding on items keeps getting more and more money to bid with.

PREVENTING INFLATION WITH A GOLD STANDARD

Until the early 1970s, printing new bills was difficult because most of the world's paper currencies were backed by a valuable metal, such as gold. Under this system, every piece of paper money circulating in the economy was convertible into a specific quantity of gold so that anyone holding cash could redeem their cash for gold whenever they wanted. For instance, in the United States, you could bring \$35 cash to the U.S. Treasury and get exactly 1 ounce of gold.

This *gold standard* made it difficult for the government to devalue the currency by printing too much money because it first had to get more gold with which to back the new money. Because purchasing gold is expensive, governments were effectively restrained from increasing their money supplies.

But in 1971, President Nixon took the United States off the gold standard and on to the *fiat system*. Under fiat, paper currency isn't backed by anything. People just have to accept the currency as though it has value. In fact, *fiat* is Latin for "Let it be!" So when you say *fiat money*, you're basically referring to how a government creates money simply by ordering it into existence. The problem with a fiat money system is that nothing limits the number of little pieces of paper that the government can print to pay its debts.



REMEMBER

If a government gets into the habit of rapidly printing new money to pay its bills, inflation can soon reach or even surpass 20 or 30 percent per month, a situation referred to as *hyperinflation*. Economists (and everyone else) hate hyperinflation because it greatly disrupts daily life and ruins the investment climate. Hyperinflation does the following:

- » **Causes people to waste huge amounts of time trying to avoid the effects of rising prices:** During the Weimar hyperinflation in Germany (which I discuss in the sidebar "Hyperinflation and Hitler"), men working at factories were paid onsite two or even three times a day because money lost its value so quickly. Their wives waited outside the factories to immediately take the money to the nearest shops, trying to spend the pay before it lost most of its value. Shopping may be fun, but not when you're desperately racing against outrageously rising prices!
- » **Destroys the incentive to save:** The only sensible thing to do with money during hyperinflation is to spend it as quickly as you can before it loses even more of its value. People whose life savings were in German marks during the

Weimar hyperinflation soon found that what they had worked so hard to amass had become worthless. And people thinking about saving for the future were greatly discouraged because they knew that any money they saved would soon lose all value.

» **Harms businesses:** The discouragement of saving causes major business problems because if people aren't saving, no money is available for businesses to borrow for new investments. And without new investments, the economy can't grow.

HYPERINFLATION AND HITLER

History's most infamous hyperinflation hit Germany in the 1920s during the economically incompetent Weimar Republic. It so badly ruined the German economy that Germans would later vote Adolph Hitler into power because he promised to fix things.

At the end of World War I, Germany faced the prospect of paying off massive debts taken on during the conflict in addition to all the ongoing costs of running a government. Most of its debts were in its own currency, the German mark.

Because the German government had the exclusive right to produce German marks, the debt proved an irresistible temptation to begin printing money to pay its bills. If the government owed a billion marks to a certain firm, it simply printed up a billion crisp, new mark bills and handed them over. If a bunch of schoolteachers hadn't been paid the previous month, the Weimar government simply printed up enough new cash to pay them.

Soon all the new money caused a wild hyperinflation. In fact, the rate of inflation in Weimar Germany in 1922 was well over 100 percent per month — and it reached nearly 6,000 percent by the end of that year!

Then things *really* got out of control. Prices went up 1,300,000,000,000 times (this is not a misprint!) in 1923. That year, Germans paid 200,000 marks for a loaf of bread and 2 million marks for a pound of meat. Prices rose so rapidly that waiters at restaurants had to pencil in new prices on menus several times a day. And if you ate slowly, you were sometimes charged twice as much as what was printed on the menu because prices had gone up exponentially while you were eating! In some places in Germany, people stopped bothering to count out money. Instead, they tied paper bills into huge bricks and weighed the bricks of cash. For instance, it may have cost 2 pounds of cash to buy a chicken.

Feeling printing press pressures: Inflation politics

Even if the government isn't trying to use inflation to increase tax revenues (see the preceding section), a certain political constituency will always pressure it to circulate more money. You may even be a member of this group — they're called *borrowers*. To understand the politics of inflation, understand that one of the functions of money is to act as a *standard of deferred payment*. What does that mean?

Well, imagine that you borrow \$1,000 to invest on your farm, promising to pay the bank back \$1,200 next year. For the past several years, prices in the economy have been stable, and, in particular, the pigs that you raise have sold for \$100 each. Essentially, your loan lets you borrow the equivalent of 10 pigs with the promise to pay back 12 pigs next year.

But you've got an idea. You lobby your congressperson to in turn lobby the government to print more money. All that new money causes inflation, after which the price of pigs rises to \$200 each. Now you have to sell only six pigs to pay back the \$1,200 loan, leaving you with more pigs. You pig!

Lenders, of course, oppose the inflationary desires of borrowers. If you were the bank, you would do everything in your power to stop the inflation. If it goes through, not only are your profits ruined, but you're an outright loser. In the first year, your loan of \$1,000 is the equivalent of ten pigs. But after the inflation, you get paid back the equivalent of only six pigs. You take a 40 percent loss on the value of your loan. Too much inflation, and a lender ends up being a pig in a poke.

As long as economies use money, lenders and borrowers will always be lined up against each other, both trying to sway the government.

Stimulating the economy with inflation: Monetary policy

A legitimate reason for governments to print more money has the respectable name of monetary policy. *Monetary policy* refers to the decisions a government makes about increasing or decreasing the money supply to stimulate or slow down the economy.

I go into monetary policy in detail in Chapter 17, but the basic idea is that if the economy is in a recession, the government may print up some new money and spend it. All the goods and services it buys with the new money stimulate the economy immediately. All those businesses that received money from the government can now also go out and spend that new money themselves. And whoever receives the money from them will also go out and spend it to buy things. In fact, this can theoretically go on forever and stimulate a lot of economic activity — enough to lift an economy out of a recession.



REMEMBER

If this sounds too good to be true, it is. Why? One word: inflation. When people start spending all that new money, it drives up prices. Eventually, the only effect of the government's good intentions is that prices rise and no additional goods are sold. For example, if the government doubles the money supply, businesses will double the prices they charge because each piece of money is worth half as much as before. Consequently, the total amount of goods and services sold will be the same as before because although twice as much money is being spent, prices are also twice as high.

The sad upshot is that an increase in the money supply stimulates the economy only when it's a surprise. If the government can print the money and start spending it before people can raise prices, you get an increase in the amount of goods and services sold. Eventually, of course, people figure it out and raise prices, but until they do, the monetary stimulus works.

Unfortunately, fooling people over and over again is tough. You can surprise people once, but it's much harder to surprise them the second time and even harder the third time. In fact, if the government keeps trying to surprise people, people begin to anticipate the government's actions, and they raise prices even before the government prints more money. Consequently, most modern governments have decided against using this sort of monetary stimulus and now strive for zero inflation or very low inflation.

Tallying up the effects of inflation

In the United States, prices normally rise only a small amount each year. However, even moderate inflation causes problems by cutting into the practical benefits of using money instead of barter. You can get a better sense of this fact by looking at the four functions that economists generally ascribe to money and the ways in which inflation messes with each of them:

» **Money is a store of value.** If I sell a cow today for one gold coin, I should be able to turn around and trade that gold coin back for a cow tomorrow or next week or next month. When money retains its value, you can hold it instead of holding cows, or real estate, or any other asset.

Inflations weaken the use of money as a store of value because each unit of currency is worth less and less as time passes.

» **Money is a unit of account.** When money is widely accepted in an economy, it often becomes the unit of account in which people write contracts. People start using phrases like "\$50 worth of lumber" rather than "50 square feet of lumber," or "\$1 million worth of shirts in inventory" instead of "20,000 shirts in inventory."

This practice makes sense if money holds its value over time, but in the presence of inflation, using money as a unit of account creates problems because the value of money declines. For instance, if the value of money is falling fast, how much lumber, exactly, is “\$50 worth of lumber”?

» **Money is a standard of deferred payment.** If you want a cow, you probably wouldn't borrow a cow with the promise to repay two cows next year. Instead, you'd be much more likely to borrow and repay in terms of money. That is, you'd borrow one gold coin and use it to buy a cow, after promising to pay back two gold coins next year.

The progressive devaluing of money during a period of inflation makes lenders reluctant to use money as a standard of deferred payment. Suppose a friend asks to borrow \$100, promising to pay you \$120 in a year. That seems like a good deal — after all, it's a 20 percent interest rate. But if prices are rapidly rising and the value of money is falling, how much will you be able to buy with that \$120 next year?

Inflations make people reluctant to lend money. Potential lenders fear that when the loans are repaid, the repayment cash won't have the same purchasing power as the cash that was lent. This uncertainty can have a devastating effect on the development of new businesses, which rely heavily on loans to fund their operations.

» **Money is a medium of exchange.** Money is a *medium* (literally meaning “something in the middle”) of trade between buyers and sellers because it can be directly exchanged for anything else, making buying and selling much easier. In a barter economy, an orange farmer who wants to buy beer may have to first trade oranges for apples and then apples for beer because the guy selling the beer wants only apples. Money can eliminate this kind of hassle.

But if inflation is bad enough, money is no longer an effective medium of exchange. During hyperinflations, economies often revert to barter so buyers and sellers don't have to worry about the falling value of money. For example, in a healthy economy, the orange seller can first sell oranges for cash and then trade the cash for beer. But during a hyperinflation, between the time he sells the oranges for cash and buys the beer, the price of beer may have skyrocketed so high that he can't buy very much beer with the cash. During hyperinflation, economies have to resort to cumbersome bartering.



REMEMBER

Another effect of inflation is that it functions as a giant tax increase. This seems strange because you normally think of governments taxing by taking away chunks of people's money, not by printing more money. But a tax is basically anything that transfers private property to the government. Debasing the currency or printing more money can have this effect.

Suppose that the government wants to buy a \$20,000 van for the post office. The honest way to go about this is to use \$20,000 of tax revenues to buy a van. But a sneakier way is to print \$20,000 in new cash to buy the van. By printing and spending the new cash, the government has converted \$20,000 of private property — the van — into public property. So printing new cash works just like a tax. Because printing new money ends up causing inflation, this type of taxation is often referred to as an *inflation tax*.

Not only is the inflation tax sneaky, it unfairly targets the poor because they spend nearly all their incomes on goods and services, the prices of which go up greatly during an inflation. By contrast, because the rich have the opportunity to save a lot of their incomes, proportionately they're less affected by an inflation tax. By investing their savings in assets whose prices go up during an inflation (such as real estate), the rich can insulate themselves from a great deal of the harm caused by inflation.

Measuring Inflation

Inflation can cause lots of problems, so for the government to keep inflation under control, it needs a way to measure inflation accurately.

The value of money is determined by the interaction of the *supply* of money with the *demand* for money (as I explain in the earlier section “Buying an Inflation: When Too Much Money Is a Bad Thing”). The supply of money is under the government's control, but the government can't directly ascertain the demand for money; it has to look at how supply and demand interact to determine how much to increase or decrease the money supply:

- » If inflation is in effect, the government knows that the supply of money is increasing faster than the demand for money. If it wants to tame the inflation, it should reduce the supply of money.
- » If deflation is in effect, the government knows that the demand for money is increasing faster than the supply of money. If it wants to end the deflation, it should increase the supply of money.



REMEMBER

Economists arbitrarily define some large collection of goods and services and refer to it as a *market basket*. They then measure inflation by finding out how much money it takes to buy this basket at various times. The best-known market basket is monitored by the Bureau of Labor Statistics. This basket is called the *consumer price index*, or CPI. It consists of what the Bureau thinks a typical family of four buys in the United States each month.

In this section, I show you how this process works by creating a market basket, noting how it can be used to measure inflation, and normalizing it to a given base year so that calculating inflation rates between any two years is a breeze. (If I’ve piqued your interest with this talk of market baskets and the CPI, feel free to check out the Bureau of Labor Statistics CPI website at www.bls.gov/cpi.)

Creating your very own market basket

The CPI involves a large number of products and services — it’s a big market basket. Understanding price indexes is easier if you create a simplified index with a small market basket. In this section, I look at a small market basket containing pizza, beer, and textbooks — typical purchases of college students. I call it the collegiate price index.

For each of the three items in the collegiate price index, I’ve created prices for 2027, 2028, and 2029 and listed them in Table 15-1.

TABLE 15-1

The Collegiate Price Index

Item	Number Bought	2027	2028	2029
Pizza	5	\$20	\$18	\$18
Beer	30	\$4	\$4	\$4.50
Textbooks	1	\$120	\$160	\$170

In 2027, one medium cheese pizza costs \$20, a cold bottle of beer costs \$4, and an overly long, incomprehensible introductory economics textbook costs \$120. The next year, the price of a medium cheese pizza actually falls to \$18 because a new pizza parlor opens up next to the old one, causing a price war. Beer still costs \$4, but the college bookstore decides that it can really stick it to students, raising the price of the textbook to \$160. (Don’t worry about the 2029 column yet. I give you a chance to dig in and calculate inflation using the 2029 numbers later in the chapter.)

So far, so good. But in evaluating the index, you also have to keep track of how many of each item the typical student buys each year. For the sake of simplicity, assume that a typical student buys 5 cheese pizzas, 30 beers, and 1 economics textbook each year.

Calculating the inflation rate

To calculate how much inflation your college economy has (or deflation, if the price of this basket happens to go down), first total how much the market basket costs each year using Table 15-1. In 2027, it costs \$340: \$100 on pizza (5 pizzas at \$20 each), \$120 on beer (30 beers at \$4 each), and \$120 on economics textbooks (1 textbook at \$120). The cost of buying the same market basket in 2028 is \$370, so the cost of buying the same market basket has gone up by \$30.

Now that you've done the addition, you need some simple algebra. Economists use the capital letter P to denote how many dollars the defined market basket costs. So in this case, P_{2027} means the cost of buying the market basket in 2027, and P_{2028} is the cost of buying the market basket in 2028. Economics has a standard practice of denoting the rate of inflation with the Greek letter π (pronounced "pie").

To calculate the rate of inflation, you use a simple formula:

$$\pi = \frac{P_{\text{SecondYear}} - P_{\text{FirstYear}}}{P_{\text{FirstYear}}} \quad (1)$$

In the case of the collegiate price index, the formula becomes:

$$\pi = \frac{P_{2028} - P_{2027}}{P_{2027}} \quad (2)$$

Substituting in $P_{2027} = \$340$ and $P_{2028} = \$370$, you find that $\pi = 0.088$. Convert this number into a percentage by multiplying by 100, and inflation in the collegiate price index is 8.8 percent between 2027 and 2028. Based on this number, a student needs 8.8 percent more money in 2028 to buy the simple market basket.

Setting up a price index

The collegiate market basket (see Table 15-1) is a simple example. When government statisticians compute the CPI, they basically do the same thing, just using a lot more goods. They also introduce the concept of a *price index* (or *price level index*) to make calculating and interpreting inflation rates over several years much easier. To set up a price index, statisticians first establish a base year, or index year. Continuing the example from Table 15-1, suppose that 2027 is the base year for the collegiate price index. You can then make a handy mathematical transformation so that the price level in 2027 is fixed at the number 100 and the price levels of every other year are set up so that they're relative to the 100 of the base year.



TIP

To make $P_{2027} = \$340$ your base year, divide it by itself. That, of course, gives you 1, which you then multiply by 100 to get 100 ($100 \times 1 = 100$). This may seem like an idiotic thing to do until you realize that if you do the same thing to the other years, you end up with something very useful. Divide P_{2028} by P_{2027} and then multiply that

quotient by 100 to get 108.8. This number is easy to interpret: It's 8.8 percent larger than 100. Or, put differently, the price level in 2028 is 8.8 percent larger than the price level in 2027.

You can keep going, using the numbers for 2029 that appear in Table 15-1. For instance, $P_{2029} = \$395$. If you divide P_{2029} by P_{2027} and multiply by 100, you get 116.2; the price level in 2029 is 16.2 percent bigger than the price level in 2027.

Figuring the rate of inflation between 2028 and 2029 using these index numbers is also easy. Because the price index level for 2028 is 108.8 and the price index level for 2029 is 116.2, inflation is simply $\frac{116.2 - 108.8}{108.8} = 0.068$, or 6.8 percent. (You're using equation [1] here, but you're inputting index numbers instead of the actual costs of market baskets.)

Figure 15-1 charts the actual values of the CPI from 1983 to 2022. The index was set to a level of 100 using prices that consumers paid on average over the two-year period from 1982 to 1984.

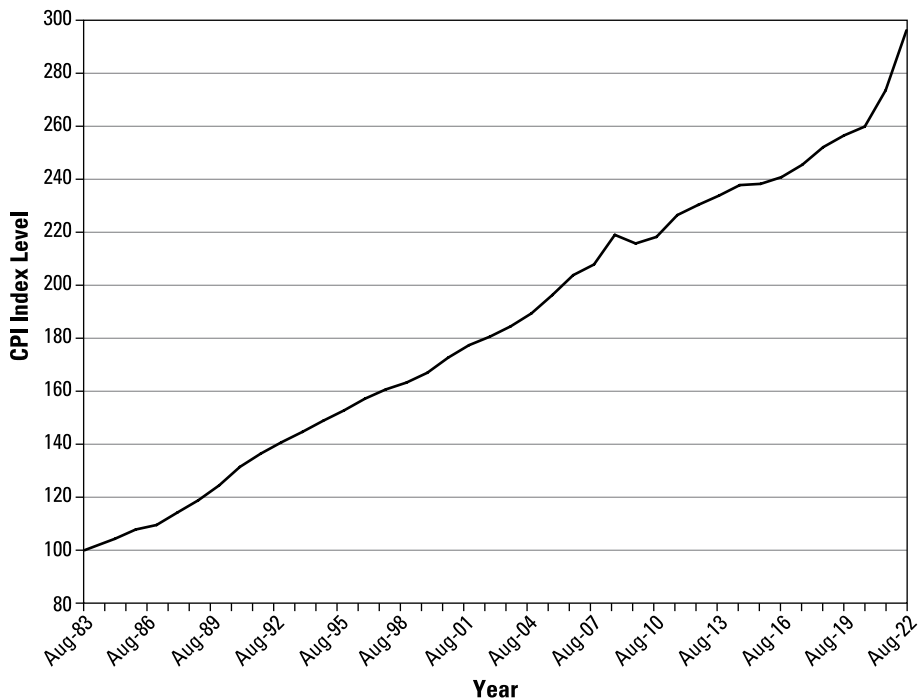


FIGURE 15-1:
Consumer
Price Index,
1983–2022.

© John Wiley & Sons, Inc.

You can see that the CPI grew from its initial level of 100 in 1983 to a level of 296 in 2022. That is, to buy what a typical family of four consumes, you would

have needed 196 percent more money in 2022 than you did in 1983. Increases in the money supply caused the price level in the United States to nearly triple over those 39 years. The average annual inflation rate over those 39 years was 2.8 percent per year.

Determining the real standard of living with the price index



REMEMBER

Beyond making inflation easy to measure and interpret, price indexes also make it simple to measure the important difference between real prices and nominal prices. *Nominal prices* are simply money prices, which can change over time due to inflation. Because nominal prices can change, economists like to focus on *real prices*, which keep track of how much of one thing you have to give up to get another thing, no matter what happens to nominal prices.

For example, suppose that in 2027 you make \$15 an hour working at a youth camp and that the cost of a museum ticket is \$30. The *real cost* of a museum ticket to you is two hours of work. Suppose that the next year, the prices of all goods double. But your wages also double, so you're earning \$30 an hour, and a museum ticket costs \$60. The result is that you still have to work two hours to buy a museum ticket. So although the *nominal* price of a museum ticket has doubled, its *real* price in terms of labor — how much labor you have to give up to purchase a museum ticket — hasn't changed.

By constructing price indexes like the CPI, economists can tell how the *real standard of living* changes for people from year to year. In the collegiate price index example from preceding sections (using data from Table 15-1), inflation is 8.8 percent between 2027 and 2028, meaning that the cost of living of a typical college student went up 8.8 percent. So if at the same time student incomes go up only 5 percent, students are actually worse off because costs have gone up faster than incomes. Real living standards — measured in terms of how much stuff you can buy with your income — have fallen.

Identifying price index problems

Using price indexes to track the cost of living isn't a flawless system. Here are three big issues:

- » **The market basket can never perfectly reflect family spending.** The Bureau of Labor Statistics tries to keep track of what a typical family of four

purchases when calculating the CPI. But families differ greatly, not only in terms of *what* they buy but also in terms of *how many* of each thing they buy.

- » **The market basket becomes outdated.** The Bureau of Labor Statistics often waits way too long before including new types of goods in the market basket. For instance, the Bureau took years to include online video-streaming services like Netflix even after they became very popular. The same was true of ignoring the rise of Uber and Lyft for several years while still putting substantial weight on traditional taxi rides. If the CPI fails to include popular new products, it's not fully capturing the price changes that matter to consumers.
- » **The market basket can't account for quality.** Price isn't the only thing that matters to consumers. For example, what if a beer stays the same price but improves in quality from one year to the next? You're getting better beer for the same price, but this isn't reflected in the data. This problem is especially severe for things like computers, smartphones, and video games. For these products, quality improves dramatically year after year while prices often either stay the same or go down.

Each of these problems troubles government statisticians, who are constantly coming up with better price indexes and statistical methods to try to overcome them. In the meantime, the Federal Reserve (the government agency charged with regulating the money supply in the United States) has estimated that the CPI overstates inflation by 1 to 2 percentage points per year. Most of the overstatement comes from the failure of the CPI to account for new goods and quality improvements.

The main consequence of this overstatement is that the government is overly generous with the cost-of-living increases it grants workers and retirees. Each year, government workers and retirees receive pay increases based upon increases in the CPI. These pay increases are designed to ensure that people's real incomes aren't eroded by inflation, but because the CPI is most likely overstating the rate of inflation each year, the cost-of-living increases are overly generous.

On the other hand, the rise of popular free internet services has caused an opposite problem. Millions of people derive great enjoyment from the services that Facebook, Google, and other internet companies provide. But those services are offered free of charge. So it's not at all clear how one would be able to incorporate them into a price basket or keep track of how inflation in the overall economy affects the demand or supply for these services — or how to account for quality improvements to free online services.

Pricing the Future: Nominal and Real Interest Rates

Because inflation erodes the value of a loan repayment (see the earlier section “Tallying up the effects of inflation”), economists have to distinguish between nominal interest rates and real interest rates. *Nominal interest rates* are simply the normal money interest rates that you’re used to dealing with; they measure the returns to a loan in terms of money borrowed and money returned. *Real interest rates*, however, compensate for inflation by measuring the returns to a loan in terms of units of stuff lent and units of stuff returned. This distinction is important because the *real* interest rate is what makes people want to save and invest. After all, what lenders really care about isn’t how much money they get back but how much stuff they can buy with it.

Suppose that you borrow \$1,000 with the promise to pay \$1,100 to the lender in a year. Your nominal interest rate is 10 percent because you’re paying back an additional \$100, or 10 percent more dollars than you borrowed. But if inflation occurs, the amount of stuff that \$100 can buy will decrease over time.

Say a nice meal for two with a bottle of wine costs \$100 right now but will cost \$105 next year. The lender is giving up 10 of these very good meals (\$1,000 divided by \$100 per meal) to give you the loan. Next year, when the lender is repaid \$1,100, they can buy 10.47 meals at the price of \$105. They’re giving up 10 meals now in exchange for 10.47 meals next year, meaning that the real rate of interest on the loan is 4.7 percent. Because of inflation, the real interest rate on the loan is substantially less than the nominal rate.



REMEMBER

When lenders and borrowers negotiate a nominal interest rate on a loan, they both try to estimate what the inflation rate will be over the period of the loan. This *expected rate of inflation* is denoted algebraically as π^e . (Don’t confuse expected inflation, π^e , with actual inflation, π . Expected inflation is what people expect to happen ahead of time, whereas actual inflation is what actually ends up happening.)

In this section, I show how to estimate and use the expected rate of inflation.

Using the Fisher equation

Economist Irving Fisher came up with a simple formula that links nominal and real interest rates. Here’s the *Fisher equation*, using i to denote the nominal interest rate and r to denote the real interest rate:

$$i = r + \pi^e \quad (3)$$

This equation simply says that the nominal interest rate is the real interest rate plus the expected rate of inflation. This relationship is important to borrowers and lenders because although all loan contracts specify a nominal rate of interest, their goal is to achieve a specific real rate of interest, even after any subsequent inflation reduces the value of money. By using the Fisher equation, the borrowers and lenders can determine what nominal interest to charge now to achieve a given real rate of return, taking into account the expected rate of inflation.

Suppose that a borrower and lender agree that 6 percent is a fair real rate of interest and that inflation is likely to be 3.3 percent over the course of one year. Using the Fisher equation, they write the loan contract with a 9.3 percent nominal interest rate. A year later, when the borrower repays the lender 9.3 percent more money than was borrowed, that money is expected to have only 6 percent more purchasing power than the borrowed money, given the expected increase in prices.

Realizing that predictions aren't perfect

Negotiations like the one described in the preceding section depend crucially on estimating the expected inflation rate, π^e . Lots of economists have job descriptions consisting primarily of trying to predict future inflation rates. Their predictions are widely reported in the business media, but every person comes up with inflation forecasts in their own way. Some listen to the experts, and others make estimates based on their own daily experiences.

During the period of low inflation starting in the early 1980s and extending through the many years of recovery after the Great Recession of 2007–2009, people's inflation expectations were remarkably accurate. For example, consumers polled by researchers at the University of Michigan each January were on average no more than a couple of percentage points off the actual rate of inflation over the next 12 months. In addition, the majority of those polled had similar predictions about inflation. In that sort of an environment, people trusted their own inflation expectations and tended to agree with other people's inflation expectations. Agreeing on the nominal interest rates to be charged on particular loans was consequently relatively easy for borrowers and lenders.

But during periods of high inflation, actual rates of inflation tend to be quite unpredictable and volatile — with, say, a 35 percent rate one year, followed by a 15 percent rate the next year, and then a 45 percent rate the year after. Thus, in periods of high inflation, individuals tend not to trust their own inflation forecasts or to agree with the inflation forecasts of others. In that sort of environment, borrowers and lenders find it difficult to agree on the nominal interest rates to be charged on particular loans. The resulting declines in lending and borrowing can reduce investment, and consequently, the economy's ability to produce goods and services.

- » Visualizing the business cycle
- » The ideal: Letting price adjustments eliminate recessions
- » The reality: Coping with sticky prices and lingering recessions
- » Linking slow price adjustments to slow wage adjustments
- » Introducing the Keynesian model

Chapter 16

Understanding Why Recessions Happen

Macroeconomists' most daunting task is to try to prevent — or at least shorten — *recessions*, those periods of time during which the economy's output of goods and services declines. Economists, politicians, and most other people who work for a living despise recessions because of the toll they exact in human suffering. That's because when output falls, firms need fewer workers. The typical result is massive layoffs, which cause significant increases in unemployment. In large countries such as the United States, millions of workers lose their jobs and their ability to support themselves and their families.

In this chapter, I use the *aggregate supply/aggregate demand* model to show you how economists analyze recessions. Typically, recessions begin with what economists like to call *shocks* — unexpected bad events such as terrorist attacks, natural disasters, the introduction of bad government policies, or sudden spikes in the cost of important natural resources such as oil.

The first big lesson of this chapter is that if the prices of goods and services in the economy were free to adjust to the changes in demand and supply that shocks cause, the economy would typically be able to recover quite swiftly. Unfortunately,

the second big lesson is that not all real-world prices are free to adjust. Rather, some important prices are quite slow to adjust. Economists like to say that they're *sticky*. As a result, recessions can linger and cause a lot of harm unless the government intervenes in ways that can help the economy recover more quickly. (In Chapter 17, I discuss the best ways for governments to intervene.)

Introducing the Business Cycle



REMEMBER

Economies go through alternating periods during which the total output of goods and services expands and then contracts. The alternating pattern of economic expansion and contraction, illustrated in Figure 16-1, is often called the *business cycle* because businesses are so greatly affected by the changes in output.

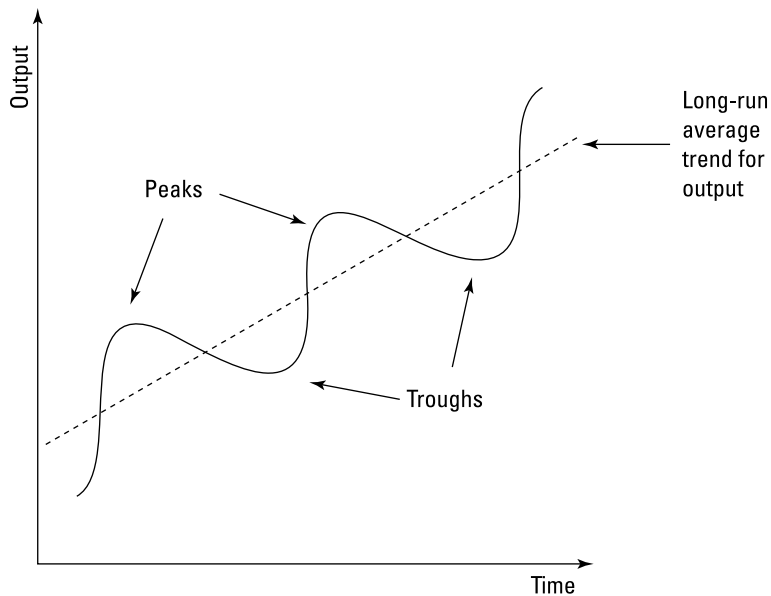


FIGURE 16-1:
The business cycle.

© John Wiley & Sons, Inc.



REMEMBER

The solid line in Figure 16-1 represents how the economy's total output of goods and services varies over time. It alternates between troughs and peaks, which helps you identify periods of recession and recovery. Here's how you can distinguish between the two:

» **Recessions:** Recessions, or contractions, are the periods of time during which output falls — that is, after a peak and before the next trough.

» **Recoveries:** Recoveries, or expansions, are the periods of time during which output increases — that is, after a trough and before the next peak.

The dotted line in Figure 16-1 represents the long-run, average growth trend for an economy's total output of goods and services. I've drawn Figure 16-1 so that it has an upward-sloping average growth trend, capturing the fact that the economies of most countries now have sustained economic growth. In other words, on average, total output tends to rise year after year. Because recessions still happen, however, the actual path given by the solid line fluctuates around the long-run growth path given by the dotted line.

In Figure 16-1, you can see that macroeconomic policy has two natural goals:

- » **Make the long-run average growth line as steep as possible.** The steeper it is, the faster (on average) total output and living standards rise.
- » **Reduce the size of business cycle fluctuations around the long-run average growth line.** Smaller distances between peaks and troughs translate into fewer people suffering through bouts of unemployment when output falls.

In Chapter 17, I explain the policies that economists think are best for achieving these two goals. But for Chapter 17 to make sense, I must first explain what causes the business cycle — especially recessions and the high rates of unemployment that accompany them. After all, if you don't understand what's wrong, you can't sensibly fix it.

Striving for Full-Employment Output

Before you can say whether an economy is doing well or doing poorly, you need some objective standard of what “doing well” is. Economists use the concept of *full-employment output* (which is represented by the symbol Y^*) as their measure of how well an economy should be doing.



REMEMBER

The idea of full-employment output revolves around the concept of *full employment*, by which economists mean a situation in which everyone who wants a full-time job can get one. Full-employment output is the total amount of goods and services produced in the economy when there's full employment in the labor market.



WARNING

Please don't confuse full-employment output with the economy's *maximum output*, which is the larger amount of goods and services that would be produced if everyone worked as much as humanly possible. Also, don't assume that full employment is the same thing as having a zero unemployment rate. Even when everyone who wants a job can get one, some unemployment will always exist because people voluntarily quit one job to search for a better one. For the duration of their job search, these people are counted as unemployed. Economists call this situation *frictional unemployment*, as though the delay in finding a better job is due to some sort of friction that causes a delay in how quickly job seekers can find a new job.



TIP

As technology improves, full-employment output (Y^*) grows because better technology means that a fully employed labor force can produce more output. But to simplify their analyses, economists usually ignore the long-term growth trend and look only at whether actual output, Y , is currently above or below their best estimates of Y^* at that particular moment. (I follow this convention, too, for the rest of the chapter. You will find out how the economy adjusts to situations in which output is either above or below potential output at a given point in time.)

As I show you in this chapter, the economy naturally tends to adjust back to Y^* anytime it deviates from Y^* . If that adjustment process were rapid enough, you wouldn't have to worry about business cycles, recessions, and unemployment. If the economy reverted back to Y^* fast enough, recessions would be too brief to cause any serious negative consequences. Unfortunately, the natural adjustment process can be slow, which means recessions can be quite lengthy and awful.

Returning to Y^* : The Natural Result of Price Adjustments

After an economic shock, such as a natural disaster or a spike in the cost of natural resources, price adjustments tend to return an economy to producing at full-employment output (Y^*). (See “Striving for Full-Employment Output” earlier in this chapter for details on full employment.) That's right, I said *price adjustments* — not the president, and not the chairman of the Federal Reserve Board. Don't believe me? Read on.

Consider a situation in which the overall demand for goods and services in the economy falls off: Individuals, firms, and the government demand and buy less than the economy is currently producing. The result is an excess supply of output that, in turn, leads to lower prices. After all, what does any business do when it

can't sell off everything it's producing at the prices it's currently charging? It has a sale. It lowers prices. The lower prices attract more buyers, and soon the business is able to sell off the rest of its output.



REMEMBER

This process repeats itself all over the economy during an economic downturn. When aggregate demand falls off due to an economic shock, firms lower prices to make sure they sell off their outputs. This process eventually leads to two outcomes:

- » Prices all over the economy fall.
- » The economy again produces at full-employment output, Y^* .

For this process to work well, prices must be able to change quickly. If they can, the economy quickly returns to Y^* . If, however, price adjustments are slow, the economy may produce less output than Y^* for a significant amount of time. In other words, if prices don't adjust quickly, you can get a recession. And until prices do adjust, the recession lingers.

Responding to Economic Shocks: Short-Run and Long-Run Effects

Economists like to break the time period after an economic shock into two parts, which they call the *short run* and the *long run*:

- » **The short run:** The period of time in which firms haven't yet made price changes in response to an economic shock
- » **The long run:** The period of time after which firms have made all necessary price changes in response to an economic shock

These definitions are intentionally vague because the speed at which firms adjust prices varies from shock to shock. In this section, I show you the major differences between what happens in the short run and the long run.

Defining some critical terms

To see the difference between an economy responding to a shock in the short run versus the long run, begin by looking at Figure 16–2, which is a model of the macroeconomy. The horizontal axis measures the dollar value of the output of goods

and services sold in the economy (Y). This number is the same as a country's gross domestic product (GDP), which I discuss in Chapter 14. The vertical axis measures the overall price level in the economy, P .

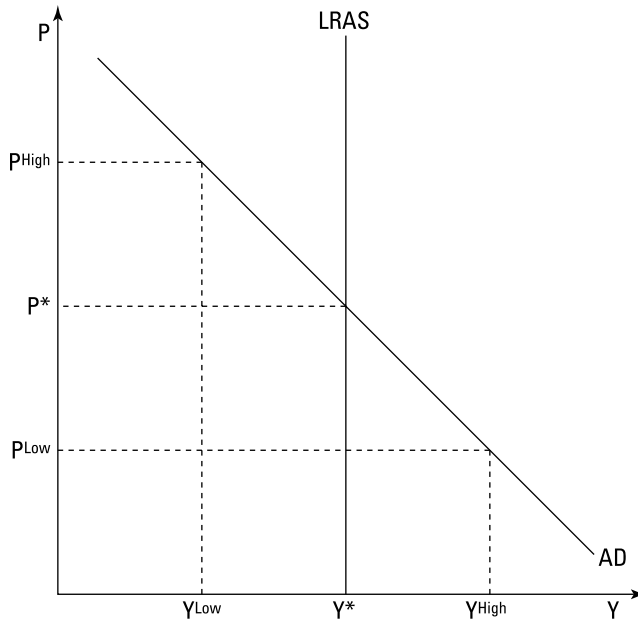


FIGURE 16-2:
A model of a
macroeconomy.

© John Wiley & Sons, Inc.

To understand the meaning of P , consider this: Although each good and service has its own price, and some of those prices may be going up while others are going down, an overall trend in prices exists for the economy as a whole. P is simply a measure of how the prices of goods and services behave as a whole. If P goes up, then on average prices are rising; if P goes down, then on average prices are falling. And if prices stay the same, then P (of course) stays the same. (See Chapter 15 for details about how economists measure P .)

Now that you understand the meaning of P , you're ready to tackle the two curves that appear in Figure 16-2:

» **Aggregate demand curve:** This curve represents the total amount of goods and services (Y) that people want to buy at any particular price level P , holding all other factors constant.

Notice that in Figure 16-2, the AD curve slopes downward. That's because there's an inverse relationship between the price level (P) and the amount of stuff that people want to buy (Y). *Inverse relationship* simply means that at the

higher price level (P^{High}), people want to buy a low level of output (Y^{Low}). But if prices fall to P^{Low} , people demand a much greater amount of output (Y^{High}). The downward slope of the AD curve captures the fact that at lower price levels, people want to buy more output.

» **Long-run aggregate supply curve:** This curve represents the amount of goods and services that an economy will produce when prices have adjusted after an economic shock.

In Figure 16-2, you can see that the $LRAS$ is a vertical line — it isn't a curve at all! (Do you feel cheated?) The $LRAS$ is drawn above the point on the horizontal axis that represents the full-employment output level, Y^* . Why? Because in the long run, changes in prices *always* return the economy to producing at the full-employment output level.

Contemplating the Tao of P: Looking at price adjustments in the long run

In Figure 16-2, P^* represents the *equilibrium* price level in the economy. P^* is determined by the intersection of the *long-run aggregate supply curve* ($LRAS$) with the *aggregate demand curve* (AD). P^* is both the price level at which consumers demand exactly the amount of full-employment output (Y^*) as well as the price level to which the economy will naturally return if the price level ever deviates from P^* .

To see why the price level will always move back to P^* , look at price level P^{High} and its corresponding aggregate demand level, Y^{Low} . Obviously, Y^{Low} is less than the economy's full-employment level of output (Y^*). That's important because firms would rather produce at output level Y^* . In fact, they've invested in factories and equipment that'll be wasted if they produce at lower levels of output. Consequently, their response is to cut prices to increase sales. And they continue to cut prices until the overall price level in the economy falls down to P^* because that's the price level at which consumers want to buy exactly Y^* worth of output.

Are you worried that all these price cuts will cause firms to lose money? Take heart: Firms don't necessarily lose profits in this situation because their costs are falling at the same time. That's because when the economy is producing at less than Y^* , there are a lot of unemployed workers and a lot of unused productive inputs, such as iron and oil. Unemployment puts downward pressure on wages; in other words, having lots of unemployed labor readily available means you can hire people at lower wages. And the larger the piles of unused productive inputs, the more their prices will fall too.

Okay, so the lower prices attract more customers, increase sales, and cause firms to hire back unemployed workers. This process continues until prices fall all the way to P^* , at which point the economy is operating at full employment again, meaning all workers who want full-time jobs can get them.

In a similar fashion, prices can't remain below P^* for long. At price level P^{Low} , people want to buy Y^{High} worth of output. But that's more than firms can produce at full employment. The only way to produce that much output is if employees work more than the standard 40-hour work week. The only way to get them to do so is to pay them more, and the only way to give them higher wages is for firms to raise prices. So with demand exceeding supply, prices are raised until they reach P^* , at which price level the quantity demanded by consumers is exactly equal to the full-employment output level, Y^* .

As you can see, if prices have enough time to adjust, the economy always returns to producing at output level Y^* . Because I'm calling the time required for prices to adjust the *long run*, it makes sense to call the vertical line above Y^* the *long-run aggregate supply curve* because it shows how much output the economy will supply after prices have had enough time to adjust to equalize the supply and demand for goods and services. (For much more on supply and demand, see Chapter 4.)

A shock to the system: Adjusting to a shift in aggregate demand

What causes prices to be too high or too low in the first place? A shock to *aggregate demand* — the total amount of goods and services that people are willing to buy.

First, visualize what a shock to aggregate demand looks like: Figure 16-3 shows the aggregate demand curve shifting to the left from AD_0 to AD_1 . A leftward shift of aggregate demand is called a *negative demand shock*, and it could be caused, for instance, by a decline in confidence in the economy that makes people want to save more and consume less. (A rightward shift of AD would be called a *positive demand shock*.)

The original price level, P_0 , is determined by where the original AD_0 curve intersects the vertical $LRAS$ curve. In the long run, after firms adjust to the demand shock, the new price level, P_1 , will be where the new AD_1 curve intersects the vertical $LRAS$ curve.



REMEMBER

The new price level (P_1) is less than the original price level (P_0). Why? Because the demand for goods and services decreases as a result of the negative demand shock. The only way to entice consumers to again purchase full-employment levels of output (Y^*) is to lower the cost of buying that much output, so the price level has to fall. Firms may take a while to make the necessary price reductions, but when they do, the economy will again produce at Y^* in the long run.

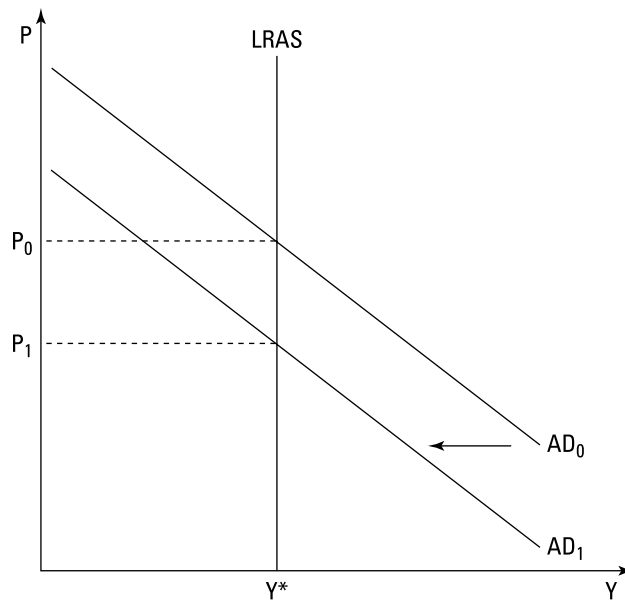


FIGURE 16-3:
A negative shock
to aggregate
demand.

© John Wiley & Sons, Inc.

Dealing with fixed prices in the short run

After an economic shock happens, prices eventually adjust to return the economy to full-employment output, Y^* . (See the preceding sections.) However, this process may take a while because in the short run, prices are essentially fixed. Even the managers of the most nimble firms need some time to decide how much to cut prices. And some firms aren't quite as nimble.

Suppose that a firm has printed up catalogs listing the prices of the things it sells. This firm distributes catalogs only once a year, which means it's committed to selling to customers at these prices until the next catalog is sent out. In such a situation, a firm adjusts its production to meet whatever amount of demand happens to come along at these fixed prices. If a lot of people show up to buy at these prices, the firm increases production, typically by hiring more employees. If very few people show up to buy, the firm reduces production, typically by hiring fewer employees.

Figure 16-4 depicts a situation in which firms have committed to a fixed set of prices and can respond to changes in demand only by adjusting their production levels. The figure shows the horizontal *short-run aggregate supply curve* (SRAS), which is actually a straight line. This “curve” corresponds to price level P_0 because the firms, in the short run, can't adjust their prices. Movements right and left along the SRAS curve capture the increases and decreases in output that firms have to make as the demand for their products varies at the fixed price level.

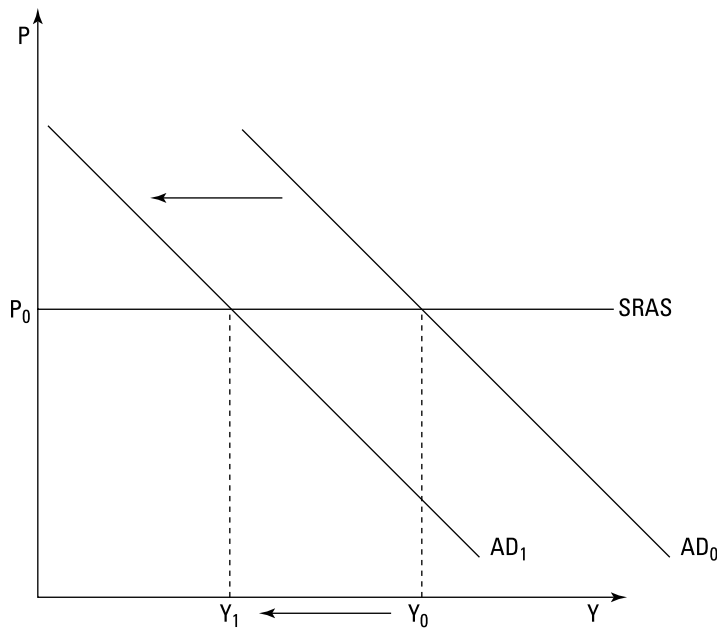


FIGURE 16-4:
The short-run
aggregate supply
curve (SRAS).

© John Wiley & Sons, Inc.



Economists have various complicated ways of explaining how an economy adjusts to a demand shock. I'm cutting the discussion down to the bare bones, which means thinking of the SRAS curve as a horizontal line. In more elaborate explanations, the SRAS curve is an upward-sloping line. But don't be confused: The basic idea in either case is that the economy moves from having a perfectly horizontal aggregate supply curve right after a shock, to an upward sloping one a little later, to a perfectly vertical one — the LRAS — in the long run. I'm skipping the middle part to make the discussion as direct as possible. I use only the initial, horizontal curve and the final, vertical curve, calling the former the SRAS and the latter the LRAS.

Figure 16-4 also has two aggregate demand curves, AD_0 and AD_1 , that again show what happens when aggregate demand is reduced as the result of a negative demand shock. The initial level of output that firms produce, Y_0 , is determined by the intersection of the original aggregate demand curve, AD_0 , with the SRAS curve. In other words, at price level P_0 , people demand output level Y_0 , and firms respond by supplying it.

When the negative demand shock strikes, it shifts aggregate demand leftward to AD_1 . Reduced demand means that at the fixed price level, customers are willing to buy less output. Because firms can't change prices, their only recourse is to reduce production to match the decrease in demand; this reduced level of output (Y_1) appears on the graph where the SRAS curve intersects AD_1 . Because lower output

means that firms need fewer workers, you end up with a recession. Output falls, and unemployment rises.

If you compare Figures 16–3 and 16–4, you can see that the leftward shift in aggregate demand has very different effects in the short run and the long run:

- » **In the short run, when prices are fixed:** Output falls and unemployment rises.
- » **In the long run, when prices can fall:** Output returns to the full-employment level and the unemployment rate returns to the low level associated with an economy in which the only kind of unemployment is frictional unemployment (due to people being voluntarily unemployed for a while as they take time off to search for better opportunities).

Why the huge difference between the short run and the long run? Firms aren't forever stuck with their original catalog prices. Eventually, they print new catalogs with lower prices. The lower prices entice customers to purchase more, and soon the economy can return to producing at the full-employment output level, Y^* .

Putting together the long and short of it

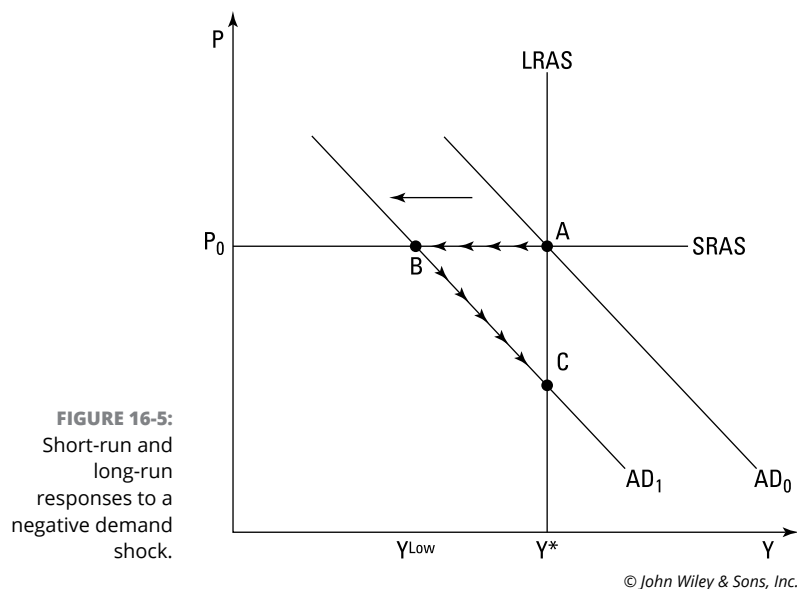
If you have both long-run and short-run responses to an economic shock tucked under your belt, you're ready to put the two very different responses into one big picture. Figure 16–5 lets you see how an economy adapts to a negative demand shock both in the short run and in the long run. The economy begins at point A, where the original aggregate demand curve, AD_0 , intersects both the $LRAS$ and the $SRAS$ curves. At Point A, the economy is in equilibrium because at price level P_0 , the aggregate demand for output equals the full-employment level of output, Y^* . There is neither a surplus nor a shortage that could cause prices to change.



REMEMBER

The $SRAS$ curve is horizontal at price P_0 to reflect the fact that after the economy reaches its equilibrium (where AD_0 intersects the $LRAS$ at output level Y^*), the prices that are determined at that level are fixed in the short run; they can't change immediately, even if a demand shock happens to come along.

For instance, suppose that the aggregate demand curve shifts left from AD_0 to AD_1 because of a negative demand shock of some sort. Because prices are fixed in the short run at P_0 , the economy's first response is to move from Point A to Point B. In other words, because prices are fixed, production falls from Y^* down to Y^{Low} as firms respond to decreased demand by cutting production. (Small arrows indicate the movement of the economy from Point A to Point B.)



At Point B, the economy is operating below full employment, implying that a lot of workers are unemployed. This high level of unemployment causes wages to fall. As wages fall, firms' costs also fall, allowing them to cut prices to attract more customers.

Falling prices cause increased aggregate demand for goods and services, which eventually moves the economy all the way from Point B to Point C. (Arrows on the graph indicate this movement.) When the economy reaches point C, it's once again producing at full employment, Y^* .



REMEMBER

The short-run and long-run effects of a negative demand shock are basically total opposites of each other:

- » **Short run:** In the short run, prices are fixed while output decreases.
- » **Long run:** In the long run, prices decrease while output returns to Y^* .

If prices don't stay fixed for very long, the economy can quickly move from A to B to C. But if prices are slow to adjust to the negative aggregate demand shock, the economy can take a long time to get from A to B to C. In such cases, there's a long-lasting recession, during which output remains below Y^* and many people are unemployed.

Heading toward Recession: Getting Stuck with Sticky Prices

When the economy encounters a negative demand shock like the one depicted in Figure 16–5, price flexibility (or lack of flexibility) determines both the severity and the length of any recession that may result. If prices were infinitely flexible — if they could change within seconds or minutes after a shock — the economy would immediately move from Point A to Point C, and all would be right with the world. But if prices are fixed for any period of time, the economy goes into a recession as it moves from Point A to Point B before prices eventually fall and bring it back to full-employment output at Point C.

In the real world, prices are indeed somewhat slow to change, or as economists like to say, prices are *sticky*. Interestingly, prices tend to be stickier when going downward than upward, meaning that prices appear to have a harder time falling than rising.

The major culprit seems to be one particular price: wages. *Wages* are the price employers must pay workers for their labor. Unlike other prices in the economy, people are emotionally attached to wages and how they change over time.



REMEMBER

Employees don't like to see their wages cut. They have a strong sense of fairness concerning their wages and usually retaliate against any wage cut by working less hard. As a result, managers typically find lowering wages to be counterproductive, even if a firm is losing money and needs to cut costs.

This section explains how firms' worries about worker motivation lead to the sticky output prices that prevent the economy from rapidly recovering from recessions. When sales fall as a result of the recession, output prices can't fall very much because firms choose to lay off workers rather than cut wages.

Cutting wages or cutting workers

During a recession, you see a large increase in unemployment but little decrease in wage rates. The fact that managers are unwilling to cut wages, however, has a nasty side effect: Not cutting wages makes it hard for firms to cut the prices of the goods and services they sell. (For more on this topic, see the next section.)

Suppose that a negative demand shock hits an economy and greatly reduces sales at a particular company. The firm is losing money, so managers need to figure out a way to cut costs. About 70 percent of this company's total costs are labor costs (wages and salaries). Naturally, labor costs are an obvious target for cuts.

But the managers of the firm realize that if they cut wages, employees will get angry and work less hard. In fact, their productivity may fall off so much that cutting wages may make the firm's profit situation worse. Output may fall so much that sales revenues will decrease by more than the reduction in labor costs. Therefore, cutting wages isn't really a good option.

So instead, the managers lay off a large chunk of their workforce to reduce labor costs. For instance, if sales are down 40 percent, the firm may lay off 40 percent of the workforce. However, any workers who remain employed get to keep their old wages so that they aren't angry and their productivity doesn't fall.

Adding up the costs of wages and profits

Obviously, firms need to turn a profit to stay in business. And that means making sure that the price per unit that they charge for their products exceeds the cost per unit of making them.

During a recession, lower aggregate demand means that firms reduce production and sell fewer units. Wages are the largest component of most firms' costs — in fact, they're a full 70 percent of the average firm's costs. If a firm can't cut wages for fear of causing worker productivity to drop, it can't reduce its per-unit production costs very much either. In turn, the firm can't cut its prices very much because prices have to stay above production costs if firms are to break even and stay in business. What does all this mean?



REMEMBER

When demand drops off, prices are typically sticky. They stay high even though there's less demand for output in the economy. That's the reason behind the economy's moving horizontally from Point A to Point B in Figure 16-5 after the negative demand shock. With prices sticky because firms can't quickly or easily cut wages, the negative demand shock results in a recession, with output falling and unemployment rising because so many workers get fired.

Worse yet, unless prices can somehow begin to fall, the economy won't be able to move from B to C to get back to producing at the full-employment output level (Y^*). Prices do *eventually* fall, but this process can take a long time, meaning that the negative demand shock can cause a long-lasting recession.

One way around this slow adjustment process is for the government to try to offset the negative demand shock. The next section covers how such attempts may be able to speed recovery by avoiding the need for prices to adjust to bring the economy back to producing at the full-employment output level.

Returning to Y^* with and without government intervention

In Chapter 17, I explain in great detail how the government can use monetary and fiscal stimuli to get around the sticky prices problem by boosting aggregate demand. Here, I want to give you a preview of how that process works.

Imagine that after the negative demand shock depicted in Figure 16–5 moves aggregate demand leftward from AD_o to AD_1 , the government doesn't hang around waiting for prices to fall. Instead, it stimulates aggregate demand so that the aggregate demand curve shifts right and returns to where it started, at AD_o . Taking this action returns the economy to producing at full employment without having to wait for prices to fall.

What if the government doesn't act to stimulate aggregate demand in that fashion? What if the economy is at Point B and the government doesn't intervene? In such cases, prices do eventually fall because firms' production costs eventually fall.

Labor costs are slow to fall because managers don't want to risk alienating workers by cutting their wages. (See the earlier section "Cutting wages or cutting workers" for details.) But because so many workers are unemployed when the economy is at Point B, wages eventually decline. Some firms hire unemployed people at lower wages, which reduces the firms' costs, meaning that they can undersell firms that keep wages high. Eventually, such competitive pressures mean that all firms end up cutting wages.

Other costs also decline because, during a recession, with output so diminished, a significant portion of the economy's productive capacity is unused. There are unused factories, unused trucks, unused train cars, and unused ships. There are also large amounts of unused lumber, iron, oil, and other productive inputs.

The owners of these unused inputs lower their prices to try to sell them. As their prices fall, so do production costs, thereby allowing firms to reduce the selling prices of their output. And as these selling prices fall, the economy moves from Point B to Point C in Figure 16–5, restoring the economy to producing at the full-employment output level (Y^*). See how nicely it all (eventually) works out?

Achieving Equilibrium with Sticky Prices: The Keynesian Model

John Maynard Keynes was the most influential economist of the 20th century. Why? He was the first economist to realize that sticky prices (caused by sticky wages) are the culprit behind recessions, an insight that changed the way people studied economies.

What inspired Keynes to have this insight? He was led to the idea by the horrible state that the economy reached during the Great Depression of the 1930s. Just the name itself — *Great Depression* — gives you some idea how bad things got. Normal economic downturns are called *recessions*. Really bad recessions are called *depressions*. What happened in the 1930s was so bad that people started calling it the *Great Depression* to indicate how severe it was.

The Great Depression started with a lingering recession from 1929 to 1933. The United States didn't see its output return to its 1929 level until after entering World War II in 1941. To put the Great Depression in perspective, Table 16-1 gives data for the eight recessions the United States has experienced since 1960, plus (on the first line) the same data for the Great Depression.

TABLE 16-1 **The Great Depression and U.S. Recessions between 1960 and 2022**

Start (Month/Year)	End (Month/Year)	Duration (Months)	Highest Unemployment Rate	Change in Real GDP (%)
8/1929	3/1933	43	24.9	-28.8
4/1960	2/1961	10	6.7	-1.1
12/1969	11/1970	11	5.9	-0.2
11/1973	3/1975	16	8.5	-3.2
1/1980	7/1980	6	7.6	-2.2
6/1981	11/1982	16	9.7	-2.9
6/1990	3/1991	8	7.5	-1.4
3/2001	11/2001	8	6.0	-0.4
12/2007	6/2009	18	9.9	-4.3
3/2020	4/2020	2	14.7	-10.1

Source: NBER, *Economic Report of the President*, Bureau of Labor Statistics

As you can see, the Great Depression was far, far worse than any normal recession. Nearly 25 percent of the labor force was unemployed, and the initial downturn lasted about four times longer than the 11.6-month average duration of post-1960 recessions.

Total economic output as measured by real GDP (which I discuss in Chapter 15) also fell much more than in a normal recession. Because real GDP adjusts for inflation, it captures changes in the physical quantity of output produced. In recent recessions, real output has fallen at most by 10.1 percentage points, with less than 3 percentage points being typical. During the Great Depression, real GDP fell 28.8 percent and the U.S. economy didn't fully recoup that decline in output until 1942. By contrast, the U.S. economy took less than 18 months to regain the 10.1 percent decline in output that it experienced during the very brief COVID-19 recession of 2020.

As a witness to the Great Depression, Keynes obviously wanted to figure out what could cause such a drastic economic downturn — and what could prevent such devastation from happening again. Keep reading for more on the Keynesian model.

Adjusting inventories instead of prices

Not only did Keynes figure out that sticky prices cause recessions, but he also developed a hugely influential model whose core ideas dominate macroeconomics even up to the present time. This model is a small part of a larger approach to managing the macroeconomy that came to be called *Keynesianism* — an approach that favored large government interventions into the economy rather than the sort of *laissez-faire* policies of nonintervention preferred by other people. (For a discussion of the costs and benefits of having the government intervene in the economy, see Chapter 3.)

To be fair, I have to point out that Keynesianism has attracted a lot of critics and isn't the be-all-end-all of macroeconomics. But the part of it I present here isn't controversial. It explains how an economy adjusts to *equilibrium* — a place where aggregate supply matches aggregate demand — in the extreme short run after an economic shock, when prices can't change at all.

Look back at Figure 16-4 for a moment. The Keynesian model elaborates on exactly how an economy moves from producing at output level Y_o to producing at output level Y_1 when a shock to aggregate demand happens and prices are fixed at level P_o .



REMEMBER

Keynes's model focuses your attention on firms' inventories of goods that have been made but not yet sold. According to Keynes, changes in inventories guide firms to increase or decrease output during situations in which prices are sticky and can't serve as signals of what to do.

To see the novelty of Keynes's inventory idea, understand that if prices could change, then prices (not inventories) would guide firm decisions about how much to produce:

- » If prices were rising, a firm would know that its product was popular and that it should increase output.
- » If prices were falling, the firm would know that the product wasn't doing well and that it should probably cut output — and maybe get into another line of business!

In an economy with fixed prices, however, firms need some other way of deciding whether to increase or decrease production. Keynes realized that the guiding force would be changes in inventories, as you can see in the following subsections.

Keeping an eye on target inventory levels

Inventories are constantly turning over, with goods flowing both in and out. New production increases inventories, while new sales decrease inventories. The two factors interact to determine if inventories are rising, falling, or staying the same. For instance, if new production equals new sales, inventory levels stay constant. If new production exceeds new sales, inventories rise.



REMEMBER

The interaction of new production and new sales is important because each firm has a *target level* of inventories that it likes to keep on hand to meet situations in which sales temporarily run faster than the firm can produce output. The costs and benefits of having a bigger or smaller inventory on hand determine the target level.

Having less inventory than the target level is dangerous because the firm may not be able to keep up with sales spikes. Having more inventory than the target level is wasteful because there's no point in having stuff sitting around unsold, year after year. Each firm weighs these costs and benefits to come up with its own target inventory level.

Target inventory levels may vary from year to year, depending on whether firms are expecting strong or weak sales. If managers are expecting strong sales, they may increase inventories, whereas if they're expecting weak sales, they may decrease them.



REMEMBER

Keynes realized that aggregate demand shocks (which are, by definition, unexpected) would show up as unexpected changes in firm inventories:

- » Unexpectedly low aggregate demand means that sales slow so much that inventories increase to levels higher than firms had planned on.
- » Unexpectedly high aggregate demand means that sales increase so much that inventories decrease and reach levels lower than firms had planned on.

Increasing or decreasing output as inventories fluctuate

Unexpectedly large changes in inventories cause firms to change their output levels as follows:

- » If inventories rise above target levels, firms respond by cutting production. By reducing production rates to less than sales rates, inventories begin to fall down toward target levels.
- » If inventories fall below target levels, firms respond by raising production. By increasing production rates to more than sales rates, inventories begin to rise toward target levels.

The changes in output levels caused by changes in inventories are hugely important because they determine not only whether output (Y) is increasing or decreasing but also whether unemployment is rising or falling. For instance, if firms increase production because inventories have fallen below target levels, they need to hire more workers so that unemployment falls. If, on the other hand, firms decrease production because inventories rise above target levels, they need to lay off workers, so unemployment rises.

Adjusting inventories based on planned and actual expenditures

The Keynesian model differentiates between planned expenditures and actual expenditures as follows:

- » **Planned expenditures:** The amount of money that households, businesses, the government, and foreigners would like to spend on domestically produced goods and services.
- » **Actual expenditures:** What households, businesses, the government, and foreigners end up spending on domestically produced goods and services. The sum total of all these actual expenditures is equal to GDP.



REMEMBER

What happens when actual expenditures are different from planned expenditures? Inventories automatically change. For instance, if more money is spent on goods and services than planned, people are buying more output than is currently being produced. This situation is possible because firms sell (from their inventories) goods that were produced in previous periods. On the flip side, if people spend less money on goods and services than planned, firms' inventories rise because firms have to store all the output that they can't sell.

Keynes represented planned expenditures, PE , algebraically with the following equation:

$$PE = C + I^P + G + NX \quad (1)$$

What do all these letters mean? Here's the short version. (I discuss them in detail in Chapter 14.)

- » C stands for the amount of output that consumers want to consume.
- » I^P stands for the amount of output that firms plan to buy as investment goods, such as new factories and equipment, as well as any inventory changes that firms plan to make.

If, later on, firms have to increase or decrease inventories more than they planned, then actual investment, I , won't equal planned investment, I^P .
- » G stands for how much output the government wants to spend on things like building schools, paving roads, and ensuring an adequate supply of paper for paperwork.
- » NX stands for *net exports* — the value of a nation's exports minus the value of its imports. NX tells you the net demand that the foreign sector of the economy has for stuff that your country makes domestically.

For actual expenditures, Y , Keynes used the same equation that you use to calculate gross domestic product (which I discuss in Chapter 14):

$$Y = C + I + G + NX \quad (2)$$



TIP

Why can you use the GDP equation to calculate actual expenditures? Actual expenditures are equal to national income because every cent of expenditure made in the economy is income to somebody. Furthermore, actual expenditure is also equal to the dollar value of all goods and services produced in the economy because every bit of output that's produced is sold to someone. (Any output that a firm makes but can't sell to customers is counted as being "sold" by the firm to itself as it's placed into inventory. These inventory changes are known as *inventory investment* and are totaled up in GDP as part of the total investment, I .)

Having three ways of looking at Y is handy as you become familiar with the Keynesian model. Sometimes understanding the model is easier if you think of Y as being actual expenditures; at other times, understanding is easier if you think of Y as being national income; and at yet other times it's most helpful to think of Y as being the dollar value of output (that is, gross domestic product). I switch freely among these three definitions whenever doing so helps make understanding the model easier.



REMEMBER

The only difference between the right sides of Equations (1) and (2) is the investment variable, which is *planned investment* (I^P) in the first equation and *actual investment* (I) in the second. In other words, Y and PE differ only because of differences in investments caused by inventories unexpectedly increasing (decreasing) whenever sales are more (less) than planned.

Bringing some algebra into the mix

You knew it was coming: It's time to get algebraic so you can identify the Keynesian model's economic equilibrium using mathematical superpowers. (Now, where did I put those?)

First you need to define a *consumption function* — a way to calculate total consumption — that you can substitute into Equation (1). In Chapter 14, I present the following formula for calculating consumption:

$$C = C_o + c(1-t)Y \quad (3)$$

For now, what you really need to know about this formula is that higher income (Y) leads to higher consumption (C). If you substitute Equation (3) into Equation (1), you get the following:

$$PE = C_o + c(1-t)Y + I^P + G + NX \quad (4)$$

If you look carefully, you can see that this equation shows that the total planned expenditures on goods and services in the economy (PE) depends on the total income in the economy (Y). The higher the total income, the more money people are going to plan to spend.



TIP

A good way to simplify the planned-expenditures equation is to create a variable called A and to define it as follows:

$$A = C_o + I^P + G + NX$$

If you do that, Equation (4) looks a little more palatable:

$$PE = A + c(1-t)Y \quad (5)$$

The variable A stands for *autonomous expenditures*, by which economists mean the part of planned expenditures that doesn't depend on income (Y). The part of planned expenditures that does depend on income, $c(1-t)Y$, is known as *induced expenditures*.

To understand induced expenditures, realize that because t stands for the income tax rate, $(1-t)Y$ is what people have left over to spend after the government taxes them. And of that amount, the fraction c gets spent on consumption, so $c(1-t)Y$ tells you how much expenditure is “induced” by an income of size Y .

Figure 16-6 graphs Equation (5) and labels it the *planned expenditures line*.

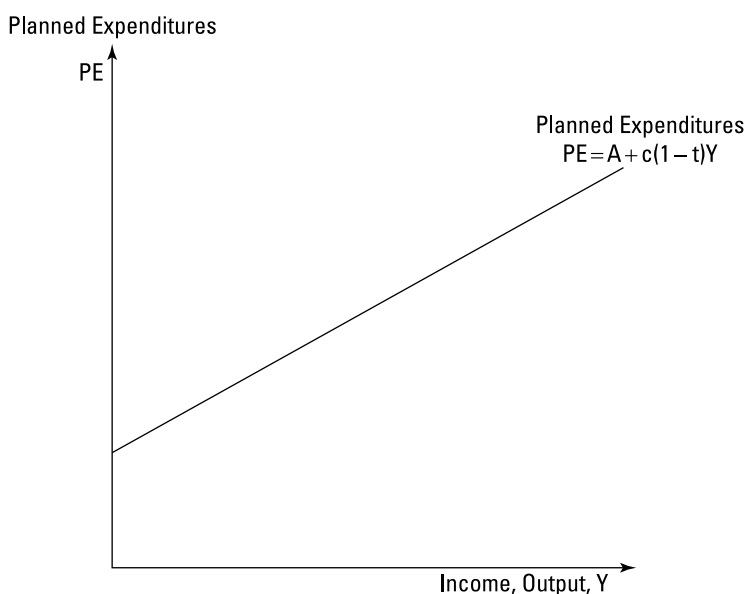


FIGURE 16-6:
The planned
expenditures line.

© John Wiley & Sons, Inc.

To find the specific equilibrium of the Keynesian model, realize that the following equation captures all possible equilibriums:

$$PE = Y \quad (6)$$

You can read this equation as “planned expenditures equal actual expenditures.” (Remember that Y equals both total income and total expenditure in the economy because all expenditures are income to somebody.)



REMEMBER

Any situation where $PE = Y$ is an equilibrium. Why? Because if the economy could get to the point where $PE = Y$, then nobody would have a reason to change their behavior. Consumers would be consuming as much as they planned to consume (C). The government would be buying up as much output as it wanted to buy (G). Foreigners would be buying as much stuff from your nation as they intended (NX). And, most importantly, firms would be spending exactly as much on investment as they planned — implying that inventories aren't changing unexpectedly.

If planned expenditures equal actual expenditures, you truly have an equilibrium because everybody is getting what they want, and no one has an incentive to change their behavior.



TECHNICAL
STUFF

You can solve for the equilibrium value of output, which I call $\tilde{Y}$, by substituting Equation (5) into Equation (6). If you do so, you get the following:

$$\tilde{Y} = \frac{1}{1-c(1-t)} A \quad (7)$$

Showing equilibrium graphically

If Equation (7) from the preceding section looks intimidating, stick with me. Finding the Keynesian model's equilibrium is much easier graphically. To generate the necessary figure, you plot the $PE = Y$ equation on the same graph as the $PE = A + c(1-t)Y$ equation, as I do in Figure 16-7. The point where the two lines cross is the equilibrium. At that point, planned expenditures are exactly equal to actual expenditures in the economy.

This equilibrium is *stable*, by which I mean that if the economy starts out at any income level other than $\tilde{Y}$, it soon moves back to $\tilde{Y}$. The element that returns the output to $\tilde{Y}$ is inventory changes.

To see why this is true, look at Figure 16-8, which exploits a nifty geometric trick about the $PE = Y$ line to show how the economy behaves when it's not producing at the equilibrium output level, $\tilde{Y}$.



TIP

The trick is that the $PE = Y$ line shows up on the graph at a 45° angle, meaning that you can use it to draw squares — boxes whose sides all have the same length. That means you can transpose any value of Y onto the vertical axis. To do so, take any value of Y , go straight up until you hit the 45° line, and then go straight to the left, sideways, until you hit the vertical axis. The point you hit represents as many dollars vertically as Y represents horizontally.

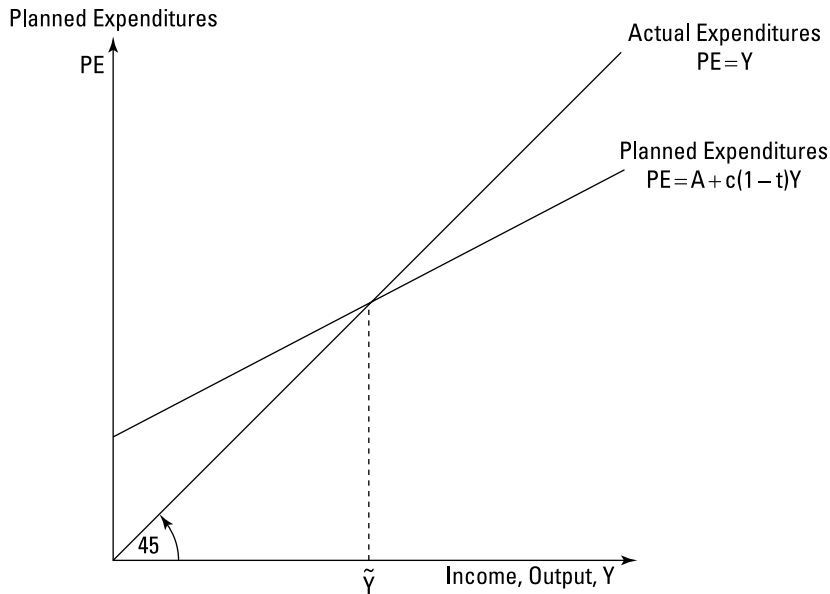


FIGURE 16-7:
The Keynesian
model's
equilibrium, $\tilde{Y}$.

© John Wiley & Sons, Inc.

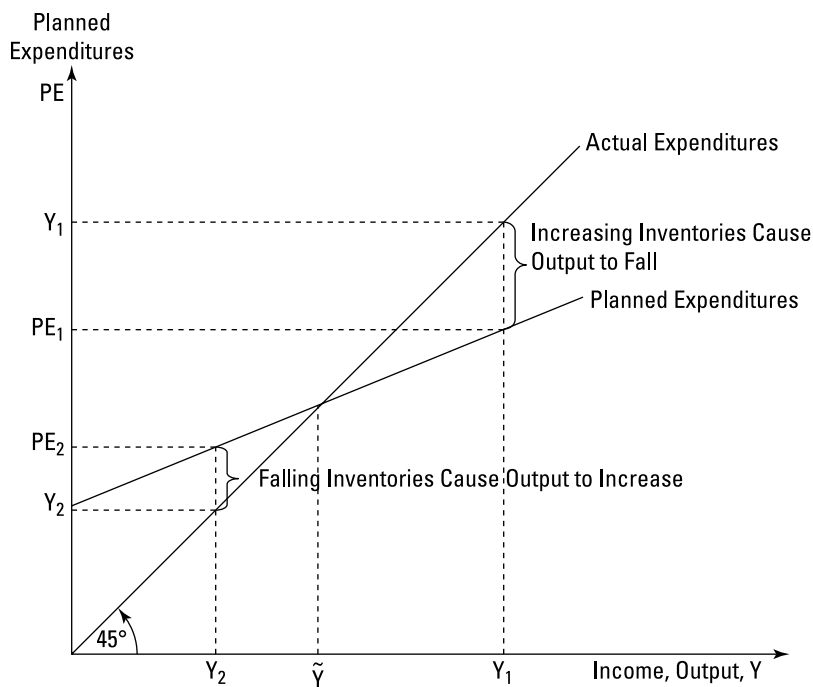
For instance, in Figure 16-8, start on the horizontal axis at output level Y_2 , which is less than the equilibrium output level $\tilde{Y}$. If you go up vertically to the 45° line and then to the left, you can plot output level Y_2 onto the vertical axis. Why is this useful? Because you can then compare Y_2 directly to the level of planned expenditures, PE_2 , that you get by starting at output level Y_2 on the horizontal axis.

As you can see, $PE_2 > Y_2$, meaning that planned expenditures exceed output in the economy. This means that inventories will unexpectedly drop as firms sell part of their stockpiles of inventory to make up for the fact that people are buying up more stuff than firms are currently producing.

This drop in inventories will return the economy to equilibrium. To see why, note that as inventories fall unexpectedly, firms will increase production. As a result, Y increases. Furthermore, actual expenditures continue to increase until Y reaches $\tilde{Y}$ because for any value of $Y < \tilde{Y}$, planned expenditures will continue to exceed output, as you can see from the graph.

Inventory adjustments also return the economy to equilibrium if it starts out at an output level like Y_1 , which is greater than $\tilde{Y}$. As you can see in Figure 16-8, by using the 45° line, actual output, Y_1 , exceeds planned expenditures, PE_1 . In other words, people are buying less (PE_1) than firms are currently producing (Y_1), so inventories will start to rise.

FIGURE 16-8:
How inventory
adjustments
always move
output back
toward $\tilde{Y}$.



© John Wiley & Sons, Inc.

Firms respond to increases in inventories by reducing output. They lay off workers and cut production. As a result, Y falls. It continues to fall until it reaches $\tilde{Y}$ because for any value of $Y > \tilde{Y}$, output will continue to exceed actual expenditures, as the graph shows.

Boosting GDP in the Keynesian model

Keynes didn't just invent his model to explain how economies with sticky prices reach a stable equilibrium. What he really wanted to do was show what governments could do during a recession to make things better.

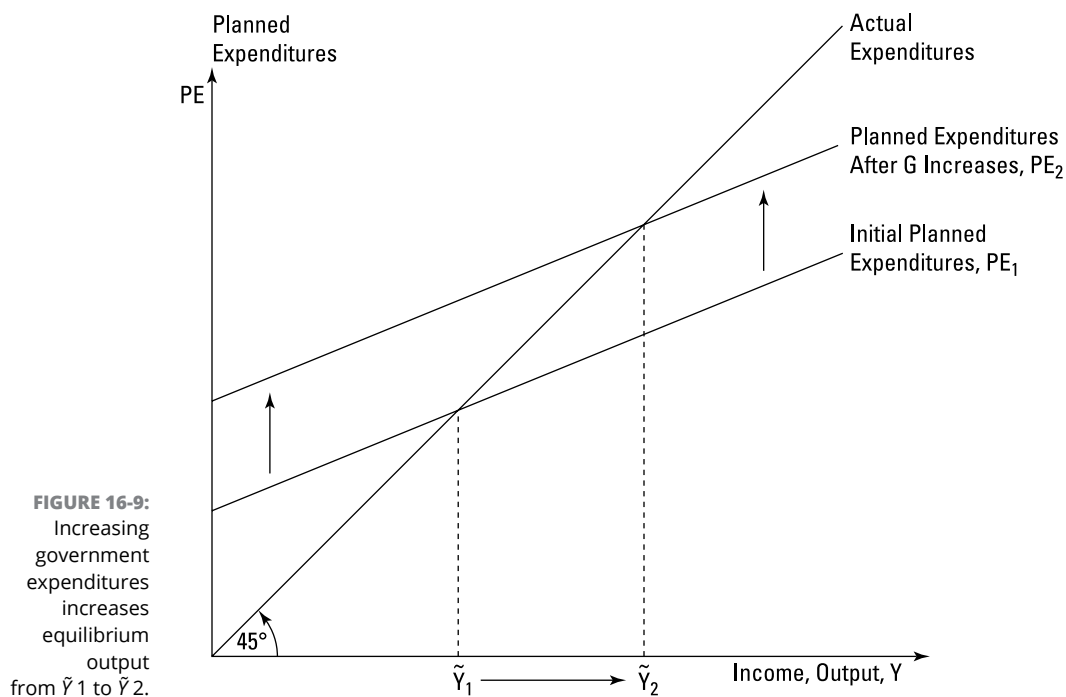
For instance, consider Figure 16-8. Suppose that inventory adjustments have carried the economy to equilibrium income, $\tilde{Y}$, but that $\tilde{Y}$ is less than the economy's full-employment output level, Y^* . In such a case, Keynes asked, what — if anything — governments should do.



REMEMBER

Governments could choose to do nothing. Eventually, because $\tilde{Y} < Y^*$, prices will fall and the economy will return to full employment (as it does moving from Point B to Point C in Figure 16-5). But Keynes argued that governments could speed up the recovery by boosting planned expenditures.

For instance, suppose that the government decides to increase G , government spending on goods and services. If it does so, then PE in Equation (4) clearly gets bigger. Because G is part of autonomous expenditures (A), the increase in G means an increase in A in Equation (5). Graphically, a larger A means that the planned expenditure line shifts vertically from PE_1 to PE_2 , as in Figure 16-9. Given that the actual expenditure line ($PE = Y$) doesn't change, the vertical shift in the planned expenditure line causes equilibrium output to increase from $\tilde{Y}_1$ to $\tilde{Y}_2$.



© John Wiley & Sons, Inc.

What Keynes suggested doing was using government policy to increase planned expenditures by whatever amount was necessary to increase the economy's short-run, sticky-price equilibrium, $\tilde{Y}$, all the way to the full-employment output level, Y^* . In Chapter 17, I discuss such policies in greater detail, including why they don't always work so well in practice.

IN THIS CHAPTER

- » Using monetary and fiscal policy to stimulate the economy
- » Facing the fact that too much stimulus causes inflation
- » Realizing that rational expectations can frustrate monetary and fiscal policy
- » Getting the details behind monetary and fiscal policy

Chapter **17**

Fighting Recessions with Monetary and Fiscal Policy

Monetary and fiscal policy are two of the most important functions of modern governments. *Monetary policy* focuses on increasing or decreasing the money supply to influence the economy, and *fiscal policy* uses government spending and the tax code to influence the economy.

Thanks to the development of good economic theory, governments can use monetary and fiscal policy to mitigate the duration and severity of recessions. This theory is hugely important because it gives governments the chance to make a positive difference in the lives of billions of people. Good economic policy can make a nation prosperous, but bad economic policy can ruin it.

That being said, monetary and fiscal policy aren't without problems. In this chapter, I show you how well they can work in the best-case scenario as well as their limits and problems in the real world, where best-case scenarios are often wishful

thinking. By seeing the whole picture, you can decide when and how governments should use monetary and fiscal policy.

The information in this chapter can put you two steps ahead of many politicians and help you judge when politically biased economists are trying to pull a fast one. As Joan Robinson, one of the great economists of the 20th century, said, “The purpose of studying economics is not to acquire a set of ready-made answers to economic questions but to learn how to avoid being deceived by an economist.” I totally agree. But don’t worry, you can trust *me*.

If you haven’t read Chapter 16, I encourage you to do so before tackling this chapter. Although my goal with this book is to make each chapter its own entity so that you can jump in and out wherever you need, I introduce and explain much of the terminology you encounter in this chapter in Chapter 16. You may find tackling monetary and fiscal policy easier if you have a basic understanding of how recessions work, which is the focus of that chapter.

Stimulating Demand to End Recessions

Before looking at monetary and fiscal policy in detail, realize that the purpose of both is to alter the aggregate demand for goods and services. (The *aggregate demand* is the total demand for goods and services in an economy.) In particular, both monetary and fiscal policy can increase aggregate demand during a recession. This section explains that strategy more in depth.

Aiming for full-employment output

The ability to use monetary and fiscal policy to stimulate the economy is important because you always want to end a recession and return the economy to producing at the full-employment output level as quickly as possible.

The full-employment output level, Y^* , is the amount of output the economy produces at full employment, which occurs when every person who wants a full-time job can get one. If the economy goes into recession and produces less than Y^* worth of output, millions of people lose their jobs because firms need fewer workers to produce the smaller amount of output.

Worse yet, the unemployment rate remains high until output returns to the full-employment level. Monetary policy and fiscal policy are useful precisely because they can help return the economy to producing at Y^* as soon as possible and shorten the frustration and misery the unemployed have to endure.

Look at Figure 17-1, which shows how monetary and fiscal policy can stimulate aggregate demand and return an economy to producing at Y^* as quickly as possible after the economy is hit with a negative demand shock. (*Negative demand shocks* [see Chapter 16] are events that unexpectedly decrease aggregate demand, such as a drop in consumer confidence.)

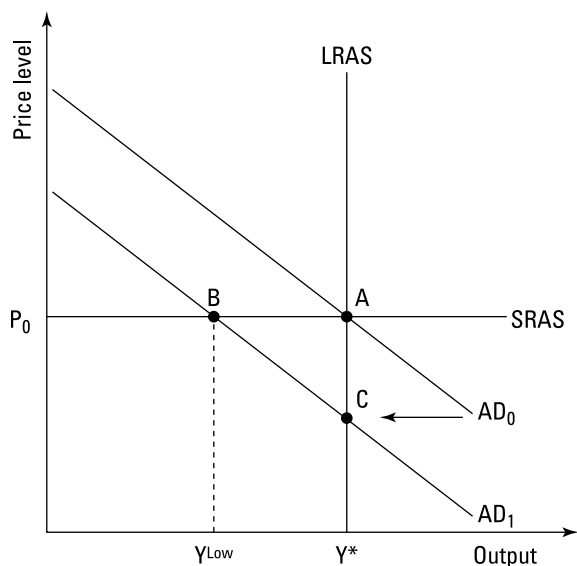


FIGURE 17-1:
How a negative
demand shock
affects the
economy.

© John Wiley & Sons, Inc.

In Figure 17-1, the economy begins in equilibrium at Point A, where the downward-sloping aggregate demand curve AD_0 intersects the vertical long-run aggregate supply curve, $LRAS$. Prices in the economy are fixed in the short run, so the short-run aggregate supply curve, $SRAS$, is horizontal at the initial price level (P_0), which is determined by the intersection of AD_0 and $LRAS$. (As in Chapter 16, for simplicity's sake I'm using horizontal $SRAS$ curves rather than the upward-sloping curves many other books use. If you're used to seeing upward-sloping $SRAS$ curves, take a peek at that chapter.)

When the negative demand shock comes along, here's what happens:

- » The aggregate demand curve shifts left to AD_1 , reflecting the reduction in spending on goods and services.
- » With prices fixed at P_0 in the short run, the economy's equilibrium shifts leftward from Point A to Point B, and output in the economy falls from Y^* down to Y^{Low} .
- » As output falls, unemployment rises because firms don't need as many workers.

As you can see, the overall result of the demand shock is a recession: a period of declining output and increasing unemployment. Unfortunately, a recession can take a long time to resolve. If the government takes no action to end a recession, the only way for the economy to return to producing at the full-employment output level is for prices to drop so that the economy's equilibrium can slide down the AD_1 curve from Point B to Point C. That process is typically slow because of sticky (slow-to-change) prices, especially sticky wages. As a result, the economy has high unemployment and takes a long time to get back to producing at Y^* unless the government becomes involved. Refer to Chapter 16 for more in-depth information.

Back to work: Shifting the AD curve right



REMEMBER

The trick that both monetary and fiscal policy accomplish is to increase aggregate demand, which eliminates the need to endure the slow adjustment process that takes the economy from Point B to Point C (see Figure 17-1). They do this by shifting the aggregate demand curve to the right.

For instance, if the government were able to shift the aggregate demand curve from AD_1 back to AD_0 , the economy would jump back to equilibrium Point A. That's nice because it gets the economy back to producing at Y^* without having to go through the slow adjustment process needed to get an economy to move from B to C. This means that unemployment ends much sooner for millions of workers, who can once again find jobs and provide for themselves and their families. Unfortunately, implementing aggregate demand shifts to fight recessions isn't easy. Several problems can creep up involving inflation and people's expectations about how increases in aggregate demand affect prices.

Generating Inflation: The Risk of Too Much Stimulation

The best way to begin to understand the limitations of economic policies that stimulate aggregate demand is to note that in the long run, such policies can change only the price level, not the level of output. Why? No matter where the aggregate demand curve happens to be — no matter how much stuff consumers are willing (or unwilling) to buy — prices eventually adjust until the economy is once again producing at full-employment output (Y^*). The economy simply doesn't want to stray from Y^* indefinitely.

I explain the economy's affection for Y^* in Chapter 16, and you can see it in Figure 17-1. The negative demand shock shifts the aggregate demand curve from AD_0 to AD_1 . If the government doesn't use some sort of stimulus, the economy

slowly adjusts on its own from Point A to Point B to Point C. At Point C, the price level has fallen, and output has returned to Y^* .

But even if the government applies some sort of stimulus to move the aggregate demand curve to the right of AD_0 , the long-run result is always that the economy comes to equilibrium at the point where the aggregate demand curve intersects the long-run aggregate supply (LRAS) curve. And as I show in Chapter 16, the LRAS is a vertical line that corresponds to the full-employment output level, Y^* .

In this section, I explain why changes in wages and prices prevent government stimulus from permanently pushing output above Y^* . In particular, I show that an economy that has been temporarily stimulated into producing more than Y^* is an economy in which workers are overworked and demanding wage increases — wage increases that will eventually drive up firm production costs, lower firm profits, and cause firms to reduce production. That feedback process continues until output falls back to Y^* .

Trying to increase output beyond Y^*

Because the economy always returns to producing at full-employment output (Y^*), the government can't for any significant period of time keep the economy producing more output than Y^* . To see why this is true, suppose that the government uses monetary or fiscal policy to shift the aggregate demand curve from AD_0 to AD_1 , as Figure 17-2 shows.

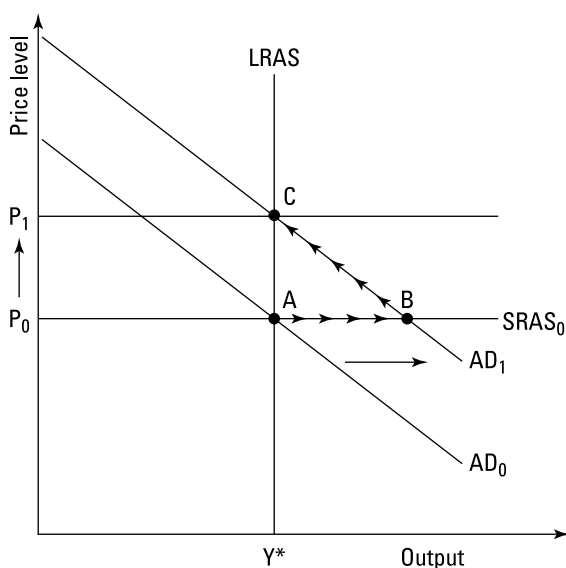


FIGURE 17-2:
The results of
increasing
aggregate
demand.

© John Wiley & Sons, Inc.

Before the shift, the economy is in equilibrium at Point A, where the original aggregate demand curve, AD_0 , intersects the long-run aggregate supply (LRAS) curve, which is a vertical line above Y^* . At that initial equilibrium, the price level is P_0 , and because prices are sticky in the short run (see Chapter 16), the short-run aggregate supply ($SRAS_0$) curve is a horizontal line at P_0 .

When the government stimulates the economy and shifts the aggregate demand curve to the right from AD_0 to AD_1 , the economy initially shifts from Point A to Point B. That is, because prices are fixed in the short run, the economy adjusts to a temporary equilibrium at B (where AD_1 intersects $SRAS_0$).

The economy's output level at Point B is greater than the full-employment output level, Y^* . Point B is only a temporary equilibrium because the only way the economy can produce more than Y^* is by using more labor than is used at Y^* . This can happen in only two ways, both of which increase wages:

- » **Firms convince workers to work overtime.** To get existing workers to consistently work overtime, firms must pay them high overtime wages.
- » **Firms increase the total number of workers by tempting people like retirees, not normally in the labor force, to take jobs.** To tempt people like retirees to join the workforce, firms must increase wages (because, obviously, these people weren't tempted to work at the old wages).

Either way, production costs rise. And as they do, firms pass them on to consumers by raising the prices they charge for goods and services. That's why the economy moves from Point B to Point C in Figure 17-2. As prices rise because wages are increasing, the economy moves up the AD_1 curve (as the arrows indicate). Wages, and hence prices, continue to rise until the economy is once again producing Y^* worth of output at Point C. At that point, there's no need for further wage increases; the economy is once again producing at Y^* , and firms don't have to increase wages to try to produce more than that level.

Tracing the movement of real wages

In the movement from A to B to C in Figure 17-2 caused by the government's stimulus, you can see that the only long-run consequence is an increase in the price level from P_0 to P_1 . After a period of increased production, the economy is back to producing at the full-employment output level, Y^* .

You can take two critical lessons away from this example:

- » The government can't permanently keep output above Y^* .
- » The government can't permanently keep more people employed than the number employed at Y^* .

These two lessons are true because of *real wages* — measured not in terms of money but by how much stuff workers can buy with the money they're paid.



REMEMBER

Real wages are crucial to understanding how government stimulus affects the economy because people don't work hard for money in and of itself — they work hard for the things money can buy. This distinction is important because as the economy reacts to the government's shifting of the aggregate demand curve from AD_o to AD_1 , real wages increase only temporarily. While they're higher, workers supply more labor. But when wages fall back down to their original levels, workers go back to supplying their original amount of labor.

Raising nominal wages while prices are stuck

To see the importance of real wages, consider the situation of a banana-loving worker named Ralph. When the economy is at Point A in Figure 17-2, Ralph is paid \$10 per hour, and his favorite food, bananas, costs \$1 per pound. This implies that his *real wages* — his wages measured in terms of what they can buy — are 10 pounds of bananas per hour. At that real wage, Ralph is willing to work full-time.

When the government stimulates the economy and shifts the aggregate demand curve from AD_o to AD_1 , workers like Ralph initially benefit because real wages initially rise. That's because to produce more output than Y^* , firms have to raise *nominal wages* (wages measured in money) to get workers to produce more. Because prices are initially sticky at price level P_o , the increase in nominal wages means an increase in real wages.

In Ralph's case, suppose that the price of bananas remains at \$1 per pound because of sticky prices, but Ralph's nominal wage rises to \$12 per hour because the company he works for needs more labor. Ralph's real wage increases from 10 pounds of bananas per hour to 12 pounds of bananas per hour.

This increase in real wages motivates workers to supply all the extra labor that's required to produce higher levels of output. (In Figure 17-2, this is what's going on at Point B.) Because nominal wages have gone up but prices haven't, the resulting increase in real wages causes workers to supply more labor, which in turn allows firms to produce an output level greater than Y^* .

Moving back to Y^* and to original real wages

Unfortunately, as firms begin to pass on the costs of increased nominal wages as higher prices, real wages begin to fall. Suppose that because of higher labor costs, the price of bananas rises to \$1.10 per pound. At that price, Ralph's real wage falls from 12 pounds of bananas per hour down to 10.91 pounds of bananas per hour. (To get 10.91, divide Ralph's \$12-per-hour money wage by the \$1.10-per-pound price of bananas.)



REMEMBER

In Figure 17-2, the decrease in real wages happens as the economy moves along the aggregate demand curve from Point B to Point C. As prices rise, real wages fall. Prices continue to rise until real wages return to where they were at Point A before the government stimulated aggregate demand.

In Ralph's case, the price of bananas continues to rise until they cost \$1.20 per pound. Now his higher nominal wage of \$12 per hour once again buys him 10 pounds of bananas per hour; his real wage is back where it started.

This boomerang effect in the real wage makes sense. Because the economy returns to producing at Y^* , you only need to motivate workers to supply enough labor to produce Y^* , not anything extra. Workers like Ralph were willing to supply that much labor at Point A for a real wage of 10 pounds of bananas per hour. After the economy has moved to Point C, these workers will once again be willing to supply that amount of labor for the same real wage.

Not every worker is as fixated on bananas as Ralph is, but you get the idea: If both wages and prices rise by 20 percent, real wages remain unchanged, and consequently, the amount of labor that workers supply ends up unchanged.



REMEMBER

Because the amount of labor reflects real wages, government stimulus policies that shift aggregate demand from AD_0 to AD_1 (as in Figure 17-2) can't permanently increase the amount of labor firms employ. Nor can these policies permanently increase workers' real wages. These effects are at best temporary; they last only as long as the economy takes to adjust from A to B to C.

Failing to stimulate: What happens when a stimulus is expected

Unfortunately, if people know about a stimulus ahead of time, the economy may adjust directly from A to C and eliminate the ability of the aggregate demand shift to stimulate the economy even temporarily. In this section, I show you that prices may adjust so quickly that the stimulus may fail to increase output at all, even temporarily.

Respecting the importance of price stickiness

As Figure 17-2 shows, any increase in output after aggregate demand shifts rightward from AD_0 to AD_1 depends on prices being sticky in the short run. In other words, the economy moves from Point A to Point B along the horizontal short-run aggregate supply curve, $SRAS_0$, only if the price level is fixed at P_0 in the short run.

A lot of evidence shows that prices have a hard time falling during a recession. (See Chapter 16 for details.) In particular, firms don't like to cut wages and insult

their workers. The firms' managers know that if they cut wages, workers will become angry and refuse to work hard, and the resulting decline in productivity will make the firm's profit situation even worse.

As a result, the economy has a lot of *downward wage stickiness*, which means that nominal wages rarely decline. Downward wage stickiness leads to *downward price stickiness* because firms can't cut their prices below production costs if they want to break even and stay in business. (Keep in mind that labor costs are, for most businesses, the largest part of production costs. If firms can't cut wages, they can't cut the price of their output.)

Realizing that prices aren't very sticky upward

In the preceding section, I talk only about *downward* stickiness; I don't say anything about prices or wages having trouble rising. In fact, very little in the economy can cause *upward* wage stickiness or *upward* price stickiness.

HELPING EMPLOYMENT WITH A LITTLE INFLATION

Economists have thought a lot about the best way to use monetary policy. Many have concluded that it should always be just a little bit overstimulating so that there's always a modest 1 or 2 percent inflation rate. The idea is that modest inflation helps smooth out the labor market by giving firms a sneaky way to increase profits if they run into a temporary slowdown in sales.

Wages are typically sticky downward because if you cut workers' wages, they get mad and give less effort. The result is that when the demand for a firm's output slows and labor costs need to be cut to restore profitability, managers usually fire a portion of the workforce and keep the remaining workers at their old wages instead of keeping all the workers on the job at lower wages.

The pressure to make such layoffs is lower in the presence of inflation because inflation drives up the selling price of the firm's output. If managers keep nominal wages fixed while that's happening, profits improve and lessen the need to fire anyone.

But workers' real wages will fall because while nominal (money) wages are fixed, the cost of living keeps rising due to the inflation. And because inflation helps firm profits along, firms see less of a need to lay off workers. So although the workers lose in some sense by their real wages falling, many still have jobs, whereas otherwise they would've been laid off.

Wages and prices seem quite free to rise if demand increases relative to supply. Business and labor contracts may limit price and wage increases for a while, but as soon as these contracts expire, prices and wages are free to rise.

Anticipating (and undermining) a stimulus



REMEMBER

The lack of upward price stickiness implies two important ideas for any government attempting to stimulate the economy into producing more than the full-employment output level (Y^*):

- » **If prices and wages can rise quickly, the economy will produce more than Y^* only briefly.** That is, the economy will move from A to B to C in Figure 17-2 quickly — so quickly that the stimulus will cause output and employment to rise above Y^* only briefly.
- » **If people see a stimulus coming, that stimulus (which attempts to increase output beyond Y^*) is likely to generate only inflation and no increase in output whatsoever.** If people anticipate an increase in aggregate demand, the economy may jump directly from Point A to Point C so that the price level rises without even a temporary increase in output.

This phenomenon is an example of *rational expectations*, a term that economists use to describe how people rationally change their current behavior in anticipation of future events. In this case, firms rationally decide to raise prices immediately when they find out that the government will be increasing aggregate demand from AD_0 to AD_1 in the future.

Indeed, firms' only rational course of action is to raise prices immediately because if firms were to leave prices alone at P_0 , they'd be volunteering for the decrease in profits that results when the economy moves from Point A to Point B (when nominal wages rise while prices stay constant). By immediately raising prices and shifting the economy directly from A to C , firms can avoid that situation altogether.



REMEMBER

Rational expectations is one of the most important ideas in macroeconomics because it tells you that there are strong limits on the government's ability to control the economy. People don't just sit around like potted plants when the government announces a policy change; they change their behavior. And sometimes their behavioral changes completely ruin the government's ability to stimulate the economy — as is the case with firms' immediately raising prices when the firms' managers find out that the government is about to attempt a stimulus.

To see why this is true, suppose that the government preannounces a big stimulus package that will shift aggregate demand from AD_0 to AD_1 in a few months' time (see Figure 17-2). Because workers and businesses can learn macroeconomics just as well as the politicians running the government, workers and businesses realize that the only long-run effect of the upcoming stimulus will be for prices to rise from P_0 to P_1 .

In addition, workers understand that real wages will remain unchanged in the long run because both their nominal wages and their cost of living (given by the price level) will increase by equal amounts. As a result, workers know that in the long run, the stimulus won't help them at all. Indeed, their only hope for gains is based entirely upon the short run, when nominal wages should go up and the price level should stay the same. In other words, they hope to benefit from the movement from A to B in Figure 17-2.

But firms aren't stupid. They don't want to have their profits reduced because wages are rising while prices are fixed, so they simply anticipate everything. Because prices eventually have to rise from P_0 to P_1 and wages eventually have to rise by an equal amount, firms get ahead of the wage increases by raising prices as soon as they can.

Nothing prevents firms from raising prices because nothing in the economy causes upward price stickiness. So if firms can see the stimulus coming ahead of time, they simply raise prices as soon as they can to make sure that prices and wages are going up at the same pace. As a result, the price level jumps from P_0 to P_1 .

Of course, at the same time, firms raise wages by an equal percentage to keep real wages the same. They want to keep workers motivated to supply the labor necessary to produce Y^* worth of output.

As you can see, if a government tries to stimulate the economy past producing at Y^* , and if everyone in the economy understands and anticipates the stimulus, it may not work at all. Prices and wages may simply jump from Point A to Point C, meaning that the stimulus fails to stimulate because output stays constant at Y^* while prices and wages go up simultaneously.

I explain monetary and fiscal policy in more detail in the rest of the chapter, where you can see other examples of rational expectations limiting the effectiveness of government policy. Be sure to notice how, in every case, changes in people's behavior reduce the impact of government policy initiatives.

Figuring Out Fiscal Policy

Fiscal policy concerns itself with how governments tax and spend. It overlaps macroeconomics because modern governments have many opportunities to increase aggregate demand by making changes in fiscal policy. These changes fall into two main categories:

- » **Indirect:** Increasing aggregate demand indirectly by lowering taxes so that consumers have larger after-tax incomes to spend on buying more goods and services
- » **Direct:** Increasing aggregate demand directly by buying more goods and services

The first category involves decreasing government revenues, and the second involves increasing government spending. Because the government's budget deficit is defined as tax revenues minus spending, both types of fiscal policy are likely to increase government budget deficits. This idea is important because large and ongoing budget deficits may lead to many economic problems, including inflation. As a result, the fear of large budget deficits constrains the magnitude of fiscal policy initiatives.



As you read about fiscal policy, keep the fear of large budget deficits in mind because it limits the size of the aggregate demand shifts that a government can undertake. For instance, if you look back at Figure 17-1, the government may want to use fiscal policy to shift aggregate demand rightward from AD_0 to AD_1 , but if doing so would involve an overly large budget deficit, the government may have to settle for a smaller shift that moves the economy only part of the way back to producing again at full-employment output (Y^*).

Increasing government spending to help end recessions

If an economy gets into trouble, one of the first things that politicians call for is increased government spending. The idea is that if people are unemployed and unsold goods are sitting around gathering dust, the government can come in with a lot of money and buy up a lot of unsold products. The result is that the government generates so much demand that businesses start hiring the unemployed to increase output to meet all the new demand.

The hope is that this stimulus jump-starts further demand. When people who were unemployed start getting paychecks again, they start spending more money,

which means demand rises. When this happens, the economic recovery should be self-sustaining so the government doesn't need to continue to spend so much money.

Paying for increased government spending

Politicians like suggesting increases in government spending because such increases make politicians look good, especially if they can get some of the new spending earmarked for their constituents. However, nothing in life is free.



REMEMBER

The government can pay for increased spending in only three ways:

- » **The government can print more money.** Printing lots of new money to pay for increased government spending leads to large bouts of inflation, which bring with them economic chaos and recession. (See Chapter 16.) Consequently, governments nowadays rarely resort to printing money to pay for increased government purchases of goods and services.
- » **The government can raise taxes.** Raising taxes is problematic because if you're trying to get out of a recession, you want consumers to spend as much as possible on goods and services. If you raise taxes, consumers reduce their spending. You may offset some of the decreased private spending by immediately turning around and spending all the tax revenue, but clearly this isn't the way to stimulate aggregate demand in the long run. The government may as well just let its citizens spend their money in the first place.
- » **The government can borrow more money.** To combat recessions, governments have to figure out a way to increase their spending without decreasing private spending. The solution is borrowing.

Borrowing and spending: The most common solution

By borrowing and spending money during a recession, the government can increase its purchases of goods and services without decreasing the private sector's purchases. Who does it borrow from? You and other people like you.

At any given moment, people want to save a certain part of their incomes. These people can use these savings to buy many different kinds of assets, including stocks and bonds issued by corporations, real estate, mutual funds, and annuities. But people can also use their savings to buy government bonds, which are, in essence, loans to the government.



REMEMBER

By offering more bonds for sale, the government can redirect some of the savings that people are making away from purchases of other assets and into purchases of government-issued bonds. By selling bonds, the government can get hold of lots of money that it can spend on goods and services, thereby turning what otherwise would've been private spending on assets into public spending on goods and services.

Dealing with deficits

Increasing government spending and financing it through borrowing is clearly a good way to increase the overall demand for goods and services. But it has the potentially nasty side effect of creating a *budget deficit*, which is the dollar amount by which government spending exceeds tax revenues during the current year. Any current budget deficit adds to the *national debt*, the cumulative total of all the money that the government owes lenders.

The problem with budget deficits and the national debt is that they have to be paid back someday. Consider a ten-year bond that pays a 6 percent rate of return. When you buy the bond from the government, you give it \$1,000. In return, the government promises to do two things:

- » Give you back your \$1,000 in ten years
- » Give you \$60 a year (a 6 percent return) until you get your \$1,000 back

So the government gets \$1,000 now to spend on goods and services to boost the economy, but it has to figure out where to get \$60 per year for your interest payments and where to get \$1,000 in ten years when the bond matures.

In this section, I explain how tax revenues and the government's ability to print money ensure that it should always be able to pay back its bonds.

Relying on the security of future tax revenues

Obviously, the only reason people are willing to lend the government money by buying bonds is that the buyers believe that the government will eventually pay them back. They have confidence in repayment because governments have the exclusive right to tax things. Essentially, all government borrowing is secured by future tax revenues.

But the link between taxes and bond repayments isn't direct. Just because a government has a lot of bonds coming due, it doesn't necessarily have to raise taxes

suddenly to get the money to pay off the bonds. That's because governments often refinance the bonds that are coming due by issuing new bonds to get enough cash to pay off the old bonds. This process is referred to as *rolling over the debt* and is routinely practiced by governments everywhere.

But don't think that this is all just a huge scam to defer paying off the debt indefinitely. The only reason investors are willing to participate in a rollover is that they have confidence that the government can always use its future tax revenues to pay off its debts. Investor confidence allows governments to keep on borrowing, whether to fund new borrowing or to roll over old debt.

Paying the debt by printing money: A devastating choice

Sometimes investor confidence in the government turns out to have been misplaced. Governments have another (rather diabolical) way to pay off their bonds besides using tax revenues: They can print lots of money.

A \$1,000 bond obligates the government to pay you back \$1,000 worth of money. The bond doesn't say where that \$1,000 comes from. So the government is free to print \$1,000 worth of new bills and hand them to you. This solution may seem okay at first, but when you and all the other bond holders with newly printed cash go out into the economy and start spending that new money, you drive up prices and cause inflation.

Big inflations destroy economic activity, as Chapter 15 explains. Prices lose much of their meaning, and people are much more mistrustful and reluctant to engage in long-term contracts or make long-term investments because they don't know how much money will be worth in the future.



REMEMBER

Knowing the potential horrors of inflation, people tend to worry anytime they see a government running large budget deficits or piling up a large debt. They worry that the government may find itself in a position in which it can't raise taxes high enough to pay off its obligations (or it isn't willing to anger voters by raising taxes that high). Investors worry that if this situation occurs, the government may resort to printing money to pay off its debts — and doing so would ruin the economy.

Printing money to pay government debts would also badly hurt most bondholders because most of them would get their cash after prices have gone up, meaning their cash won't buy much stuff. Consequently, when people really begin to worry that a government may start printing money to pay off its debts, finding anyone willing to buy government bonds gets harder. In such a situation, the only way for

the government to get anyone to buy its bonds is to offer higher and higher interest rates as a compensation for people's worries that the money they'll eventually get back won't be worth much. These higher interest rates make the government's situation even more desperate because any debt rollovers have to be done at the higher interest rates.

Furthermore, because inflation affects all bonds, not just the ones the government issues, interest rates all across the economy rise if people fear inflation is coming. This situation can have bad economic consequences immediately because higher interest rates dissuade consumers from borrowing money to buy things like cars and houses, and the interest rates discourage firms from borrowing money to buy new factories and equipment. Consequently, just the expectation that a government may print money at some point in the future to pay off its bonds can cause immediate harm to the economy. (This is another example of rational expectations in action; see the section "Anticipating (and undermining) a stimulus," earlier in the chapter, for details.)

Most governments try to keep their debt level and their deficits under control so that no one seriously worries that the government will ever be tempted to print money to pay off its bonds.

Dissecting Monetary Policy

Monetary policy is the manipulation of the money supply and interest rates to stabilize or stimulate the economy. In modern economies, monetary policy is regarded as the most powerful mechanism that governments have at their disposal to fight recessions and reduce unemployment — even more powerful than fiscal policy.

Governments put monetary policy into practice by changing the supply of money to manipulate interest rates. Because interest rates affect everything from consumers' demand for home mortgages to businesses' demand for investment goods, interest rates have a huge and pervasive effect on stimulating or depressing economic activity.

To give you a complete picture of how monetary policy functions, I first explain what money is. I then show you that having too much money is possible (say what?) and how that idea is related to interest rates and inflation. That, in turn, gives you the insight necessary to understand how the government can affect interest rates by changing the amount of money floating around in the economy.

Identifying the benefits of fiat money over the gold standard

Money is an *asset*, meaning that it's a type of property that has value and retains value over time. Other assets include real estate, precious metals such as gold, and financial assets such as stocks and bonds. But money is unique because it's the only asset that's universally acceptable as a means of payment for goods and services.

Money makes an economy much more efficient because it eliminates the need to engage in barter. But the need to verify the authenticity of money (so people are willing to accept it) means the responsibility for producing money and suppressing counterfeits falls to governments. That, in turn, brings up potential problems because governments always face the temptation to print more money to pay off old debts or buy lots of newly produced goods and services.

Historically, one way to limit governments' ability to print more money to pay bills was to put money on a *metallic standard*. Under such a system, governments couldn't print more bills without backing them with a precious metal, such as gold. For instance, the United States used to have a gold standard under which people could redeem \$35 for 1 ounce of gold. You could actually bring \$35 of bills to the U.S. Treasury and exchange it for an ounce of gold.

What this meant for monetary policy was that the government couldn't arbitrarily increase the supply of paper money because for every \$35 of new bills it wanted to print, it had to buy an ounce of gold with which to back them. The high cost of buying gold limited the money supply. Such a system is great for preventing big inflations because the only way you ever get a big inflation is if the government prints a huge amount of new money. (When that new money begins circulating, it drives up prices.)



WARNING

Preventing inflations is a good thing, but using a metallic standard turns out to have some big drawbacks. That's because using a metallic standard causes the supply of money to be pretty much fixed over time, meaning that even if the economy could use a little bit more or a little bit less money to make it work better, the government can't do anything because the supply of money is fixed by the amount of gold the government has in its vaults.

In particular, the metallic standard means that you can't use monetary policy to stimulate your economy if it gets into a recession. One of the reasons the Great Depression was so bad around the world was that nearly every country was on a gold standard when the calamity began. This meant that governments were unable to increase their money supplies to help their economies. It also explains why the countries that quit their gold standards earliest had the shortest and mildest

recessions; after they quit, they were free to print new money to stimulate their economies. On the other hand, countries that stubbornly stuck to their gold standards, such as the United States and England, had the most prolonged and painful economic downturns.

Largely because of that experience and the desire to use monetary policy if needed, every country in the world has abandoned the gold standard in favor of *fiat money*. Under a fiat money system, the government simply prints up as many bills as it likes, declares them to be money, and puts them out in the economy. (*Fiat* means “let it be” in Latin.) The great benefit of this system is that the government can arbitrarily increase or decrease the money supply in whatever way will best help to regulate the economy.



I use M to denote the total supply of money floating around the economy. For instance, “ $M = \$1.3$ trillion” means that the sum of the face values of all the bills and coins in the economy is \$1.3 trillion.

Realizing you can have too much money

I know what you’re thinking: More money is always better than less money, and you can never have too much. But hear me out.

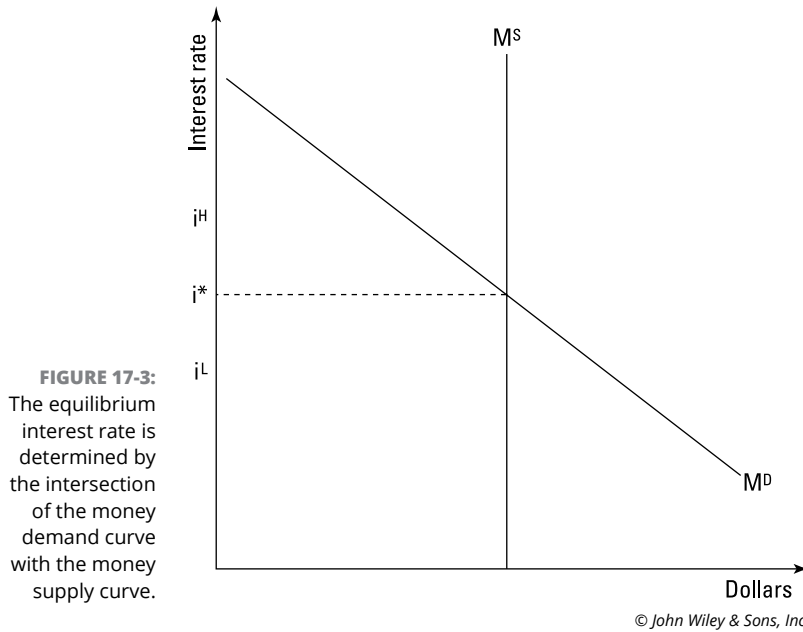
Monetary policy manipulates the supply of money to change the price of borrowing money, the interest rate. The key to making monetary policy work is that the demand for money depends on the interest rate.

Imagine that you have \$1 million and can do whatever you want with it. Suppose you’re frugal and decide to save every penny, at least for a year, because you think that’ll be enough time to figure out how to best blow the money. My question to you is this: Should you keep it all in cash? If you answered no, you’re right.

Holding wealth in cash isn’t smart because it earns no interest. Even if you put the money into a checking account at a bank, you’d get at least a tiny bit of interest. Even 1 percent of interest on a million dollars is \$10,000. Why would you give that up? Even better, if you use the cash to buy government bonds, you may get 5–6 percent in normal times. That’s \$50,000 or \$60,000 more than you’d get if you kept your wealth in cash.

Clearly, the higher the interest rate you can get on other assets, the more incentive you have to convert your cash into other assets. In fact, the only thing preventing people from converting all their wealth to other assets and never holding cash is the fact that money lets them buy things. Beyond that function, money isn’t better than any other asset; in fact, it’s worse in terms of its rate of return because the rate of return on cash is always zero.

Figure 17-3's graph demonstrates how much money people demand to hold at any particular interest rate. I denote money demand as M^D . The nominal interest rate, i , is on the vertical axis. (For an explanation of nominal interest rates, see Chapter 15.) The horizontal axis is measured in dollars.



As you can see from the downward slope of the money demand curve, the higher the interest rate, the less money people want to hold. This graph simply represents the idea that cash, with its zero interest rate, is a worse and worse place to park your wealth if you can get higher and higher returns in alternative assets. In other words, the higher the interest rate on other assets, the more you're going to want to economize on your cash holdings.

Figure 17-3 also shows the vertical money supply curve, where M^S stands for money supply. This curve is vertical because the government can decide how much money it wants to print and circulate without regard to the interest rate.

The M^D and M^S curves cross at interest rate i^* . This interest rate is the *equilibrium* interest rate because it's the only one at which the total number of dollars of money that people want to hold is equal to the total number of dollars that the government has circulated. More importantly, i^* is a *stable equilibrium*, meaning that if interest rates ever deviate from it, market forces will push them back to i^* .

Getting the basics about bonds

The best way to see why the equilibrium interest rate constitutes a stable equilibrium (see the preceding section) is to understand how interest rates are determined in the bond market. Pay close attention because bond markets are *the* place where interest rates for the whole economy are determined. Bond markets have a huge effect on everything else that goes on in the economy.

A *bond* is a financial asset for which you pay a certain amount of money right now in exchange for a series of payments in the future. Payments come in two flavors: face value payments and coupon payments:

- » **Face value payment:** The face value is printed on the face of the bond certificate, and this payment comes on the date the bond expires.
- » **Coupon payments:** These payments are typically made twice per year until the bond expires. They're called *coupon payments* because before computerized recordkeeping, you'd literally clip a coupon off the bottom of the bond certificate and mail it in to receive your payment.

Typically, bonds expire after 1, 5, 10, or 20 years.



REMEMBER

Bonds don't guarantee any sort of rate of return. They promise only to make the coupon and face value payments on time. The rate of return depends on how much you pay for the right to receive those payments.

If you think I'm speaking in tongues right now, bear with me. Imagine a really simple kind of bond called a *zero-coupon bond* (so named because there are no coupon payments). The only payment this bond will ever make is the face value payment that comes when the bond expires. And to make things really simple, suppose that this bond will pay its owner exactly \$100 one year from now.

If you're the bond owner, the rate of return the bond will pay depends on how much you pay for it right now. Suppose that you were naive enough to pay \$100 for the bond right now. Your rate of return would be zero percent because you paid \$100 for something that will give you \$100 in a year.

On the other hand, suppose you pay only \$90 for the bond right now. Your rate of return will be about 11 percent because $\frac{\$100 - \$90}{\$90} = 0.111$, or 11.1 percent. If you could buy the bond for only \$50, your rate of return would be 100 percent because you'd double your money in a year's time.



TIP

The rate of return on a bond *varies inversely* with how much you pay for it. Because the amount of money you get in the future is always fixed, the more you pay for it right now, the less is your rate of return. Higher bond prices imply lower rates of return.

Seeing the link between bond prices and interest rates

The fact that bond prices vary inversely with interest rates is the key to understanding why i^* is a stable equilibrium in Figure 17-3. First, consider interest rates that are higher than i^* , such as i^H . When interest rates are higher than i^* , the amount of money supplied exceeds the amount of money demanded. This means people have been given more of the asset called *money* than they want to hold. So what they do is try to reallocate their portfolio of assets by using the excess money to buy other assets.

Another asset is bonds. But with all this new money being thrown at the limited supply of bonds, the price of bonds rises. Now be careful. What happens to interest rates when bond prices rise? They *fall*. That's why if you start out at an interest rate that's higher than i^* , interest rates will fall back toward i^* . Excess money drives up the price of bonds, which lowers interest rates.

On the other hand, for interest rates like i^L that are lower than i^* , the amount of money demanded exceeds the amount supplied. Because people want more money than they have, they're going to try to get it by selling noncash assets such as bonds to convert those assets into the cash they want.

Imagine that everybody does this by trying to sell their bonds. With all the selling, bond prices fall, meaning interest rates will *rise*. In fact, bond prices will continue to fall and interest rates rise until they're back at i^* , because that's the only rate of interest at which people are satisfied holding the amount of money M^S that the government has decided to circulate.



REMEMBER

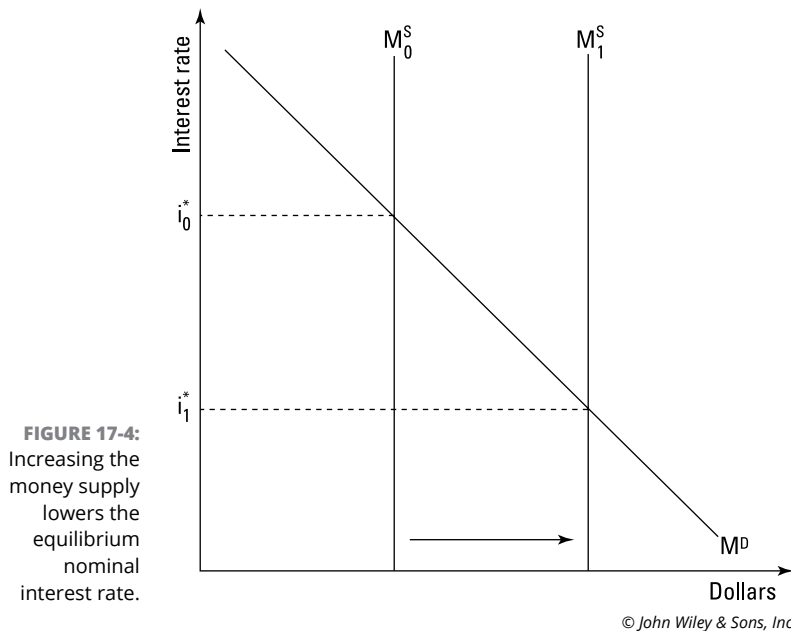
Movements back to the equilibrium interest rate, i^* , are quick. Any excess money demand or excess money supply never lasts very long because rapid adjustments in the price of bonds move the interest rate to its equilibrium.

An important consequence of the fact that interest rates adjust so quickly is that the government can print whatever amount of money it wants to, knowing that interest rates will adjust to get people to want to hold exactly that amount. This gives the government a useful policy tool to manage the economy because it can think one step ahead and create whatever interest rate it wants by printing the appropriate amount of money.

Changing the money supply to change interest rates

Monetary policy works because governments know that interest rates adjust to get people to hold whatever amount of money the government decides to print. The interest rate is, in some sense, the price of money, and it reacts in a way similar to other prices. That is, if the money supply suddenly increases, the price of money falls, and vice versa.

You can see this in Figure 17-4, in which the government increases the money supply from M_0^S to M_1^S . This action shifts the vertical money supply line to the right and lowers the equilibrium nominal interest rate from i_0^* to i_1^* .



In the United States, changes in the money supply are controlled by the Federal Reserve Bank, which is often just referred to as *the Federal Reserve* or *the Fed*. The Fed has the exclusive right to print currency in the United States, which means that it could make M^S as big as it wanted to by printing more money and handing it out. However, the Fed actually relies on a more subtle method for changing the money supply, a method that economists call *open-market operations*.



REMEMBER

Open-market operations refers to the Fed's buying and selling of U.S. government bonds. That is, open-market operations are transactions that take place in the public, or open, bond market. Depending on whether the Fed buys or sells bonds, the money supply out in circulation in the economy either increases or decreases:

- » **If the Fed wants to increase the money supply, it buys bonds.** To buy bonds, the Fed must pay cash, which then circulates throughout the economy.
- » **If the Fed wants to decrease the money supply, it sells bonds.** The people to whom the Fed is selling the bonds have to give the Fed money, which the Fed then locks away in a vault so that it no longer circulates.

By buying or selling bonds in this way, the amount of money out in circulation (M^S) can be precisely controlled, meaning that the Fed can, in turn, keep tight control over interest rates.

Lowering interest rates to stimulate the economy

The basic idea behind monetary policy is that lower interest rates cause both more consumption and more investment, thereby shifting the aggregate demand curve to the right. Here's how:

- » **Consumption spending:** Lower interest rates stimulate consumption spending by increasing the attractiveness of taking out loans for buying things such as automobiles and houses.
- » **Investment spending:** Lower interest rates stimulate businesses' investment spending because at lower interest rates, a larger number of potential investment projects become profitable. That is, if interest rates are 10 percent, businesses are only willing to borrow money to invest in projects with rates of return of more than 10 percent. But if interest rates fall to 5 percent, all projects with rates of return higher than 5 percent become viable, so firms take out more loans and start more projects. (For more on how interest rates affect investment, see Chapter 14.)



REMEMBER

When trying to remember how monetary policy works, keep in mind that it's actually a simple three-step process. When the Fed wants to help increase output, it initiates the following chain of events:

1. **It buys U.S. government bonds to increase the money supply.**
2. **The increased money supply causes interest rates to fall because the prices of bonds get bid up.**

3. Consumers and businesses respond to the lower interest rates by taking out more loans and using the money to buy more goods.

The hard part is remembering that higher bond prices mean lower interest rates, which seems counterintuitive. But if you have a hard time remembering that, don't be embarrassed. Many economists get stuck on it, too.

Adding In the Administered Rates

The Fed has another set of policy tools available for controlling the amount of money circulating outside the Fed in the economy at large. They're collectively known as the *administered interest rates*, and they work by incentivizing large financial institutions to park money at the Fed.

There are two administered interest rates. The first is called the *interest rate on reserve balances*, or the *IORB rate*. The Fed pays this to banks for any cash they opt to keep on deposit ("on reserve") at the Fed.

As background, please understand that every commercial bank in the United States has an account at the Fed into which it can deposit money rather than lending that money out to borrowers in the form of business loans, credit card loans, home loans, and so on. Whether banks are more inclined to park cash at the Fed or lend it out to individuals and firms depends on the IORB rate that the Fed sets for banks.

If, for example, the Fed sets the IORB rate at 5 percent per year, then banks would limit their lending to individuals and firms to only those loans on which they could charge more than 5 percent. Banks would do that because depositing money at the Fed is riskless; the Fed can always print up whatever amount of money it might need to pay banks back on whatever amount of money they deposited at the Fed. By contrast, lending by banks to individuals and firms comes with the risk of default. Thus, all other things equal, banks would rather deposit money at the Fed at any given interest rate than lend it to individuals or firms at that same interest rate. The result is that if the Fed offers to pay banks an IORB rate of 5 percent on any money that they deposit at the Fed, banks would limit their lending into the private sector to only those loans that could sustain an interest rate of more than 5 percent.

The Fed knows that, and thus knows that it can control how much money banks will lend into the private sector. The higher the IORB rate, the less banks will lend to the private sector. The lower the IORB rate, the more banks will lend to the

private sector. Thus, by raising or lowering the IORB rate, the Fed has an additional way to control how much money is circulating in the economy. The higher the IORB rate, the more money is on deposit at the Fed and thus not circulating in the economy. The lower the IORB rate, the less money is on deposit at the Fed and thus circulating in the economy.

The second administered interest rate is called the *over-night reverse repo rate*, or *ON RRP rate*. It's the interest rate that the Fed pays to large nonbank financial firms that have the option of depositing money at the Fed rather than circulating that money around the economy. These large nonbank financial firms include money-market mutual funds and the mortgage-lending giants known informally as Freddie Mac and Fannie Mae. The higher the ON RRP rate offered by Fed, the more they'll choose to deposit money at the Fed, thereby reducing the amount of money circulating around the economy.

By raising or lowering the IORB and ON RRP rates, the Fed has an additional method for increasing or decreasing the economy's money supply, M^S . In practice, the Fed uses the administered rates in conjunction with open-market operations to control the overall money supply. In some cases, more emphasis is given to changes in the administered rates, whereas in other cases the Fed relies mostly on open-market operations.

Understanding how rational expectations can limit monetary policy

The government's ability to use increases in the money supply to stimulate the economy is limited by rational expectations and the fears that people have about inflation. Specifically, investors understand that increases in the money supply can cause inflation (as I discuss in Chapter 15). Whenever the Federal Reserve increases the money supply to lower nominal interest rates, it has to do so with some moderation to avoid causing inflationary fears that can offset the stimulatory effect of increasing the supply.

Graphing the results of money supply increases

Look at Figure 17-5, which shows an economy in recession at Point A where aggregate demand curve AD_0 intersects short-run aggregate supply curve $SRAS_0$, which is fixed at price level P_0 . The Federal Reserve then increases the money supply to lower interest rates and stimulate the economy, which causes the aggregate demand curve to shift rightward to AD_1 .

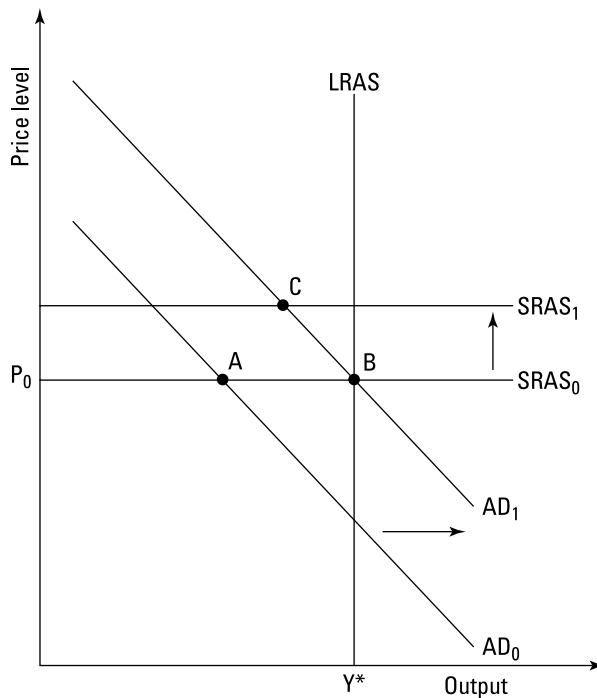


FIGURE 17-5:
The result of
increasing the
money supply
depends on
inflationary
expectations.

© John Wiley & Sons, Inc.

At this point, two things can happen, depending on inflationary expectations:

- » If people believe that the price level will remain fixed at P_0 , the rightward shift in aggregate demand will move the economy's equilibrium rightward along the $SRAS_0$ curve from Point A to Point B.
- » If people believe the price level will jump in response to the increase in the money supply, the short-run aggregate supply curve will shift up vertically by the amount that the price level is expected to increase. That means that the economy's equilibrium will move from A to C, where AD_1 intersects the new short-run aggregate supply curve, $SRAS_1$.

Because output increases less if the economy moves from A to C than if it moves from A to B, the Fed obviously has to be careful about inflationary expectations when trying to stimulate the economy by increasing the money supply. If people expect inflation to occur, their actions can offset some of the stimulus that an increased money supply is expected to bring with it.

Realizing how inflationary expectations affect interest rates



REMEMBER

The underlying problem with monetary policy is that the Fed has only partial control over interest rates. In particular, it controls money supply but not money demand. This is a problem because if people think that an increase in the money supply will cause inflation, they increase their money demand because they're expecting to need more cash to buy things at higher prices.

So while the increase in the money supply tends to lower interest rates, as in Figure 17-4, the increase in money demand caused by inflationary fears tends to increase interest rates. Because higher interest rates tend to decrease investment, any increase in interest rates caused by inflationary fears works against the stimulus that the Fed is attempting to apply to the economy by increasing the money supply.

This decrease in the effectiveness of monetary stimulus is why the big shift in aggregate demand in Figure 17-5 doesn't shift the economy back to producing at Y^* . With people expecting inflation, part of the stimulus ends up causing inflation rather than stimulating the economy to produce more output.

Keeping inflationary expectations low to help monetary policy work well

Since the 1970s, most countries have been cautious when using monetary policy. That's because during the 1970s, countries learned the lesson that if people believe an increase in the money supply is going to cause inflation, an increase in the money supply may mostly end up causing inflation rather than providing stimulus.

You can see an extreme case of this situation in Figure 17-6, where output remains unchanged at the recessionary level Y^{Low} despite an increase in the money supply that causes aggregate demand to shift rightward from AD_0 to AD_1 . The problem is that higher inflationary expectations cause the short-run aggregate supply curve to shift up vertically from $SRAS_0$ to $SRAS_1$, fully offsetting the increase in aggregate demand. The short-run equilibrium shifts from A to B, but the only effect is a higher price level with no increase in output. Economists came to refer to the situation in Figure 17-6 as *stagflation*, in which the economy has a stagnant output level coupled with inflation.



REMEMBER

The experience of stagflation during the 1970s taught the Federal Reserve (and its equivalents in other countries) that monetary policy works best if people believe that the Fed is *not* going to cause inflation. Consequently, these days, the Fed makes only moderate increases in the money supply when it wants to stimulate the economy. These increases end up being more effective than larger increases because they don't trigger inflationary fears.

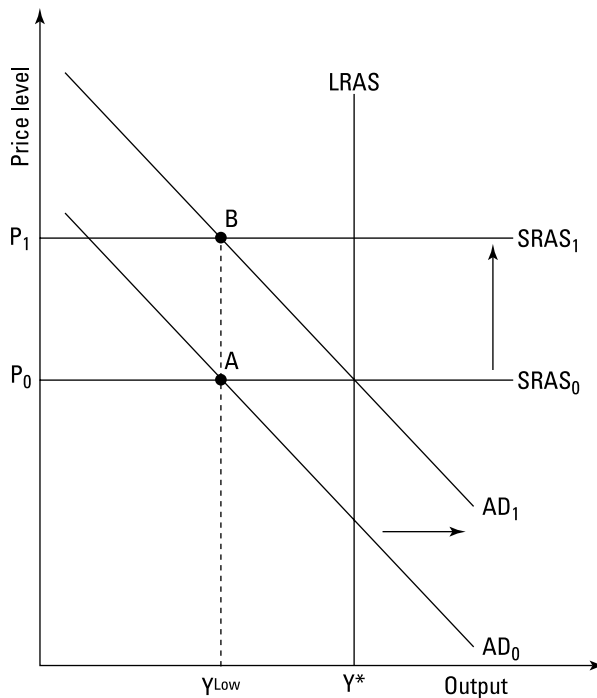


FIGURE 17-6:
An example of
stagflation.

© John Wiley & Sons, Inc.

Examining quantitative easing and the Great Recession

The worldwide Great Recession of 2007–2009 began when a housing bubble in the United States popped in 2006. Trillions of dollars had been invested in the financial markets on the premise that residential housing prices would never decline significantly. As the bubble burst and home prices began to plummet, dozens of large banks as well as many hundreds of financial firms were threatened with bankruptcy.

Lending ceased not only for home mortgages but for business loans, and if it hadn't been for aggressive interventions by governments and central banks, the entire worldwide financial system might have collapsed, such that nobody would have been able to obtain a loan for anything. Because the world economy is highly dependent on borrowed money to finance everything from credit card purchases to factory construction, another Great Depression loomed just over the horizon.

To calm the immediate crisis, the Fed and other central banks became lenders of last resort, ensuring that businesses could still obtain financing directly from the Fed even if banks were reluctant to loan. The Fed also protected banks from

banking panics by increasing the insurance limits offered to checking account depositors. These and other steps limited the recession to being only moderately severe.

That being said, the Great Recession was still much more severe than your typical recession. So it wasn't surprising that the Fed and other central banks undertook innovative new monetary policy measures to stimulate the economy. These came to be known as *unconventional monetary policy*.

Under conventional monetary policy, central banks like the Fed use open-market operations to purchase short-term (less than 1-year maturity) government bonds to increase the money supply and drive short-term interest rates toward zero to stimulate the economy. (For more, see the earlier section "Lowering interest rates to stimulate the economy.")

But given the severity of the Great Recession, the Fed and other central banks found that conventional monetary policy wasn't going to be enough to drive short-term interest rates to zero. Additional stimulus was needed. So the Fed and other central banks began to purchase trillions of dollars' worth of longer-term government bonds (with 5-year, 10-year, and 20-year maturities), private-sector bonds (including mortgage-backed securities), and even stocks. The goal in all cases was to increase the quantity of money available to be loaned out. It was hoped that more borrowing and lending would take place, thereby stimulating both consumption and investment.

These policies became known as *quantitative easing* (or QE) because their goal was to ease the constraints on lending and borrowing by increasing the quantity of money in circulation. By contrast, conventional monetary policy targets not the quantity of money in circulation but rather the *price* of money — the interest rate. Conventional monetary policy attempts to stimulate the economy by lowering the price of loans. Unconventional monetary policy (quantitative easing) attempts to increase the sheer volume of loans.

Another unconventional monetary policy was setting an explicit target for inflation so that people would know how aggressive central banks were going to be with changes in the money supply. The inflation target was set at 2 percent in the United States, which could only be met if the Fed were increasing the money supply faster than any increase in money demand. By setting a 2 percent target, the Fed was committing itself to continually pushing the AD curve to the right and thus always attempting to stimulate the economy. This commitment was intended to give consumers and businesses confidence that the Fed would stimulate as long as necessary until the Great Recession was over and the economy had fully recovered.

IN THIS CHAPTER

- » Understanding that borrowing and debt drive price bubbles and financial crises
- » Seeing how rising asset prices feed bubbles by increasing the value of loan collateral
- » Examining how popping bubbles can take down the banking system
- » Realizing that the post-bubble economy is hamstrung by high debt levels
- » Identifying why fiscal and monetary policy don't work very well after a financial crisis

Chapter **18**

Grasping Origins and Effects of Financial Crises

A *financial crisis* is a period of economic instability triggered by the failure of one or more major financial institutions to fulfill their promises. For instance, banking crises are triggered when banks fail to honor their legal obligation to redeem deposits on demand. And currency crises are triggered when national central banks break promises to maintain fixed exchange rates.

Financial crises can cause *recessions*, periods of decline in the total output of goods and services produced by the economy. Please understand, though, that not every

bank failure or broken promise about an exchange rate triggers a recession. A recession follows a financial crisis only if previous events have led the economy into such a precarious situation that bank failures or exchange-rate collapses make a recession almost inevitable. That happened in the United States and Europe in 2007, when the failure of many banks and investment companies to repay money that they'd borrowed to invest in real estate led to the most severe recession in more than 70 years.



REMEMBER

So what sets the stage for a crisis? Massive amounts of debt. Financial crises are always preceded by economy-wide borrowing binges that fuel unsustainable asset-price bubbles, typically in stocks or real estate. These borrowing binges also fuel economic booms because a great deal of the borrowed money is used to purchase not only stocks or real estate but also goods and services.

As a boom continues, people start thinking it's permanent and that their incomes will keep rising rapidly. This encourages them to take on even more debt, so when the bubble eventually bursts, they find themselves with huge debt but much-lower-than-expected incomes. The economy then goes into a recession as people retrench, directing more of their reduced incomes to paying down debt and less to buying goods and services.

This chapter tells you how debt-driven bubbles develop, how they can cause misallocation of resources in the economy, and why recessions that occur after debt bubbles burst are so severe and hard for governments to deal with.

Understanding How Debt-Driven Bubbles Develop

Debt contracts, such as bonds and mortgages, are promises to repay particular amounts of money. When negotiating such contracts, lenders normally believe that those to whom they're lending will be able to repay, and borrowers typically believe that they'll be able to repay. Otherwise, neither lenders nor borrowers would sign the contract.

However, a circular, self-reinforcing process can develop between the total amount of borrowing and the anticipated ability of borrowers to repay their loans. These situations drive *asset-price bubbles*, in which speculative purchases financed with borrowed money drive the price of an asset (such as real estate) upward.

As long as the bubble continues to expand and prices continue to rise, nearly every loan is paid off — leading both borrowers and lenders to erroneously conclude

that lending and borrowing are very safe. That drives the demand for even more borrowing and lending, further inflating the bubble. This continues until the bubble finally pops and prices collapse. In this section, I give more detail on how bubbles get started and expand, and I discuss price collapses later in “Seeing the Bubble Burst.”

Embracing borrowing in a booming economy



REMEMBER

During the periods preceding financial crises, the expectations of both lenders and borrowers tend to be overly optimistic, usually because the economy has enjoyed a period of sustained growth in output and living standards. With the demand for goods and services high, jobs are plentiful and wages tend to grow strongly. Thus, both borrowers and lenders come to believe that the financial prospects of borrowers are very strong — and therefore that borrowing or lending successively larger amounts of money at successively lower interest rates wouldn’t be overly risky.

The problem is that the lending and borrowing facilitated by the initial optimism can quickly become self-justifying. When people spend borrowed money, that money stimulates the economy even more and causes an even higher level of optimism that justifies making even more loans.

Firms use the loans to expand factories, and individuals use borrowed money to purchase houses, cars, and durable goods. All this economic activity makes it easy for both borrowers and lenders to conclude that the boom will go on indefinitely and that paying back loans will always be easy for borrowers.

Offering larger loans as collateral values rise

The sharp increase in lending and borrowing that precedes a financial crisis is exacerbated by the fact that lending can drive up the value of the assets (such as real estate) that are used as collateral for loans. *Collateral* is property that a borrower pledges to a lender as security for a loan. People refer to the collateral as *security* because if the borrower fails to repay the loan, the lender receives the collateral. Thus, the collateral provides some financial security (safety) for the lender.

For instance, when a home buyer takes out a mortgage, the contract with the bank from which they’re borrowing specifies that the house itself serves as collateral for the mortgage loan. Thus, if the borrower defaults on the loan, the home becomes the property of the bank, which can then sell the home at auction to raise money to pay off the defaulted loan.

Relaxing lending standards

Beyond increasing collateral values, rising prices also cause lenders to become lax about their lending standards in an additional way. During normal times, lenders verify that borrowers can pay off their loans with their labor income. If you have a low labor income, you normally aren't allowed to borrow much money because you can't afford large monthly loan payments.

This logic is shunted aside during periods of rapidly rising real estate prices. In fact, banks become increasingly willing to lend large amounts to anybody, including people with low incomes, because rising real estate prices imply that borrowers should always be able to sell their houses for more than they paid for them.

For example, consider how willing a bank will be to give an economically disadvantaged person a loan of \$200,000 to purchase a house if the price of the house is expected to rise to \$225,000 over the course of the next year. With the market price of the house expected to rise by \$25,000, the bank will assume that the chance that the borrower will default on the loan is virtually zero because they'll always be able to sell the house for more money than they borrowed. Thus, when home prices are expected to keep rising, banks become willing to lend even to those with low incomes.

Borrowing more in hopes of profit

The willingness of banks to lend when home prices are expected to rise is matched by borrowers' desire to take out loans in attempts to profit from the rising real estate prices. For instance, if home prices are expected to rise from \$200,000 to \$225,000 over the next year, potential borrowers will assume that they can easily repay a \$200,000 mortgage loan by selling the house the next year for \$225,000. In fact, they'll expect to come out almost \$25,000 ahead on the deal, even after accounting for accumulated interest on the loan.

The risk is that the house prices won't rise by as much as expected and may even fall, but during a housing-price bubble, many people become so convinced that home prices will rise that they feel that they're not really taking on risk when they borrow money to purchase real estate.

Indeed, because the potential \$25,000 gain looks like free money, many millions of people are tempted to take out loans to purchase houses that they'd otherwise have no reason to care about. Their goal isn't to live in those houses or to rent them out to tenants but rather to use borrowed money to "buy low and sell high," hoping to profit if home prices continue to rise.

Watching the process gain momentum

Both borrowers and lenders are willing to participate in a real estate bubble. The lenders think that rising real estate prices almost certainly guarantee that loans will be repaid, and borrowers think that rising real estate prices almost certainly guarantee that buyers can make money for nothing.

As soon as that process gains momentum, it can become temporarily self-sustaining. Lots of borrowed money drives up real estate prices, causing even more people to think that price increases will be sustained, which tempts both borrowers and lenders to increase borrowing and lending, thereby driving even further price increases. Sadly, though, there's no such thing as a party that never ends — or a party for which nobody has to pick up the tab.

Seeing the Bubble Burst

Debt-driven asset-price bubbles eventually pop because, at some point, borrowing enough money to keep prices rising becomes impossible. Imagine that home prices start off at \$100,000. At that price, potential buyers must be able to borrow \$100,000 if they're to purchase a home. But if home prices are driven up over time to \$300,000, then any potential buyers must get their hands on \$300,000 of borrowed money to purchase a home. And if home prices rise to \$500,000, potential buyers need loans of \$500,000.



REMEMBER

Given the limited amounts of money available for borrowing, buyers will at some point face the reality that they simply can't find enough new money to borrow to drive up prices any higher. At that point, prices will peak.

In this section, I explain why asset-price bubbles pop quickly rather than deflate slowly. The end is rapid because the end of rising prices causes both a massive increase in supply as well as a massive decrease in demand. With supply rising and demand falling, prices fall rapidly.

Deleveraging: Trying to ditch debt as prices fall

People who've borrowed money to purchase assets are in a *leveraged position*, meaning that they've augmented (or in other words, *leveraged*) their own purchasing power with that of the money they've borrowed.

When a price bubble peaks, being in a leveraged position is extremely precarious. The problem is that if you're a borrower, you probably don't have enough labor income to pay off the loan. That's because you took out your loan in anticipation of future price increases. Your plan had been to sell your house for more than you paid for it, thereby getting enough money to pay off the mortgage loan while leaving any extra money as profit.

With prices unable to rise any further (due to new borrowers' being unable to borrow enough money to keep prices rising), you have no incentive to hold on to the property. You want to sell it, pay off your loan, and *deleverage* while the price of houses is still above your purchase price. Unfortunately, so does everyone else in your predicament.

The result is a massive increase in the supply of homes for sale. When prices start to fall due to the increase in supply, sellers begin to panic, undercutting each other on selling price because each one is desperate to sell before the price of houses falls below the price they paid (which is equal to the amount of money they borrowed).

Even worse, the demand for houses collapses at the same time. Why? Because during the bubble, the demand was mostly driven by the anticipation of future price increases. With prices now falling, speculators leave the market.

Taken together, the skyrocketing supply and collapsing demand cause prices to fall precipitously. The price bubble pops, and people who couldn't sell fast enough find themselves "underwater," owing more money on their mortgages than their properties are currently worth. Unless they have large enough labor incomes to make their monthly mortgage payments, they'll end up defaulting on their loans. At that point, the bubble can bring down the banking system and induce an economy-wide recession.

Comprehending bank collapses caused by bursting bubbles

Financial crises begin when bubbles burst. Because millions of borrowers owe more than their property is worth, thousands of banks and other lenders will be unlikely ever to collect on a large percentage of the mortgage loans they made during the bubble. Many of those lenders are likely to go bankrupt because their future income from loan repayments will be substantially less than what they need to meet their current and future financial obligations.

Consider commercial banks, which attempt to support themselves by borrowing money at a low interest rate and lending it out at a higher interest rate. In particular, the money that checking-account depositors place in their accounts at a

particular commercial bank is in fact a loan to that bank. The bank pays these depositors a small amount of interest on their deposits and then lends their money out at a higher rate of interest to finance things such as home mortgages and small-business loans.

Normally, the majority of those loans for mortgages and small-business investments would be repaid on time. But many of the loans made during a bubble will never be repaid. And that means that the bank won't have enough money coming in from loan repayments to ensure that it can return all the money that it owes to its checking account depositors.

In legal terms, the bank is *insolvent*, owing more to the people it has borrowed money from than it can ever expect to get back from the people it has lent money to. Stated simply, an insolvent bank can't honor its promises. The government typically takes over banks that have reached insolvency and uses taxpayer money to guarantee that all checking account depositors will in fact get their money back. After the takeover, those banks cease to make new loans because they don't even have enough money to pay back their depositors.

Similarly, banks that aren't yet insolvent but feel threatened by that possibility also stop making new loans. They do this so that all the income that's flowing in from repayments on the loans that haven't gone bust can be saved up to guarantee the bank's ability to meet its obligations to its checking account depositors.

The net result is that lending in the economy plummets as most banks stop making new loans either because they're already insolvent or because they may become insolvent.

Leading into a recession

After the asset-price bubble pops, the reduction in lending on the part of financially stressed banks reduces the aggregate demand for goods and services in two ways:

- » Entrepreneurs and businesses can't borrow nearly as much money as before to fund the purchase of investment goods such as machinery and computers.
- » Consumers aren't able to get nearly as many loans to finance the purchase of consumer goods.

Furthermore, many of the people who took on large amounts of debt during the bubble will begin trying to pay off their loans. Repayment is possible only by devoting a larger chunk of their paychecks to paying off loans and a smaller fraction to purchasing goods and services. This reallocation of their incomes away

from consumption and toward debt repayment reduces aggregate demand even more, thereby exacerbating the recession.

And that's how the collapse of an asset-price bubble can lead to a recession. After the bubble pops, credit dries up and reduces the total demand for goods and services. With less demand, firms sell fewer products. And then they start laying off employees who are made redundant by the low level of sales. Unemployment rises, spending slows, and GDP (gross domestic product; see Chapter 14) begins to decline.

After the Crisis: Looking at Recovery

An international historical examination of financial crises reveals that the recessionary periods that follow financial crises last several times longer than those following recessions that don't involve the building up of massive amounts of debt. This section examines why the recovery from post-bubble recessions tends to take so long and why government policies that can work well against normal recessions have trouble speeding up the recovery process after an asset-price bubble collapses. Here are the two main culprits:

- » A weak banking system that can't make many new loans
- » *Structural mismatches* between the goods and services that the economy's existing firms are capable of producing and the goods and services that consumers actually demand in the post-bubble period

Enduring a broken banking system

In the wake of a financial crisis, a nation's banking system tends to be weak and unable to extend many loans. Many banks become insolvent and are forced out of business as the borrowers default on their loans. (See the earlier section "Comprehending bank collapses caused by bursting bubbles" for details.) Other banks survive but are typically in weak financial condition because they made so many bad loans during the bubble.

As a result, the post-crisis banking system has very little capacity to make loans as the economy recovers from its post-bubble recession. The weak lending capacity prolongs the recession because even when consumers and firms regain confidence and want to borrow and spend again, they find few banks willing to make loans.

By contrast, ordinary recessions tend to do very little damage to banks because these recessions aren't preceded by asset-price bubbles. When the economy is recovering from a normal recession, loans are usually much more widely available. This speeds recovery by allowing firms and consumers to borrow and spend more freely.

Struggling with structural mismatches

Post-bubble recessions tend to be longer and more severe than ordinary recessions partly because of structural mismatches. An economy has a *structural mismatch* between its production capacity and the products consumers demand if the mix of goods and services that the economy's firms are capable of producing differs from what consumers want to purchase and consume.

Post-crisis economies often feature structural mismatches because of distortions to productive capacity that occur while bubbles are expanding. For example, consider the United States housing bubble of 2000–2006 and the subsequent recession of 2007–2009.

The U.S. housing bubble of 2000–2006 was the largest real estate bubble in world history. During the expansionary phase of that bubble, credit was easily available throughout the economy. This led to \$2 trillion of excess housing being built and to many firms borrowing trillions of dollars to fund new investment projects.

But even after the recession ended and the economy began to slowly grow again in the summer of 2009, unemployment remained high and businesses were producing at levels far below their capacity. Some experts ascribed this to low aggregate demand caused by banks' reducing lending and consumers' reducing consumption to pay down debts.

Other experts, however, feared that the U.S. economy was suffering from a structural mismatch. Their reasoning was that, during the bubble, firms had used their access to easy credit to build up lots of capacity to produce things that were popular during the boom — large and new houses, fancy malls, and plenty of SUVs and large trucks. After the bubble popped, however, consumers didn't want lots of new houses or new SUVs or additional large trucks. Instead, they wanted products such as iPads, better touch-screen cell phones, and smaller cars, as well as the ability to do more of their shopping over the internet instead of at shopping malls.

If that interpretation is correct, then recovering from the 2007–2009 recession entailed a much slower recovery process than recovering from a plain-vanilla cyclical recession. The added difficulty came from having to rejigger the economy's production capacity from making the products that were in demand while

the bubble was expanding to producing the different goods and services in demand after the bubble burst.

Making that sort of transition requires revamping old firms or starting new ones, moving workers away from dying industries into new industries, and undertaking the time and expense of retooling factories and restructuring supply chains. Making those adjustments all over the economy isn't quick.

Noting the limits of government policy

Whether a recession is preceded by a bubble or not, governments almost always attempt to use both fiscal and monetary policy to increase aggregate demand, as spelled out in Chapter 17. However, government stimulus policies often appear unable to significantly speed up the recovery process after a financial crisis. The policies are stymied by the debt that remains after the bubble pops.



REMEMBER

In particular, both fiscal and monetary policy are of limited effectiveness because post-bubble consumers want to deleverage, or get rid of their debt. Here's how the desire to deleverage hampers both fiscal and monetary policy:

» **Fiscal policy:** Post-crisis fiscal policy comes in the form of massive increases in government spending that are intended to stimulate aggregate demand by having the government purchase lots of goods and services. The hope is that by increasing people's incomes, those initial purchases will spur further economic activity that will snowball into robust economic growth as consumers return to spending confidently.

But in the aftermath of a debt-driven bubble, fiscal policy's stimulatory effects may be limited because people often use increases in income to pay down debt rather than to purchase additional goods and services.

» **Monetary policy:** The government attempts to use monetary policy to stimulate demand by lowering interest rates to encourage both consumers and firms to borrow and spend more.

However, monetary policy fails to work very well after a debt-driven bubble because highly leveraged consumers are in no mood to borrow more money. Understandably, they're more interested in paying down their current debts than in taking on more.

If this situation sounds dismal, it is. Until the debt level in the economy falls so that people don't need to devote so much of their incomes to paying off loans, economic growth is likely to remain stagnant, and government attempts at monetary and fiscal policy are likely to prove ineffective.

5

The Part of Tens

IN THIS PART . . .

Discover ten common — but false — beliefs held about economics.

Get the scoop on ten valuable insights from economics.

Find out about some of the most famous economists who ever lived.

Learn about recent Nobel laureates in economics.

- » Avoiding logical fallacies that sucker intelligent people
- » Steering clear of bad economic reasoning

Chapter **19**

Ten Seductive Economic Fallacies

In this short chapter, I outline the most attractive and compelling incorrect ideas in economics. Some are logical fallacies. A few are myopic opinions that don't consider the big picture. And others are poorly thought-out examples of economic reasoning. All are to be avoided.

The Lump-of-Labor Fallacy

The argument that there's a fixed amount of work that you can divide up among as many people as you want is often presented as a cure for unemployment. The idea is that if you convert from a 40-hour work week to a 20-hour work week, firms will have to hire twice as many workers. France once reduced its work week to 35 hours hoping that firms would hire more workers and cure the country's unemployment problem.

It didn't work. Such policies have never worked. One problem is that hiring workers involves many fixed costs, including training costs and health insurance. So

two 20-hour-per-week workers cost more to employ than one 40-hour-per-week worker. What's more, two 20-hour-per-week workers don't produce more output than one 40-hour-per-week worker.

So if laws were passed that forced firms to move from a 40-hour work week to a 20-hour work week, firms wouldn't double the size of their workforces. They'd hire fewer than twice as many workers because costs would rise.

In addition, even if cutting the work week in half actually did double the number of workers who were employed, it would only hide the overall unemployment problem by spreading it around. If, for example, 100 percent of workers are working half-time, they are all 50 percent underemployed. In terms of the total amount of output that could be produced (and hence available for people to consume), that situation wouldn't be any better than having 50 percent of the population employed full-time and 50 percent totally unemployed.



REMEMBER

You really want a situation in which every worker who wants a full-time job is able to get one. Shortening the work week doesn't achieve that goal.

The Overpopulation Fallacy

Various versions of this myth have been floating around since the late 18th century when Thomas Malthus first asserted it. He argued that living standards couldn't permanently rise because higher living standards would cause people to breed faster. He believed that population growth would outpace our ability to grow more food, so we would be doomed to return to subsistence levels of nutrition and living standards.

Even when Malthus first published this idea, lots of evidence indicated that it was bunk. For generations, living standards had been rising while birth rates had been falling. And because that trend has continued up to the present day, we're not going to breed our way to subsistence. Indeed, many nations now face an *underpopulation* problem because birthrates have fallen below the replacement rate necessary to keep the population stable. The populations of Italy and Japan (among other countries) started declining in the 2010s and China's began to decline in 2022.

A related problem is that rapidly falling birth rates are wreaking havoc on government-sponsored retirement systems because there aren't enough young workers to pay all the taxes needed to fund retirees' pensions.

The Sequence Fallacy

Post hoc ergo propter hoc is a Latin phrase that roughly translates as, “Because you see one thing precede another, you think that it causes the other.” That is, if A happens before B, you assume that A causes B. Such a deduction is false because A and B often don’t have any relationship. For instance, sometimes it rains in the morning and I get a headache in the afternoon. That doesn’t mean that the rain caused my headache.

Politicians try to pull this logical fallacy all the time when discussing the economy. For instance, suppose that politician A gets elected, and a recession begins a few months later. The two may have nothing to do with each other, but you can be sure that during the next election, politician B — an opponent of politician A — will claim that the recession was the result of politician A’s policies. The only proof offered is that one event happened before the other.

The Protectionism Fallacy

Many trade unions and politicians argue in favor of trade barriers and taxes on imports on the grounds that these policies would benefit citizens and prevent jobs from being exported. The problem is that their arguments consider only the benefits of protectionism. Trade barriers and taxes on imports *do* protect the specific jobs that they’re intended to protect. However, other jobs are often sacrificed in the process.

Another problem with protectionism is that citizens are consumers as well as producers. For instance, if the government prevents the importation of lower price, higher quality foreign automobiles, it preserves jobs in the domestic auto industry. But costs for domestic consumers rise as a result.

Protecting an unproductive industry that faces foreign competition only allows it to keep using resources that would otherwise be better used by more vibrant industries. Workers who would otherwise move to jobs in innovative, highly productive new industries instead get stuck in an industry so unproductive that it can survive only by having the government rig the economy in its favor.

Granted, the move from a dying industry to an innovative new industry can be rough for an individual worker. But rather than avoid the need for change by protecting unproductive industries, the government can help domestic workers more

efficiently by providing retraining programs for employees. (In the case of older workers who have only a few years of employment left, early retirement programs may be more viable than retraining.)

The Fallacy of Composition

Assuming that what's good for one person to do is good for everyone to do all at once is another common fallacy. For instance, if you're at a sold-out sporting event or concert and want to get a better view, standing is a good idea — but only if you're the only one who stands up. If everyone else also stands up, everyone's view is just as bad as when everyone was sitting down (but now everyone's legs are getting tired). Consequently, what was good for you to do alone is actually bad for everyone to do at the same time.

The fallacy of composition is false because some things in life have to do with relative position. For instance, if you start out as the lowest paid employee at your firm but then get a 50 percent raise while nobody else gets a raise, your relative position within the firm improves. However, if everyone gets a 50 percent raise at the same time, you're still the lowest paid person at the firm. If what matters to you is your relative standing within the firm, getting the same raise as everyone else doesn't make you any happier. (On the other hand, if you're more interested in where you stand relative to people who work at other firms, getting a 50 percent raise is good even if everyone else at your firm gets it too!)

The All-In Fallacy

We all value safety. But was a famous U.S. politician really being sensible when he said that we should spend whatever money might be necessary to make flying on commercial airlines “as safe as possible”? Economists would say, “No!” The problem is that making commercial airline travel “as safe as possible” would mean making it prohibitively expensive.

The politician failed to apply *marginalism* — the idea that the best way to approach a problem is to compare marginal benefits with marginal costs. The first few airline safety innovations (such as seatbelts and radar) are sensible to undertake because the extra, or marginal, benefit that each brings is greater than the extra, or marginal, cost required to pay for it. But after the first few safety innovations are implemented, successive innovations become more costly and less effective. At some point, additional innovations bring only small marginal increases in safety while running up high marginal costs.



REMEMBER

When the costs for the extra safety innovations exceed their benefits, they *shouldn't* be implemented. You should add safety features only as long as the marginal benefits exceed the marginal costs — which means that you'll usually stop adding safety features long before you get anywhere near making things “as safe as possible.”

The Market-Instability Fallacy

Free market prices and quantities often change extremely rapidly. They do so because supply and demand often change very quickly, causing rapid changes in equilibrium prices and quantities, which I discuss in Chapter 4. But rapidly changing prices and quantities shouldn't be seen as a problem in and of themselves. The responsiveness of markets is actually one of their great benefits. Unlike a government bureaucracy that can never react quickly to anything, markets can adjust to huge changes in world events in only minutes especially with the speed that news travels on social media.

The new equilibrium prices and quantities see to it that resources are allocated to their best uses and that society suffers from neither shortages nor gluts. So don't call markets unstable. Call them *responsive*.

The Cheap-Foreign-Wages Fallacy

Suppose that a U.S. factory pays its workers \$20 per hour while a competing factory in a developing country pays its workers \$4 per hour. People mistakenly jump to the conclusion that because the developing country's factory labor costs are so much lower, it can easily undersell the U.S. factory. But this argument fails to take into account two things:

- » What actually matters is labor costs per *unit*, not labor costs per *hour*.
- » Differences in productivity typically mean that labor costs per unit are often nearly identical despite huge differences in labor costs per hour.

To see what I mean, compare how productive the two factories are. Because the U.S. factory uses more advanced technology, one worker in one hour can produce 20 units of output. The U.S. worker gets paid \$20 per hour, so the labor cost *per unit of output* is \$1. The factory in the developing country is much less productive; a worker there produces only 4 units in one hour. Given the foreign wage of \$4 per hour, the labor cost per unit of output in the developing country is also \$1.

Obviously, the developing country's lower hourly wage rate *per hour* doesn't translate into lower labor costs *per unit* — meaning that it won't be able to under-sell its U.S. competitor.

People who focus exclusively on the difference in labor costs per hour never mention the productivity differences that tend to equalize labor costs per unit. And don't think that my example uses happy-happy numbers. Wage differences across countries really do tend to reflect productivity differences.



TIP

Keep in mind that governments can seriously screw up what would otherwise be a near equality of labor costs per unit by fixing artificially low exchange rates. For instance, if at an exchange rate of 5 Chinese yuan to 1 U.S. dollar labor costs per unit are equal, the Chinese government could make labor costs per unit look artificially low to U.S. consumers if it fixes its currency at, for instance, 8 yuan to 1 dollar. In such situations, the inability of U.S. workers to compete with Chinese workers is due to the currency manipulation, not to the lower wage rate per hour found in China.

The Tax-Incentives Fallacy

Some politicians argue for raising income taxes as though the only effect of doing so will be to raise more money. But it's been demonstrated over and over again that beyond a certain point, people respond to higher taxes by working less. And that reduction in labor denies society all the benefits that would have come from the extra work. (Because people work less, the increased tax rate also doesn't bring in nearly as much revenue as expected.)

So if you see a politician arguing for an increase in income taxes, look into the details to make sure that the disincentive effects of the tax hike don't cause more mischief than the benefits that will be derived from spending the money that the higher tax generates.

The Unintended-Consequences Fallacy

When evaluating a policy, people tend to concentrate on how the policy will fix some particular problem while ignoring or downplaying other effects it may have. Economists often refer to this situation as *The Law of Unintended Consequences*. For

instance, suppose that you impose a tariff on imported steel to protect the jobs of domestic steelworkers. If you impose a high enough tariff, their jobs will indeed be protected from competition by foreign steel companies. But an unintended consequence is that the jobs of some autoworkers will be lost to foreign competition. Why? The tariff that protects steelworkers raises the price of the steel that domestic automobile makers need to build their cars. As a result, domestic automobile manufacturers have to raise the prices of their cars, making them relatively less attractive than foreign cars. Raising prices tends to reduce domestic car sales, so some domestic autoworkers lose their jobs. The higher domestic price of steel will also hurt workers in other industries that use steel as an input, including construction, ship building, and appliance manufacturing.

Unintended consequences are far too common. Be aware of them whenever a politician tries to persuade you to see things their way. Chances are that they're mentioning only the good results of a certain policy; they may not have even thought about its not-so-good side effects.

- » Understanding ten fundamental economic principles
- » Arming yourself against the economic follies of politicians, the media, and others

Chapter 20

Ten Economic Ideas to Hold Dear

In this chapter, I list ten economic ideas that every informed person should understand and be ready to use to evaluate the policy proposals that politicians, pundits, and the media make. Some of these ideas aren't true in all situations, but because they're usually correct, be wary if someone wants you to believe that they don't apply to a particular situation. Chances are that they're wrong.

Self-Interest Can Improve Society

The idea that self-interest can improve society is basically Adam Smith's famous *invisible hand*. If all economic transactions in a society are voluntary for all the parties involved, then the only transactions that will take place are those in which all parties feel that they're being made better off. This concept doesn't mean that charitable acts are bad for society. Rather, it means that even philanthropy is generated by self-interest: People give because they enjoy helping others, so both givers and the people they help are better off. In addition, this concept motivates entrepreneurs to find ways to produce products you like at prices you like. Sellers must fear your power of choice — your right to walk away from any offer that you don't consider advantageous.

Free Markets Require Regulation

Economists firmly believe that voluntary transactions in free markets tend to work toward the common good. But they also believe that nearly every participant in the marketplace would love to rig the system in their own favor. Adam Smith was quick to argue that for markets to work and serve the common good, the government has to fight monopolies, collusion, and any other attempts to prevent a properly functioning market in which firms vigorously compete against each other to give consumers what they want at the lowest possible price. Government intervention may also be necessary to deal with asymmetric information, public goods, and externalities.

Economic Growth Relies on Innovation

The only way to have sustained economic growth and widespread increases in living standards is to invent more efficient technologies that allow people to produce more from the limited supply of labor and physical resources. To that end, societies should promote education and set up institutions such as patent rights, competitive markets, and copyright laws to promote innovation and increases in efficiency.

Freedom and Democracy Make Us Richer

Good moral and ethical reasons exist for favoring freedom and democracy. But a more bottom-line reason is that because they promote the free development and exchange of ideas, free societies have more innovation and, consequently, faster economic growth.

Education Raises Living Standards

Not only do educated people produce more as workers — and get paid higher salaries — they produce innovative new technologies. Sustained economic growth and higher living standards are possible only if you educate your citizens well. You have, of course, other good reasons for getting an education, including the ability to appreciate high art and literature. But if all you care about is rising living standards, work hard to promote education in science and engineering, sectors where revolutionary technologies are created.

Intellectual Property Boosts Innovation

People need incentives to take risks. One of the biggest risks is leaving a secure job to start a new business or develop a great new idea. *Intellectual property rights*, such as patents and copyrights, guarantee that you'll be the only one making money off your hard, innovative work. Without this assurance, few would be willing to take the risks necessary to provide new technologies and products.

Weak Property Rights Cause All Environmental Problems

Environmental problems stem from poorly defined or nonexistent property rights that allow polluters to ignore the costs they impose on others. Therefore, economists favor the creation and enforcement of property rights systems that force people to take all costs into account. People always need to do some polluting. After all, even if you don't want gas-guzzling SUVs causing lots of air pollution, you probably would still be willing to allow gas and diesel ambulances and fire trucks to operate even though they, too, pollute the environment. Their benefits to society outweigh their costs of pollution. As I discuss in Chapter 10, strong property rights are the key to ensuring that people weigh the complete costs and benefits of causing pollution. Property rights force people to consider not only personal costs but also the costs that their actions impose on others.

International Trade Is a Good Thing

Opening your country to international trade means opening your country to new ideas and innovations. Competition causes local businesses to innovate to match the best offerings of companies from around the world. Throughout history, the richest and most dynamic societies have been the ones open to international trade. Of course, what economists have in mind when they think of the benefits of international trade is *free trade*, where companies compete across borders to provide people with the best goods and services at the lowest prices. Economists strongly condemn the government subsidies and trade restrictions that impede free trade and that try to rig the game in one country's favor.

Government Can Provide Public Goods

Economists view the existence of public goods as one of the most important justifications for government intervention in the economy. Although private philanthropy can provide some public goods, many are so expensive that they can be provided only if the government uses taxes to fund them. Consequently, public goods are typically government-provided.

Private firms can provide goods and services only if they can at least break even doing so. To break even (or make a profit), whatever a firm is selling has to be *excludable*, which means that only those paying for the good or service receive it. As I explain in Chapter 11, some goods and services aren't excludable. For instance, a lighthouse provides warning services to all ships in the vicinity regardless of whether the ships' captains pay the lighthouse keeper. The private lighthouse quickly goes bankrupt because only a few captains are fair-minded enough to pay for the service. Goods and services that aren't excludable are called *public goods* because they're essentially open to the public and can't be kept private.

Governments can force people to pay for public goods by levying taxes and using the tax revenues to pay for such things as national defense, police departments, lighthouses, public fireworks displays, and basic scientific research.

Preventing Inflation Is Easy

High rates of inflation are caused when the government increases the money supply too rapidly. A growing economy always has a growing demand for money because with more stuff to buy, you need more money to buy it with. To keep the overall level of prices constant, the government needs to increase the money supply at the same rate as the demand. If the supply of money increases faster than the demand, the value of money falls, creating inflation. In other words, you need more money to buy the same amount of stuff as before, so prices go up. Modern central banks such as the Federal Reserve Bank in the United States can do this quite easily, so there's no excuse for high rates of inflation.

- » Realizing that no economist works in isolation
- » Singling out some amazing economists

Chapter **21**

Ten (Or So) Super-Famous Economists

This short chapter describes the ideas put forth by eleven of the best and most influential economists in world history. (Ten wasn't enough.) Each either radically changed the way that economics conceptualizes the world or radically changed the way that politicians and government officials formulate public policy.

Adam Smith

Adam Smith (1723–1790) developed the intuition that as long as firms are constrained by robust competition, their self-interested profit seeking inadvertently causes them to act in ways that are socially optimal — as though they're guided by an *invisible hand* to do the right thing.

But Smith wasn't naive. He believed that businesspeople prefer to collude rather than compete whenever possible and that governments have an important economic role to play in fostering the robust competition needed for the invisible hand to work its magic. He also believed that governments must provide many essential public goods, like national defense, that the private sector don't readily produce.

David Ricardo

David Ricardo (1772–1823) discovered comparative advantage and argued correctly that international trade is a win-win situation for the countries involved. Comparative advantage destroyed the intellectual respectability of *mercantilism*, the mistaken theory behind colonialism that viewed trade as being one-sided and consequently argued that trade should be set up to benefit the mother country at the expense of its colony.

Ricardo correctly analyzed the economic phenomenon of diminishing returns, which explains why costs tend to increase as you increase production levels. He was also a strong early proponent of the quantity theory of money, the idea that increasing the money supply will increase prices, all other things held equal.

Karl Marx

Karl Marx (1818–1883) was the foremost economist among 19th-century socialists. None of his major economic theories is now believed to be true, but because proponents of his Marxist ideas came to power in dozens of countries during the 20th century, he is surely one of the most influential economists who ever lived. (Marx gets the most space here not because he's the most important economist on this list, but because I have to take the time to explain his ideas before discrediting them. The ideas of the other economists on this list are already explained in detail in other places in this book.)

Marx's most important intellectual contribution was his idea that capitalism is a historically unique form of social and productive organization. In his book *Capital*, he analyzed capitalism as a brand-new form of social and economic organization based on capital accumulation and factory production. He called the owners of the factories "capitalists" and argued that they would be forced to exploit the workers who labored in their factories.

In particular, he believed that the only capitalists who would survive and whose businesses would grow were those who paid workers the minimum salaries necessary for the workers to survive. Thus, even as productivity and output rose rapidly, workers would endure permanent, grinding poverty out of which they could never rise except by means of a violent overthrow of the capitalists that would lead to the workers gaining control over the factories.

Marx argued that this violent overthrow would be facilitated by what he saw as an inevitable tendency toward concentration and monopoly. When there was only one monopoly firm in each industry, it would be much easier for the workers to revolt and take over the system.

With more than a century of hindsight, we know that Marx was wrong in his economic thinking. In particular, workers' wages *do* rise over time — in fact, they rise on average as fast as technological innovation increases productivity levels. That's because capitalists compete over the limited supply of workers, and wages are bid up as quickly as productivity improvements allow one capitalist to bid higher wages to steal workers away from other capitalists. In addition, competition does *not* lead to each industry being dominated by a single monopoly firm. Rather, competition remains robust in most industries and consequently delivers all the benefits of Adam Smith's invisible hand.

Alfred Marshall

Alfred Marshall (1842–1924) invented the supply-and-demand method for analyzing markets. Applying mathematics to economic theory, he clearly differentiated between *shifts* of demand and supply curves and movements *along* demand and supply curves. In doing so, he cleared up 2,000 years of faulty reasoning. He also made the revolutionary prediction that the market price would be where the demand and supply curves cross.

Marshall went one step further and realized that by comparing points along demand and supply curves with the market price, you could quantify the benefits consumers and producers derive from market transactions: The sum of consumer surplus and producer surplus is the total economic surplus.

This method of quantifying the benefits of production and consumption is still used today and forms the basis of *welfare economics*, which studies the costs and benefits of economic activities. This method also happens to illustrate the intuition behind Adam Smith's invisible hand. The free market equilibrium, where demand and supply cross, is exactly the same as what a benevolent social planner would choose to do to try to maximize social welfare by maximizing total economic surplus. In other words, a free market does indeed act “as if moved by an invisible hand” to promote the common good.

John Maynard Keynes

John Maynard Keynes (1883–1946) invented modern macroeconomics and the idea of government-provided economic stimuli overcoming recessions. Much of the rest of 20th-century macroeconomics was a response to his ideas.

His most famous ideas were developed in response to the long agony of the Great Depression of the 1930s. He first asserted that the Great Depression was the result of a collapse in the expenditures being made on goods and services. He then asserted that monetary policy had been ineffective in combating the decline in expenditures. Finally, he concluded, given his dismay about monetary policy, that fiscal policy was the only remaining source of salvation. In particular, Keynes believed that the best way to increase expenditures in such dire circumstances was for the government to spend heavily to pay for programs that would buy up lots of goods and services to get the economy moving again.

Keynes's policy prescriptions were adopted during the Great Depression in many countries, including the United States. And although many of his ideas about the cause of the Great Depression and the best policies for dealing with recessions are no longer embraced, his underlying idea that governments are responsible for taming the business cycle remains very much with us today.

Kenneth Arrow and Gerard Debreu

Kenneth Arrow (1921–2017) and Gerard Debreu (1921–2004) mathematically proved that Adam Smith's intuition about the invisible hand was, in fact, correct. Not only do competitive firms offer society the utility-maximizing combination of goods and services, they do so efficiently, at minimum cost. Because this proof came in the 1950s, it served to disprove the assertions of totalitarians and communists that centrally planned economies were more productive or more efficient than market economies.

Milton Friedman

Milton Friedman (1912–2006) convinced economists that the quantity theory of money is, in fact, true: Sustained inflations are the result of sustained increases in the money supply (printing too much money). This insight limited using monetary policy to stimulate the economy.

Friedman also argued that the Great Depression was chiefly a monetary disaster and that its severity was the result of a gruesomely tight money supply that kept real interest rates much too high. This diagnosis of the cause of the Great Depression is now the standard explanation, meaning that the intellectual ammunition for Keynes's solution to recessions — large increases in government spending — has lost much of the sway that it once had. It has also led many economists to conclude that monetary policy is more important than fiscal policy for regulating the economy and preventing recessions.

Paul Samuelson

Paul Samuelson (1915–2009) made many contributions. Perhaps the most important was crystallizing the idea that all economic behavior can be thought of as consumers and firms maximizing either utility or profits subject to a set of constraints. *Constrained maximization* has become the dominant paradigm that governs how economists conceive of economic behavior.

Samuelson also developed a judicious blending of Keynesian and classical ideas about the proper use of government intervention in the economy. Keynes argued for large government interventions to mitigate recessions. Smith and Ricardo argued for minimal government interventions, fearing that government interventions tend to make things worse. Samuelson's *neoclassical synthesis* states that during recessions the government should be willing to make large interventions in the economy to get it moving, but when it's operating at full potential, the role of government is to provide public goods and take care of externalities. Many economists embrace this view.

Robert Solow

Robert Solow (b. 1924) has made huge contributions to the understanding of economic growth and rising living standards. In addition to developing innovative models of how economies grow, he showed that the dominant long-run force propelling economic growth is technological innovation.

Before Solow, the economic profession believed that increases in output were the result of increases in inputs. In particular, increases in output were solely the result of using either more workers or more capital (such as bigger factories). What Solow demonstrated was that *at most* 50 percent of the long-run growth of living standards can be explained by increases in labor and capital. The rest has to be the result of technological innovation.

This created a huge paradigm shift among economists that has resulted in the study of technological innovation and how government policies like patents and copyrights can improve it. It also opens up the refreshing possibility that technological innovation will allow us to enjoy higher living standards without having to constantly increase use of the earth's resources.

Gary Becker

Gary Becker (1930–2014) was hugely influential because he pushed economics into areas that were previously immune to economic thinking. His first major contribution was to argue that free markets would tend to work *for* equality and *against* racial and gender discrimination. The intuition is that firms that refuse to hire the best qualified workers because of their race or gender put themselves at a competitive disadvantage relative to nonbiased firms. Becker backed up this intuition by showing that industries that are more competitive do, in fact, employ more minorities and women.

Another significant contribution Becker made was to model families as economic units in which family members tend to act on the basis of cost–benefit analyses. As societies became richer and paid employment became more plentiful (and better paying), Becker predicted that more women would choose to work rather than stay home. He provided an economic explanation for a huge change in the labor force that otherwise would have been explained only in terms of sociological considerations (such as changing gender roles).

Similarly, he was the first to model criminal behavior in terms of how criminals view the potential costs and benefits of committing a given crime. If the expected benefits exceed the expected costs, the criminal will most likely attempt the crime. This theory of criminal behavior is radically different from previous explanations, and it led Becker to propose the influential idea that the best way to deter crime is to raise the costs relative to the benefits.

Robert Lucas

Robert Lucas (1937–2023) showed that people are sophisticated planners who constantly modify their optimal strategies in response to government policy. That's important because if people are forward-looking and rational, then a government

policy change that's designed to modify people's behavior may fail as people adjust their optimal strategies to deal with the new situation. In particular, monetary policy may lose most of its effectiveness if people rationally plan for monetary-policy changes.

As an example, suppose the government announces that in three months it's going to double the money supply in an attempt to stimulate increased purchases of goods and services. If store owners keep prices the same even though more money is on the way, the economy will be stimulated because people will buy more stuff with that new money. But if shop owners double their prices in anticipation of the doubling of the money supply, the policy change won't result in increased goods and services sold. With prices twice as high, twice as much money will only allow customers to buy as much as before.

Lucas's idea came to be known as *rational expectations*, and it brought with it a new humility about the extent to which government policy — monetary policy in particular — can influence the world.

- » Highlighting some outstanding recent Nobel Laureates in Economics
- » Explaining their pathbreaking new ideas

Chapter 22

Ten (Or So) Recent Nobel Laureates in Economics

This chapter provides short biographies for a dozen of the most influential economists of the late 20th and early 21st centuries. Each of them either (1) substantially improved the theories that economists use to understand what's going on in our often-confusing world, or (2) developed or improved methods for analyzing and interpreting real-world data. Several of them did both.

I hope you find this list inspiring, especially if you're young. Economics has made great strides in my lifetime, but there's still so much more left to do!

Amartya Sen

Amartya Sen is an Indian economist who won the Nobel Prize in Economics for his contributions to welfare economics, which is the branch of economics that evaluates — and develops measures to improve — individual and collective well-being.

Sen's work centers on the idea that measuring a country's economic growth in terms of gross domestic product (GDP) doesn't tell the whole story. True economic development requires considering a range of factors, such as education, healthcare, and political freedoms, all of which contribute to well-being.

According to Sen, the goal of economic development should be “human flourishing.” This means providing for education and healthcare, ensuring political and civil rights, and creating an environment in which people can build capabilities and lead fulfilling lives.

Daniel Kahneman

Daniel Kahneman is an Israeli-American psychologist who revolutionized the study of decision-making with his work on prospect theory.

Prospect theory does a much better job than previous theories of human decision-making because it considers that people’s decisions are based not on expected final outcomes, per se, but on whether those potential outcomes are perceived as being gains or losses relative to a decision-maker’s current situation.

Other work by Kahneman overtured a variety of decision-making theories that assumed that people always made rational, self-interested decisions and showed that people’s choices were in fact often influenced by irrational biases that they were unaware of. Kahneman is considered to be the godfather of behavioral economics, and his work has been widely influential in economics, psychology, and related fields.

Elinor Ostrom

Elinor Ostrom was an American political scientist whose groundbreaking work in studying the management of common-pool resources challenged the conventional wisdom of the “tragedy of the commons,” under which it was believed that only property-rights systems enforced by governments could prevent natural resources from becoming overused or completely exhausted.

Ostrom’s research and writings on “governing the commons” highlighted the effectiveness of local community-based solutions and the need to empower people in managing their shared resources, while emphasizing the importance of cooperation, communication, and trust among resource users.

Ostrom was the first woman to be awarded a Nobel Prize in Economics.

Dan McFadden

The American microeconomist Dan McFadden was awarded the Nobel Prize for his work on developing innovative statistical methods that could be used to make accurate predictions about how consumers with a limited set of discrete choice options (for instance, commuting to work by bus, bicycle, train, or car) would react to changes in prices and other factors.

Those same statistical methods have also been applied in fields as diverse as political science, public health, and labor economics. Examples include understanding voting behavior, studying the factors that affect people's choices of medical treatments, and studying how employees make job and career choices in the presence of discrimination.

Vernon Smith

Vernon Smith is an American economist who won the Nobel Prize in Economics for pioneering the field of experimental economics, in which controlled experiments are used to test economic theories and hypotheses.

One of Smith's most notable experimental contributions was the confirmation of the "law of supply and demand" in experimental settings, showing that markets could be efficient and self-regulating even in the absence of government intervention.

Smith also developed the concept of "induced-value auctions," which are now commonly used in the sale of government assets, such as telecommunications frequencies.

George Akerlof

George Akerlof is an American economist who won the Nobel Prize in Economics for his work in the field of information asymmetry.

Akerlof's most famous publication, "The Market for Lemons," focused on understanding how incomplete or asymmetric information can result in the total and utter failure of markets (as we explain in detail in Chapter 11).

Akerlof has also published highly influential research on a wide-ranging set of topics, including the economics of discrimination, the role of social norms in shaping economic behavior, and the impact of government policies on employment, inflation, and other macroeconomic outcomes.

Esther Duflo

Esther Duflo is a French–American economist who won the Nobel Prize in 2019 for her experimental approach to alleviating poverty in developing countries.

She cofounded the Abdul Latif Jameel Poverty Action Lab, a network of economists committed to fighting poverty through randomized controlled trials. Her work has produced innovative strategies for education, health, and financial services, such as using data to improve immunization rates in India, studying the health effects of iron supplements in developing countries, and investigating how empowering women politically affects economic development.

Duflo was the second woman and youngest person ever to win a Nobel Prize in Economics.

Paul Krugman

Paul Krugman is an American economist who won the Nobel Prize for his work in the field of international trade theory, and in particular his pioneering work in new trade theory and new economic geography.

New trade theory explains why countries specialize in the production of certain goods, even if they have similar levels of technology and resources. This theory posits that economies of scale, or the cost advantages gained by producing large quantities of goods, drive trade patterns.

New economic geography, by contrast, explains why certain industries or economic activities cluster together in certain regions, leading to the development of economic hubs like Silicon Valley. This theory emphasizes factors like transportation costs, economies of scale, and consumer preferences to explain the spatial distribution of economic activity.

Krugman is also known for his fundamental theoretical work on currency crises.

David Card

David Card is a Canadian-American who won the Nobel Prize in Economics in 2021 for his innovative use of natural experiments, which are situations in which individuals are inadvertently “assigned” to different treatment groups either by a naturally occurring process or as the result of changes in policies implemented for entirely different purposes. Researchers can then compare the outcomes that the members of the various treatment groups experience so they can use the random variation generated by those assignments to identify causal effects.

For example, one of Card’s most influential studies looked at the impact of the minimum wage on employment using changes in minimum wage rates across U.S. states as a natural experiment.

Card also leveraged natural experiments to conduct extensive research on the economic impact of immigration. In 1980, he used the Mariel Boatlift in Miami to study the effect of a sudden influx of immigrants on wages and employment. He concluded that the arrival of Cuban refugees didn’t have a significant impact on the wages or employment of native-born workers in Miami.

Card also popularized the “difference in differences” statistical method for evaluating the effects of natural experiments.

Richard Thaler

Richard Thaler is an American economist who was awarded the Nobel Prize in 2017 for his foundational work in behavioral economics — in particular for the evidence he provided on how often people behave in ways that are neither rational nor economically optimal.

His research showed that people are influenced by a range of factors, including social norms, emotions, and cognitive biases, which can lead them to make choices that aren’t in their best interest.

Thaler is perhaps best known for his work on *nudges*, which are small changes in the environment that can influence people’s behavior toward more desirable outcomes without restricting their choices. For example, Thaler has shown that people are more likely to save money when given the option to enroll in a savings plan by default rather than having to actively sign up for it.

Thaler's work has informed policies in areas such as consumer finance, health insurance design, and environmental protection.

Robert Shiller

Robert Shiller is an American financial economist who was awarded a Nobel Prize in 2013 for providing overwhelming evidence that irrational buying and selling often propel asset prices away from their rational, fundamental values.

His research has reshaped the way economists and investors think about the behavior of financial markets and has led to new approaches to financial modeling and risk management.

If you pay attention to the financial markets, you've no doubt heard of the Case-Shiller Home Price Index, which Shiller created with his long-time research partner, Karl Case.

Ben Bernanke

Ben Bernanke is an American macroeconomist who was awarded the Nobel Prize in 2022 for his groundbreaking research about the Great Depression of the 1930s and, in particular, on how central banks like the Federal Reserve can regulate and manage the banking system so that it can continue functioning properly even if there's a major financial crisis.

The most amazing thing about Bernanke's career, however, is that he actually got to implement his ideas in person during a major, worldwide financial crisis. That's because he was serving as the chairman of the Board of Governors of the Federal Reserve when the 2007–2008 Global Financial Crisis occurred.

During his tenure, he oversaw a number of innovative monetary policy measures, including quantitative easing, the payment of interest on bank reserves, large-scale asset purchases, pushing short-term interest rates to zero, and acting aggressively as a lender of last resort throughout the financial system (including propping up the commercial paper market).

Glossary

This glossary contains common economics terms. Words set in *italic type* are terms that are defined separately in this glossary.

absolute advantage: When a person or nation can produce more output from a given amount of resource inputs than another person or nation can.

administered interest rates: One of the two *interest rates* that the Federal Reserve pays to, respectively, either banks or large nonbank financial firms to help control the economy's money supply. The two administered interest rates are the *interest rate on reserve balances*, or *IORB rate*, and the *over-night reverse repo rate (ON RRP rate)*.

aggregate demand: The total demand for goods and services in an economy.

aggregate supply: The total supply of goods and services in an economy.

allocatively efficient: When an economy's limited supply of resources is allocated to the production of the goods and services that consumers most greatly desire to consume.

anchoring: The tendency people have to anchor (base) the value that they assign an item on recently viewed, but logically irrelevant, information.

antitrust laws: Laws that regulate *monopolies* and *cartels*.

asset-price bubble: A situation in which the price of an asset rises above its true value; it's usually driven by speculative purchases financed with borrowed money.

asymmetric information: A situation in which either the buyer or the seller knows more about the quality of the good that they're negotiating over than the other party does.

behavioral economics: The branch of economic theory that deals with irrational decision-making and *systematic errors*.

capital: Either *physical capital* (e.g., machines, factories, and infrastructure used to produce output) or *intellectual capital* (i.e., knowledge about how to use physical capital and other resources such as land and labor to produce goods and services).

cartel: A group of firms that colludes and acts as a single coordinated whole to restrict output and drive up prices; formerly called trusts.

collateral: An asset pledged to guarantee the repayment of a loan if the borrower fails to make their contractually obligated loan payments on schedule.

command economy: An economy in which the government directs all economic activity.

comparative advantage: Occurs when a person or nation can produce a good or service at a lower opportunity cost (in terms of other goods and services that must be forgone) than another person or nation.

Consumer Price Index (CPI): The *market basket* (bundle of goods and services) that the Bureau of Labor Statistics used to measure changes in the prices of goods and services bought by a typical family of four.

consumer surplus: The benefit consumers get when they can buy something for less than the maximum amount that they're willing to pay for it.

deadweight loss: The amount by which *total surplus* (the sum of *consumer surplus* and *producer surplus*) is reduced whenever output is less than the *socially optimal output level*.

deflation: When the overall level of prices in the economy is falling.

demand: The full range of quantities that a person with a given income and preferences will demand at various possible prices.

demand curve: A line on a graph that represents how much of a good or service buyers will consume at various prices.

depreciation: A decrease in the economy's stock of *capital* caused by wear and tear or obsolescence (when an older machine or tool, despite being in good working order, is no longer wanted because new technologies have outdated it).

dictator game: A *behavioral economics* game in which one person, the dictator, gets to anonymously split a sum of money with a second player. Provides evidence of a human tendency for fairness because most dictators share at least some of the money with the other player.

digital platform: An internet website or software application that facilitates interactions between two or more distinct but interdependent sets of users.

diminishing marginal utility: A situation in which each additional, or marginal, unit of a good or service that you consume brings less *utility* (happiness) than the previous unit.

diminishing returns: A situation in which each additional unit of a resource used in a production process brings forth successively smaller amounts of output.

economic costs: Total costs, including money spent on production and *opportunity cost* (the value of the best alternative forgone).

economic profits: Any monies collected by a firm above and beyond what's required to keep an entrepreneur owner interested in continuing in business.

economics: The study of how people allocate scarce resources among alternative uses.

endowment effect: The tendency people have to put a higher dollar value on items they possess as compared with identical items that they don't possess. Explained by *prospect theory* and *loss aversion*.

entrepreneur: Individuals who supply *entrepreneurial ability* to firms.

entrepreneurial ability: The human resource, distinct from labor, which combines the other three *factors of production* (land, labor, and capital) to produce new products or make innovations in the production of existing products. Without entrepreneurial ability, we'd be stuck making the same things the same way, forever.

equilibrium: A situation in which every party involved in an economic interaction has no reason to change their current behavior.

externality: A cost or benefit that falls not on the person(s) directly involved in an activity but on others; externalities can be positive (benefits) or negative (costs).

factors of production: Inputs (resources) used to create goods and services, including land, labor, capital, and entrepreneurship.

financial markets: Markets in which either people trade the property rights to assets (such as real estate or stocks) or savers lend money to borrowers.

fiscal policy: A government's policy on taxes and spending; increased government spending or lower tax rates help to fight recessions.

fixed costs: Costs that have to be paid even if a firm isn't producing anything.

framing effects: Changes in decisions that are caused by making potential options that were once viewed as losses look like gains, and vice versa. Explained by *prospect theory*.

full employment: When every worker who wants a full-time job can obtain one.

full-employment output (Y^*): The quantity of output produced in the economy when the labor market has *full employment*.

gross domestic product (GDP): The value of all goods and services produced in the economy in a given period of time, usually a quarter or a year; the sum (denoted by the variable Y) of expenditures on consumption, investment, government purchases, and exports less expenditures on imports, or $Y = C + I + G + EX - IM$.

human capital: The knowledge, skills, experience, and attributes that individuals possess and that can help them produce output.

hyperinflation: A very high rate of *inflation*; it's defined by some authors as a rate of inflation in excess of 20 or 30 percent per month.

increasing returns: A situation in which each additional unit of a resource used in a production process brings forth successively larger amounts of output.

inflation: When the overall level of prices in the economy is rising.

inflation rate: A measure of how the overall level of prices in the economy changes over time; if the inflation rate is positive, prices are rising; if the inflation rate is negative, prices are falling.

insolvent: Being unable to honor one's financial obligations because they exceed the combined value of current assets and future income.

intellectual capital: Knowledge about how to use *physical capital* and other resources, such as land and labor, to produce goods and services.

interest rate: The price someone has to pay to borrow money.

interest rate on reserve balances, or IORB rate: The interest rate that the Federal Reserve pays banks on any money that they deposit into their respective accounts at the Federal Reserve.

investment: Spending to increase the economy's stock of *capital* as well as the value of any increases in inventories.

invisible hand: Adam Smith's famous idea that when firms are constrained by competition, each firm's greed causes it to act in a socially optimal way, as if guided to do the right thing by an invisible hand.

laissez faire: See *market economy*.

law of demand: The fact that for most goods and services, price and quantity demanded have an inverse relationship. (As one goes up, the other goes down.)

long-run shutdown condition: A situation in which a firm's total revenues exceed its *variable costs* but are less than its total costs; the firm will operate until its *fixed cost* contracts expire (in the long run).

loss aversion: The tendency people have to avoid potential losses relative to their current *status quo* situation. Explained by *prospect theory*.

macroeconomics: The study of the economy as whole, concentrating on economy-wide factors such as interest rates, inflation, and unemployment. Macroeconomics also encompasses the study of economic growth and how governments use monetary and fiscal policy to moderate the harm that recessions cause.

marginal cost: How much total costs increase when you produce one more unit of output.

marginal utility: The change in total *utility* (happiness) that results from consuming the next unit of a good or service. Marginal utility can be positive or negative.

market basket: A bundle of goods and services selected to measure inflation. Economists define a market basket, such as the *Consumer Price Index*, and then track how much money it takes to buy this basket from one period to the next.

market economy: An economy in which the private decisions of consumers, resource suppliers, and firms determine how resources are allocated, with only limited interventions from the government; often referred to as a *laissez-faire* ("to leave alone") economic system.

market failures: Situations in which markets deliver nonoptimal social outcomes. Common causes of market failure are *asymmetric information*, *monopoly*, *externalities*, and *public goods*.

microeconomics: The part of economics that studies individual people and individual businesses. For people, microeconomics studies how they behave when faced with decisions about where to spend their limited budgets. For businesses, it studies how profit-maximizing firms behave individually and how they behave when competing against each other in markets.

monetary policy: Using changes in the money supply to change interest rates and stimulate or slow down economic activity.

monopolistic competition: A situation in which many firms with slightly different products compete. Production costs are above what could be achieved by perfectly competitive firms, but society benefits from the product differentiation.

monopoly: A firm that has no competitors in its industry. It produces less output, has higher costs, generates a *deadweight loss*, and sells its output for a higher price than it would if constrained by competition. These negative outcomes usually generate government regulation.

myopia: The human tendency to see near-term costs and benefits clearly while giving relatively little thought to costs and benefits that lie in the future. Leads to decisions that favor present options over future ones.

natural monopoly: An industry in which one large producer can produce output at a lower cost than many small producers. It undersells its rivals and ends up as the only firm surviving in its industry.

network: A group of people or objects that are connected to each other by a service or piece of infrastructure that facilitates flows of goods, services, or information.

network effect: The phenomenon whereby the value of a *network* depends on the size of the network (number of users).

nominal interest rates: Interest rates that measure the returns to a loan in terms of money borrowed and money returned (as opposed to *real interest rates*).

nominal prices: The price as it's actually observed in current dollars; contrasts with *real prices*, which adjust for inflation.

nominal wages: Wages measured in current dollars; contrasts with *real wages*, which adjust for *inflation*.

oligopoly: An industry with only a few firms. If they collude, they form a *cartel* to reduce output and drive up profits the way a *monopoly* does.

opportunity cost: The value of the best forgone alternative option; what you give up to pursue a particular option.

over-night reverse repo rate (ON RRP rate): The interest rate that the Federal Reserve pays to large nonbank financial companies (such as money-market mutual funds) on any money that they deposit into their respective accounts at the Federal Reserve.

perfect competition: A situation in which numerous small firms producing identical products compete against each other in a given industry. Perfect competition leads to firms producing the *socially optimal output level* at the minimum possible cost per unit.

physical capital: Machines, factories, and infrastructure used to produce output.

price ceiling: A market intervention in which the government ensures that the price of a good or service stays below the free-market price.

price floor: A market intervention in which the government keeps the price of a good or service above its free-market price.

prisoner's dilemma: A situation in which a pair of prisoners (or firms) has to decide whether to cooperate. The dilemma is that although the individual incentives favor not cooperating, both players would be better off if they could figure out a way to cooperate.

producer surplus: The gain that producers receive when they can sell their output at a price higher than the minimum amount for which they're willing to make it.

product differentiation: A competitive strategy in which one firm's product is distinguished from competing products by means of its design, related services, quality, location, or other attributes (except price).

production possibilities frontier (PPF): A graph economists use to help them visualize the tradeoffs you make when you efficiently reallocate inputs from producing one thing to producing another; sometimes referred to as the production possibilities curve.

productively efficient: Producing a given good or service at the lowest possible cost.

prospect theory: A *behavioral economics* theory of decision-making that has three main features: (1) people evaluate potential outcomes relative to the *status quo* that they're used to; (2) gains are subject to *diminishing marginal utility*, whereas losses are subject to diminishing marginal disutility; and (3) people feel losses 2.5 times more intensely than gains.

public goods: Goods or services that private firms can't profitably produce because the goods or services are impossible to provide to just one person. If you provide them to one person, you have to provide them to everybody. Because all consumers hope somebody else will pay for public goods so they can get them for free, usually nobody ends up paying.

quantity demanded: How much of a good or service a consumer will demand at a specific price, given their income and preferences.

quantity theory of money: The theory that the overall level of prices in the economy is proportional to the quantity of money circulating in the economy.

rational expectations: The theory that people will optimally change their behavior in response to policy changes. Depending on the situation, their behavioral changes can greatly limit the effectiveness of policy changes.

rationality: Behaviors and decisions that maximize a person's likelihood of achieving their goals.

real interest rates: Interest rates that compensate for inflation by measuring the returns to a loan in terms of units of stuff lent and units of stuff returned (as opposed to *nominal interest rates*).

real prices: How much of one kind of thing (such as hours worked) you have to give up to get a good or service, no matter what happens to *nominal prices*.

real wages: Nominal wages adjusted for inflation; wages measured not in terms of current dollars (as *nominal wages* are) but rather in terms of how much output can be purchased with those current dollars.

recession: A period of time during which an economy's total output falls.

recovery: The period after a *recession* ends and during which an economy's total output expands. This period continues until the next *recession* begins.

scarcity: The fact that people don't have enough resources to satisfy all their wants; the phenomenon that creates the need for *economics*.

short-run shutdown condition: A situation in which a firm's total revenues are less than its *variable costs* and the firm is better off shutting down immediately and losing only its *fixed costs*.

socially optimal output level: The output level that maximizes the benefits that society can get from its limited supply of resources.

status quo: In *prospect theory*, the current situation against which people evaluate potential gains and losses.

status quo bias: The tendency people have to stick with default (status quo) situations rather than make changes. Explained by *prospect theory* and *loss aversion*.

sticky prices: Prices that are slow to adjust to shocks. Price stickiness can cause recessions to linger.

supply and demand: An economic model of markets that separates buyers from sellers and then summarizes each group's behavior with a single line on a graph. The buyers' behavior is captured by the *demand curve*, and the sellers' behavior is captured by the *supply curve*. By putting these two curves on the same graph, economists can show how buyers and sellers interact in markets to determine how much of any particular item will be sold, as well as the price at which it will be sold.

supply curve: A line on a graph that represents how much of a good or service sellers will produce at various prices.

systematic errors: Suboptimal choices that are made again and again as if a person can't learn from their mistakes.

time inconsistency: The human tendency to systematically misjudge at the present time what the future self will want to do. Caused by *myopia*.

total surplus: The sum of *producer surplus* and *consumer surplus*.

Tragedy of the Commons: The idea that if a resource is open to public use, it typically becomes rapidly exhausted or ruined because each person's personal incentive is to use it up before anyone else can. This problem can be solved by private property rights, which give owners an incentive to conserve the resource and harvest it at sustainable rates.

ultimatum game: A *behavioral economics* game involving two players in which one gets to propose how to split a sum of money and the other gets to accept or reject the proposal. If rejected, neither player gets anything. The game provides evidence that the potential for rejection enhances fair behavior because we observe that the splits offered in the ultimatum game are more generous and more equal than the splits made by dictators in the *dictator game*.

utility: A measure of happiness that economists suppose people use to compare all possible things that they may experience.

variable costs: Costs that vary with the amount of output produced.

wages: The prices paid for the use or services of labor per unit of time.

wealth: Anything that has value because it already produces or can produce a flow of desirable goods and services.

Index

A

absolute advantage, 397
abstracting, 17
accounting profits, economic profits compared with, 110–111
actual expenditures, adjusting with fluctuating inventories, 321–323
administered interest rates, 352–353, 397
adverse selection, avoiding, 217–218
advertising
 about, 221
 framing effects and, 253–254
 selling, 230
aggregate demand
 adjusting to shift in, 310–311
 defined, 397
 model for, 303
aggregate demand curve (AD), 308–309, 332
aggregate supply
 defined, 397
 model for, 303
Akerlof, George (economist), 393–394
all-in fallacy, 374–375
allocatively efficient, 13, 38, 397
altruism, 26
Amazon.com, 235
American Telephone and Telegraph Corporation (AT&T), 176
anchoring, 254–255, 397
antitrust laws, 189, 397
area-based property rights, 208
Arrow, Kenneth (economist), 386
asset-price bubble, 397
assets
 considerations for, 280–282
 streams of, 268–270
asymmetric information

 about, 14, 209–210
 defined, 397
 insurance, 215–219
 lemons problem, 211–215
 trade limits with, 210–211
autonomous expenditures, 324
availability heuristic, 251
average costs
 equaling marginal cost, 117–119
 marginal costs compared with, 34–35
 pricing, 175–176
average fixed costs (AFC), 114–115
average total costs (ATC), 115–116
average variable costs (AVC), 113–114, 115–116

B

Baidu, 236
balanced budget, 277
banking system
 bank collapses, 364–365
 broken, 366–367
bartering, 286
Becker, Gary (economist), 388
behavioral economics
 about, 14–15
 defined, 397
 gauging fairness and self-interest, 258–262
 human brain and, 248–251
 myopia, 256–257, 258
 need for, 245–246
 neo-classical economics and, 246–247
 Prospect theory, 251–256
 time inconsistency, 257–258
Ben & Jerry's, 106
benefits
 costs exceeding, 201–202
 using supply and demand to compare costs and, 132–133

Bernanke, Ben
 (macroeconomist), 396
Bing, 236
bonds, 348–349
borrowing, 361, 362
bubbles
 bursting, 363–366
 developing, 360–363
budgets
 balanced, 277
 budget deficit, 277
 budget surplus, 277
 decision-making under limited, 92–98
Bureau of Economic Analysis, 266
Bureau of Labor Statistics, 296, 300
business cycle, 304–305

C

capital
 defined, 397
 as a resource, 39
capitalism, 57
Card, David (economist), 395
cardinal utility, 88
cartels
 agreements with, 181–187
 behavior of, 179–180
 criteria for coordinating, 180
 defined, 397
cheap-foreign-wages fallacy, 375–376
Cheat Sheet (website), 4
choice models, 24
circular flow diagram, 270–271
Coca-Cola, 179
cognitive biases, 250–251
collateral, 361, 398
collude, compete compared with, 179–180

- command economy
 - about, 53
 - defined, 398
 - failure of, 54–55
- Communism, 194
- comparative advantage, 282–284, 398
- competition. *See also* monopolistic competition
 - about, 107
 - accounting profits compared with economic profits, 110–111
 - collude compared with, 179–180
 - microeconomics and, 13
 - monopolies compared with competitive firms, 166–169
 - problems caused by lack of, 13
 - requirements for perfect, 107–108
 - setting quantities, 108–110
 - taking prices, 108–110
 - at zero marginal cost, 231
- competitive free markets
 - about, 129–130
 - analyzing efficiency of, 132–134
 - causes and consequences of perfect competition, 146–147
 - graphing how profits guide firm entry and exit, 149–153
 - managing deadweight losses, 141–146
 - measuring gains with total surplus, 135–141
 - process of perfect competition, 147–148
 - traits of properly functioning markets, 131–132
- competitors, reducing redundant, 170–171
- complementary goods, 103–104
- complexity, as a factor in political and social conflicts, 48
- composition, fallacy of, 374
- concentration, 243
- confirmation bias, 250
- constraints
 - resource, 27–28
 - technology, 28
 - time, 28

- consumer behavior
 - about, 23
 - choice models and, 24
 - cost-benefit model, 30–31
 - limitations, 27–29
 - limitations of choice models, 32–35
 - pursuing personal happiness, 25–27
 - violations of choice models, 32–35
- Consumer Price Index (CPI), 295–296, 398
- consumer surplus
 - about, 240–241
 - defined, 135, 398
 - measuring for continuous goods, 137–138
 - measuring for discrete goods, 135–137
- consumption, in gross domestic product (GDP), 273, 274–275
- consumption function, 323
- consumption spending, 351
- contestable industries, 242–243
- continuous goods, measuring consumer surplus for, 137–138
- Corn Laws, 282
- cost structure, analyzing for firms, 111–119
- cost-benefit model, 30–31
- costs
 - average, 34–35, 117–119, 175–176
 - average fixed costs (AFC), 114–115
 - average total costs (ATC), 115–116
 - average variable costs (AVC), 113–114, 115–116
 - changes in, 72–73
 - exceeding benefits, 201–202
 - marginal, 34–35, 116–125
 - natural monopolies and, 171
 - opportunity cost, 29, 66–67
 - per unit of output, 112–113
 - sunk, 33
 - using supply and demand to compare benefits and, 132–133
 - of wages and profits, 316
- coupon payments, 348
- CPI (Consumer Price Index), 295–296, 398

- credit card bills, anchoring and, 254–255
- Croesus, King of Lydia, 288
- cross-price effects, 102

D

- data
 - plotting, 18–20
 - selling, 230
- deadweight losses
 - about, 141
 - defined, 398
 - monopolies and, 168
 - from price ceilings, 141–142
 - of taxes, 142–146
- Debreu, Gerard (economist), 386
- debt-driven bubbles
 - bursting, 363–366
 - developing, 360–363
- decision-making
 - about, 87–88
 - deriving demand curves from diminishing marginal utility, 99–104
 - diminishing marginal utility, 89–91
 - with limited budgets, 92–98
 - by rankings, 88–89
 - uninformed, 32–33
- deficits
 - about, 342
 - security of future tax revenues, 342–343
- deflation, 398
- deleveraging, 363–364
- demand
 - about, 62–63
 - adjusting to shift in aggregate, 310–311
 - aggregate, 303, 310–311, 397
 - defined, 398
 - downward sloping, 190–192
 - elasticity, 67–69
 - excess, 78
 - facing down, 157–158
 - factors affecting quantity demanded, 63

- graphing demand curve, 64–66
 - for money, 287–288
 - opportunity costs, 66–67
 - reacting to increases in, 79–80
 - stimulating to end recessions, 330–332
- demand curve
 - about, 18
 - aggregate, 308–309, 332
 - creating, 18–20
 - defined, 398
 - deriving from diminishing marginal utility, 99–104
 - deriving marginal revenue from, 158–160
 - drawing a, 20–21
 - graphing, 64–66
 - market, 108
 - setting slope of, 66–67
 - sliding down the, 161–162
 - using for making predictions, 20
- democracy
 - about, 57, 380
 - economic growth and, 9
- depreciation, 276, 398
- dictator game, 260, 398
- digital economy, regulating, 240–243
- digital platforms
 - about, 235–236
 - balancing user engagement against need to charge fees, 236–237
 - defined, 398
 - parsing profit maximization, 237–240
- diminishing marginal utility
 - about, 30–31, 89–91
 - defined, 398
 - deriving demand curves from, 99–104
- diminishing returns
 - about, 38, 40–41
 - defined, 398
 - determining cause of, 113
- discrete goods, measuring consumer surplus for, 135–137
- disposable income, 274–275

- dollar amount, percentage compared with, 34
- dominant firms, breaking up, 188
- downward wage stickiness, 337–338
- Duflo, Esther (economist), 394

E

- economic costs, 398
- economic fallacies, 371–377
- economic growth, innovation and, 380
- economic profits
 - accounting profits compared with, 110–111
 - defined, 147, 398
- economic systems, political systems compared with, 57
- economics. *See also* behavioral economics; Internet economics; macroeconomics; microeconomics
 - about, 7
 - defined, 398
 - future of, 10
 - history of, 8–10
 - macroeconomics, 11–17
 - microeconomics, 11–17
 - models and graphs in, 17–21
 - as science of scarcity, 11
- economies of scale, among Internet businesses, 226–231
- economists, 383–389
- economy
 - command, 53, 54–55, 398
 - effect of international trade on, 279–284
 - lowering interest rates to stimulate the, 351–352
 - market, 53, 54–55, 401
 - measuring, 15
 - mixed, 53–55
- education. *See* literacy and education
- efficiency
 - of competitive free markets, 132–134
 - losing, 169
 - measuring, 44–45

- elasticity
 - demand and, 67–69
 - economies of scale and, 228–231
 - importance of size of, 239–240
 - supply and, 73–75
- employment, inflation and, 337
- endowment effect, 255, 399
- entrepreneur, 399
- entrepreneurial ability
 - defined, 399
 - as a resource, 39
- environmental problems, property rights and, 381
- equilibrium
 - defined, 399
 - finding, 192–193
 - market, 75–85
 - showing graphically, 325–327
- equilibrium price level (P^*), 309, 318–328
- excess demand, 78
- excess supply, 77–78
- expenditures
 - actual, 321–323
 - adjusting with fluctuating inventories, 321–323
 - autonomous, 324
 - government, 273, 277–278, 340–342
 - planned, 321–323
- expert opinions, 215
- exponential scaling, of network effects, 234–235
- exports (EX), 278–279
- externalities
 - about, 199–200
 - accepting positive amounts of negative, 202–203
 - consequences of positive, 204–205
 - defined, 399
 - effects of negative, 200–202
 - negative, 200–204
 - positive, 200, 204–205, 206
 - subsidizing things that provide positive, 206
- extinctions, 207–208

F

face value payment, 348
Facebook, 231, 233, 234–235, 236, 239–240, 243
factors of production
 defined, 399
 markets for, 270
fairness
 gauging, 258–262
 ignoring, 50–51
fallacies, economic, 371–377
fallacy of composition, 374
FC. *See* fixed costs (FC)
Federal Reserve (the Fed), 350–351, 352–353
fees, balancing user engagement against need to charge, 236–237
fiat money
 about, 290
 gold standard compared with, 345–346
fiber-optic internet access monopolies, 170
financial crises
 about, 359–360
 developing debt-driven bubbles, 360–363
 origins and effects of, 359–368
 recovery, 366–368
 seeing bubble burst, 363–366
financial markets, 270, 399
fiscal policy
 about, 16, 329–330, 340
 defined, 399
 fighting recessions with, 329–357
 increasing government spending to help end recessions, 340–342
 limits of, 368
Fisher, Irving (economist), 301
Fisher equation, 301–302
fixed costs (FC)
 about, 111
 average, 114–115
 defined, 399
 economies of scale and, 226–227

fixed prices, in short run, 311–313
flow, 270–271
framing effects, 251, 253–254, 399
free markets, regulating, 380
freedom, 380
freemium pricing, 230
frictional unemployment, 306
Friedman, Milton (economist), 386–387
full and complete property rights, 199
full employment, 399
full-employment output (Y^*)
 about, 305–306
 aiming for, 330–332
 defined, 399
 increasing output beyond, 333–334
 moving back to, 335–336
 natural result of price adjustments, 306–307
 with/without government intervention, 317
funding research, governments for, 56

G

gains, measuring with total surplus, 135–141
game theory, 181
gaze heuristic, 248–249
GDP. *See* gross domestic product (GDP)
gold standard
 about, 290
 fiat money compared with, 345–346
goods
 complementary compared with substitute, 103–104
 markets for, 270
Google, 226, 230, 231, 236, 241–242
government expenditures
 in gross domestic product (GDP), 273, 277–278
 increasing to help end recessions, 340–342
government intervention
 case against, 52–53

case for, 51–52
comparing, 47–48
deciding amount of, 55
full-employment output (Y^*) with/without, 317
government policy, limits of, 368
graphing
 demand curve, 64–66
 economist use of, 17–21
 interpreting shapes of, 44
 production possibilities, 42–45
 supply curve, 69–73
Great Depression, 318–319
Great Recession, quantitative easing and, 356–357
gross domestic product (GDP)
 about, 266
 boosting in Keynesian model, 327–328
 circular flow diagram, 270–271
 defined, 399
 equation for, 273–279
 newly produced output, 271–272
 rising, 272
 streams of income and assets, 268–270
 what it doesn't include, 267
 what it includes, 267–268
group discrimination, 218
Gutenberg, Johannes, 222–223

H

happiness
 maximizing, 37–57
 pursuing personal, 25–27
Hershey, 252–253
heuristics, 248–249
hindsight bias, 251
history, of economics, 8–10
Hitler, Adolf, 291
human brain, behavioral economics and, 248–251
human capital
 clarifying, 39–40
 defined, 399
hyperinflation, 285, 289–291, 399

I

- icons, explained, 3–4
- impulse buying, 247
- in-app purchases, 231
- income
 - as a factor affecting quantity demanded, 63
 - streams of, 268–270
- increasing returns, 399
- individual identification, 218
- inequality, as a factor in political and social conflicts, 48
- inflation
 - about, 285–286
 - buying an, 286–295
 - calculating rate of, 297
 - defined, 399
 - effects of, 293–295
 - employment and, 337
 - generating, 332–339
 - measuring, 295–300
 - nominal interest rates, 301–302
 - preventing, 382
 - real interest rates, 301–302
- inflation rate, 400
- inflationary expectations, 355–356
- innovation
 - economic growth and, 380
 - intellectual property rights and, 381
 - promoting, 56–57
- An Inquiry into the Nature and Causes of the Wealth of Nations* (Smith), 26
- insolvent, 365, 400
- Instagram, 236
- insurance, asymmetric information and, 215–219
- intellectual capital
 - defined, 400
 - as a resource, 39
- intellectual property rights
 - about, 381
 - government protection of, 56
- interest, in flow of money, 269

- interest rate on reserve balances (IORB) rate, 352–353, 400
- interest rates
 - administered, 352–353
 - changing money supply to change, 350–351
 - defined, 400
 - effect of inflationary expectations on, 355
 - lowering to stimulate the economy, 351–352
 - nominal, 301–302
 - prices of bonds and, 349
 - real, 301–302
- international trade
 - about, 15, 381
 - effect on economy of, 279–284
- Internet economics
 - about, 14, 225–226
 - digital platforms, 235–240
 - economics of scale, 226–231
 - network effects, 232–235
 - regulating digital economy, 240–243
- inventories
 - adjusting, 319–327
 - target levels for, 320–321
- inverse relationship, 18, 64, 308–309
- investment
 - defined, 400
 - in gross domestic product (GDP), 273, 275–277
- investment spending, 351
- invisible hand, 13, 400
- IORB (interest rate on reserve balances) rate, 352–353, 400
- irrationality, 33–35

K

- Kahneman, Daniel (psychologist), 392
- Keynes, John Maynard (economist), 318–319, 386
- Keynesianism, 319
- Keynes's model, 16
- Khan, Kublai, 288
- Krugman, Paul (economist), 394

L

- labor, as a resource, 39
- laissez faire economy. *See* market economy
- laissez-faire policies, 319
- land, as a resource, 39
- law of demand, 400
- lemons problem, 211–215
- lending standards, 362
- Licence Raj (India), 49–50
- limitations
 - about, 27–29
 - of choice models, 32–35
 - of resources, 38
- limited liability corporations (LLCs), economic growth and, 9
- literacy and education
 - economic growth and, 9–10
 - government encouraging, 56
 - living standards and, 380
- living standards
 - education and, 380
 - raising, 9–10
- local collective management, 208
- long-run aggregate supply curve (LRAS), 309, 312, 313
- long-run effects, of recessions, 307–314
- long-run shutdown condition, 128, 400
- loss aversion, 252–253, 400
- losses. *See also* deadweight losses
 - firm exit and, 152–153
 - visualizing, 124–125
- Lucas, Robert (economist), 388–389
- lump-of-labor fallacy, 371–372
- Lyft, 236

M

- macroeconomics
 - about, 11–12, 15
 - defined, 400
 - international trade, 15
 - measuring economy, 15
 - recessions, 16–17

- macroeconomy
 - about, 265–266
 - gross domestic product (GDP), 266–272
 - gross domestic product (GDP) equation, 273–279
 - international trade, 279–284
- Malthus, Thomas, 372
- marginal costs (MC)
 - about, 116
 - average costs compared with, 34–35
 - comparing with marginal revenues, 119–125
 - defined, 400
 - economies of scale and, 227–228
 - equaling average cost, 117–119
 - pricing, 173–175
 - setting monopoly's marginal revenue equal to, 162–164
 - zero, 231
- marginal propensity to consume (MPC), 275
- marginal revenues (MR)
 - comparing with marginal costs, 119–125
 - decreasing, 157–162
 - deriving from demand curve, 158–160
 - relating to total revenues, 160–161
 - setting equal to marginal costs for monopolies, 162–164
- marginal utility
 - about, 30–31
 - defined, 90, 400
 - equal, per dollar, 97–98
 - maximizing, 92–94
- marginalism, 374
- market basket
 - about, 295
 - creating your own, 296
 - defined, 400
- market demand curve, 108
- market economy, 53, 54–55, 401
- market equilibrium
 - about, 75–76
 - adjusting to new, 79–81
 - price controls and, 81–85
 - stability of, 77–78
- market failures
 - about, 130
 - common, 14
 - defined, 401
- market power, enhancing with network effects, 233–234
- market production, 48–50
- market results, comparing, 47–48
- market supply curve, 108
- market-instability fallacy, 375
- markets. *See also* competitive free markets
 - for factors of production, 270
 - financial, 270
 - for goods and services, 270
- Marshall, Alfred (economist), 385
- Marx, Karl (economist), 384–385
- Mastercard, 236
- maximum output, 306
- MC. *See* marginal costs (MC)
- McFadden, Dan (microeconomist), 393
- medium of exchange, 294
- metallic standard, 345
- microeconomics
 - about, 11–12
 - balancing supply and demand, 12
 - behavioral economics, 14–15
 - common market failures, 14
 - competition and, 13
 - defined, 401
 - economics of the Internet, 14
 - networks, 14
 - platforms, 14
 - problems caused by lack of competition, 13
 - reforming property rights, 13–14
 - short run compared with long run in, 126
- Microsoft, 241–242
- minimum-cost production, 151–152
- mixed economy, 53–55
- models, economist use of, 17–21
- modularity, brain, 249–250
- monetary policy
 - about, 16, 292–293, 329–330, 344
 - Administered Rates, 352–353
 - bond prices in relation to interest rates, 349
 - bonds, 348–349
 - changing money supply to change interest rates, 350–351
 - defined, 401
 - fiat money compared with gold standard, 345–346
 - fighting recessions with, 329–357
 - having too much money, 346–347
 - limiting monetary policy with rational expectations, 353–356
 - limits of, 368
 - lowering interest rates to stimulate economy, 351–352
 - quantitative easing and Great Recession, 356–357
- money
 - having too much, 346–347
 - printing, 343–344
- money demand, balancing money supply and, 287–288
- money supply
 - balancing money demand and, 287–288
 - changing to change interest rates, 350–351
 - graphing results of increases in, 353–354
- monopolies
 - about, 155–156
 - benefits of, 169–171
 - breaking up, 176
 - competitive firms compared with, 166–169
 - defined, 107, 401
 - natural, 171, 226, 401
 - natural gas, 170–171
 - profit-maximizing, 156–166
 - regulating, 172–176
 - subsidizing, 172
- monopolistic competition
 - about, 107, 177–178
 - defined, 401
 - product differentiation, 189–190, 194
 - profit limits, 190–194
- Monster.com, 235
- moral hazard, mitigating, 218–219

morals, ignoring, 50–51
MPC (marginal propensity to consume), 275
MR. *See* marginal revenues (MR)
multihoming, 243
myopia, 256–257, 258, 401

N

National Income and Product Accounts (NIPA), 265
natural gas monopolies, 170–171
natural monopolies
costs and, 171
defined, 401
economies of scale and, 226
negative demand shock, 310, 331
negative externalities
about, 200
accepting positive amounts of, 202–203
effects of, 200–202
managing, 203–204
negative network effects, 232–233
neo-classical economics, behavioral economics and, 246–247
net exports, in gross domestic product (GDP), 273, 278–279
network effects
about, 229–230, 232–233
defined, 401
enhancing market power with, 233–234
exponential scaling of, 234–235
networks, 14, 401
NIPA (National Income and Product Accounts), 265
Nixon, Richard (president), 290
Nobel Laureates, 391–396
nominal interest rates, 301–302, 401
nominal prices, 299, 401
nominal wages
defined, 401
raising while prices are stuck, 335
nonexcludable public goods, 219
non-production
about, 125
short run compared with long run in microeconomics, 126

total costs exceeding total revenues, 128
variable costs exceeding total revenues, 126–128
nonrival public goods, 219

O

oligopolies
about, 177–178
cartel agreements, 181–187
cartel behavior, 179–180
criteria for coordinating cartels, 180
defined, 107, 401
prisoner's dilemma, 181–187
regulating oligopolies, 188–189
sharing power over prices, 179
ON RRP (over-night reverse repo rate), 353, 402
OPEC, 185–187
open-market operations, 351
opportunity cost
about, 29
defined, 402
demand and, 66–67
ordinal utility, 89
Ostrom, Elinor (political scientist), 392
output
adjusting with fluctuating inventories, 321
choosing level to maximize profits, 162–166
costs per unit of, 112–113
full-employment, 305–306
increasing beyond Y^* , 333–334
maximum, 306
minimum requirements for, 172–173
newly produced, 271–272
price levels and, 166–168
quotas for, 180
subsidizing monopolies to increase, 172
overconfidence effect, 251
over-night reverse repo rate (ON RRP rate), 353, 402

overpopulation fallacy, 372
overproduction, 201–202

P

patents
economic growth and, 9
encouraging innovation and investment with, 170
to turn public goods into private goods, 222–223
PepsiCo, 179
percentage, dollar amount compared with, 34
perfect competition
causes and consequences of, 146–147
defined, 107, 402
process of, 147–148
requirements for, 107–108
traits of, 146–153
perfectly elastic demand curve, 68–69
perfectly elastic supply curve, 73–75
perfectly inelastic demand curve, 68
perfectly inelastic supply curve, 73–75
permits, 208
philanthropy, to provide public goods, 220–221
physical capital
defined, 402
as a resource, 39
planned expenditures, adjusting with fluctuating inventories, 321–323
planning fallacy, 251
platforms, 14
plotting data, 18–20
political systems, economic systems compared with, 57
politics, of inflation, 292
positive demand shock, 310
positive externalities
about, 200
calculating consequences of, 204–205
subsidizing things that provide, 206
positive network effects, 232–233

- PPF (production possibilities frontier), 42–45, 47, 402
 - precommitments, self-control and, 257–258
 - predictions
 - perfection of, 302
 - using demand curve to make, 20
 - preferences, as a factor affecting quantity demanded, 63
 - price ceilings
 - about, 82–83
 - deadweight losses from, 141–142
 - defined, 402
 - price changes
 - effect on quantities demanded of, 99–101
 - graphing, 101–103
 - supply curve and, 71–72
 - price controls
 - about, 81–82
 - price ceilings, 82–83
 - price floors, 83–85
 - price discrimination, employing, 230–231
 - price floors, 83–85, 402
 - price index/price level index
 - determining real standard of living with, 299
 - identifying problems with, 299–300
 - setting up, 297–299
 - price systems, 49
 - prices. *See also* sticky prices
 - adjusting inventories instead of, 319–327
 - average cost, 175–176
 - economies of scale and, 228–231
 - as a factor affecting quantity demanded, 63
 - long-run adjustments, 309–310
 - marginal cost, 173–175
 - natural result of adjustments in, 306–307
 - output and, 166–168
 - raising for one group of users, 237–239
 - raising nominal wages while prices are stuck, 335
 - raising until they reach equilibrium, 78
 - reducing until they reach equilibrium, 77–78
 - regulating for monopolies, 173–176
 - sharing power over, 179
 - taking, 108–110
 - at zero marginal cost, 231
 - printing money, 343–344
 - prisoner's dilemma, 181–187, 402
 - private costs, social costs compared with, 200–201
 - private goods, selling to provide public goods, 222
 - producer surplus
 - defined, 135, 402
 - measuring, 138–139
 - product differentiation
 - about, 194, 243
 - benefits of, 189–190
 - defined, 402
 - production
 - about, 37–38
 - evaluating types of output, 38–46
 - sales price and costs of, 71
 - what to produce, 46–55
 - production possibilities frontier (PPF), 42–45, 47, 402
 - productively efficient, 13, 38, 402
 - profit limits, 190–194, 362
 - profit-maximizing firms
 - about, 105–106
 - analyzing cost structure of, 111–119
 - comparing marginal revenues with marginal costs, 119–125
 - competition, 107–111
 - maximizing profits, 106–107
 - non-production, 125–128
 - profit-maximizing monopolies
 - choosing output levels for, 162–166
 - decreasing marginal revenues, 157–162
 - problems caused by, 156–157
 - profits
 - accounting compared with economic, 110–111
 - cost of, 316
 - economies of scale and, 228–231
 - firm entry and, 149–151
 - in flow of money, 269
 - graphing, 149–153
 - maximizing, 106–107
 - parsing maximization, 237–240
 - visualizing, 122–124
 - zero, 151–152
 - property rights
 - about, 197–198
 - environmental problems and, 381
 - externalities, 199–206
 - poor, 207–208
 - reforming, 13–14
 - socially optimal outcomes, 198–199
 - tragedy of the commons, 206–208
 - prospect theory, 251–256, 402
 - protectionism fallacy, 373–374
 - public goods
 - about, 14, 209–210, 219
 - defined, 402
 - government production of, 382
 - philanthropy to provide, 220–221
 - ranking new technology as, 222–223
 - selling related private goods to provide, 222
 - taxing to provide, 220
- ## Q
- quantitative easing (QE), Great Recession and, 356–357
 - quantities
 - graphing changes, 101–103
 - setting, 108–110
 - quantity demanded, 402
 - quantity theory of money, 288, 402
- ## R
- ranking, decision-making by, 88–89
 - rational expectations, 16, 403
 - rationality, 246, 403
 - real cost, 299
 - real interest rates, 301–302, 403
 - real prices, 403
 - real wages

- defined, 403
- moving back to, 335–336
- tracing movement of, 334–336
- recessions
 - about, 16–17, 303–304
 - business cycle, 304–305
 - defined, 403
 - fighting with monetary and fiscal policy, 329–357
 - full-employment output, 305–306
 - increasing government spending to help end, 340–342
 - Keynesian Model, 308–328
 - leading into, 365–366
 - long-run effects, 307–314
 - natural result of price adjustments, 306–307
 - short-run effects, 307–314
 - sticky prices, 305–328
 - stimulating demand to end, 330–332
- recognition heuristic, 249
- recovery, 305, 366–368, 403
- regulating
 - digital economy, 240–243
 - free markets, 380
 - monopolies, 172–176
 - oligopolies, 188–189
- Remember icon, 3
- rent, in flow of money, 269
- reputation, building a, 214
- research
 - government funding for, 56
 - subsidizing, 223
- resources
 - allocating, 41
 - classifying, 39
 - constraints for, 27–28
 - directing automatically, 48–50
 - overexploiting commonly owned, 206–208
- returns, increasing compared with decreasing, 112–113
- Ricardo, David (economist), 282–283, 384
- Robinson, Joan (economist), 190, 330
- rolling over the debt, 343
- running a profit, 110

S

- sales price, production costs and, 71
- Samuelson, Paul (economist), 387
- scarcity, 403
- science of scarcity, economics as a, 11
- self-control, precommitments and, 257–258
- self-interest
 - about, 26–27, 379
 - gauging, 258–262
 - “selfless” actions, 25–26
- self-serving bias, 250
- Sen, Amartya (economist), 391–392
- sequence fallacy, 373
- services, markets for, 270
- Shiller, Robert (economist), 396
- short run
 - effects of recessions in, 307–314
 - fixed prices in, 311–313
- short-run aggregate supply curve (SRAS), 311–312, 313
- short-run shutdown condition, 126–128, 403
- shrinking package, 252–253
- side effects, as a factor in political and social conflicts, 48
- Smith, Adam (economist and philosopher), 13, 134, 380, 383
 - An Inquiry into the Nature and Causes of the Wealth of Nations*, 26
 - The Theory of Moral Sentiments*, 259, 261
 - The Wealth of Nations*, 258
- Smith, Vernon (economist), 393
- social costs, private costs compared with, 200–201
- socially optimal output level
 - about, 130
 - defined, 403
 - determining, 133–134
 - property rights and, 198–199
- Solow, Robert (economist), 387–388
- Sony PlayStation, 236–237
- Spotify, 236
- SRAS (short-run aggregate supply curve), 311–312, 313

- stability, of market equilibrium, 77–78
- standard of deferred payment, 292, 294
- standard of living, determining with price index, 299
- Standard Oil Company, 188
- status quo bias, 255–256, 403
- steering heuristic, 249
- sticky prices
 - about, 315
 - adding costs of wages and profits, 316
 - cutting wages or workers, 315–316
 - defined, 403
 - importance of, 336–337
 - Keynesian model, 318–328
 - Y* with/without government intervention, 317
- stimulation, risk of too much, 332–339
- stimulus, 336–339
- structural mismatches, 366, 367–368
- subsidizing monopolies, 172
- substitute goods, 103–104
- subtractive versioning, 231
- sugar subsidy, 52–53
- sunk costs, 33
- supply
 - about, 69
 - aggregate, 303, 397
 - elasticity, 73–75
 - excess, 77–78
 - graphing supply curve, 69–73
 - of money, 287–288
 - reacting to decreases in, 80–81
- supply and demand
 - about, 61–63, 69
 - balancing, 12
 - balancing money, 287–288
 - comparing costs and benefits using, 132–133
 - defined, 403
 - elasticity, 67–69, 73–75
 - factors affecting quantity demanded, 63
 - graphing demand curve, 64–66
 - graphing supply curve, 69–73

supply and demand (*continued*)
market equilibrium, 75–81
opportunity costs, 66–67
price controls, 81–85
supply curve
defined, 403
graphing, 69–73
long-run aggregate, 309
market, 108
moving along the, 71–72
shifting the, 72–73
tax shifts of, 143–144
systematic error, 246, 404

T

target inventory levels, 320–321
tastes, as a factor affecting quantity demanded, 63
taxes
deadweight losses of, 142–146
to provide public goods, 220
security of future revenues, 342–343
tax-incentives fallacy, 376
TC. *See* total costs (TC)
Technical Stuff icon, 3
Technical Support (website), 4
technology
advances in, 45–46
constraints of, 28
promoting, 56–57
ranking as a public good, 222–223
Tesla, 227
Thaler, Richard (economist), 395–396
The Theory of Moral Sentiments (Smith), 259, 261
third parties, buying information from trusted, 213–214
TikTok, 242–243
time constraints, 28
time inconsistency, 257–258, 404
Tip icon, 4
total costs (TC)
about, 109, 111, 164
average, 115–116
exceeding total revenues, 128

total revenue (TR)
about, 109, 123, 164
relating to marginal revenue, 160–161
total costs exceeding, 128
variable costs (VC) exceeding, 126–128
total surplus
about, 140–141
calculating, 139–140
defined, 404
measuring gains with, 135–141
TR. *See* total revenue (TR)
trade deficit, 279–280
trade surplus, 279, 280
traditional economy, 53
Tragedy of the Commons, 206–208, 404
trash hauling monopolies, 170
trusts, 188
Twitter, 227, 236

U

Uber, 236
ultimatum game, 260–261, 404
unconventional monetary policy, 357
unintended-consequences fallacy, 376–377
U.S. housing bubble, 367
U.S. Steel, 225
used car market, 211–215
user engagement, balancing against need to charge fees, 236–237
utility. *See also* marginal utility
defined, 88, 404
maximizing total, 94–97
measuring happiness with, 25
types of, 88–89

V

variable costs (VC)
about, 111
average, 113–114, 115–116
defined, 404
exceeding total revenues, 126–128

Vimeo, 236
Visa, 236

W

wages
cost of, 316
cutting, 315–316
defined, 404
in flow of money, 269
moving back to real, 335–336
nominal, 335
tracing movement of real, 334–336
Wannamaker, John (entrepreneur), 254
Warning icon, 4
warranties, offering, 214
wealth, 404
The Wealth of Nations (Smith), 258
websites
Bureau of Labor Statistics, 296
Cheat Sheet, 4
Technical Support, 4
Wendy's, 194
Wikipedia, 240
winner-take-all industries, 242–243
workers, cutting, 315–316

X

Xbox, 236

Y

Y*. *See* full-employment output (Y*)
YouTube, 231, 236

Z

zero profits, 151–152
zero-coupon bond, 348

About the Author

Sean Masaki Flynn, PhD, is an associate professor of economics at Scripps College in Claremont, California.

A recurring commentator on radio and television, Sean holds a BA in economics from the University of Southern California and a PhD in economics from UC Berkeley, where he completed his dissertation under the supervision of Nobel Laureate George Akerlof.

As one of America's leading economic educators, Sean is a coauthor, along with Campbell McConnell and Stanley Brue, of the world's best-selling college economics textbook, *Economics: Principles, Problems, and Policies* (McGraw-Hill). The book's popularity is such that it's also the world's best-selling college textbook on any subject.

Sean's academic research focuses on the often puzzling and seemingly irrational behavior of stock market investors, but he's also investigated topics as wide-ranging as the factors that affect tipping behavior in restaurants and why you see a lot of unionized workers only in certain industries.

His recent focus has been on incorporating proven economic solutions directly into law.

Dedication

To my mother, Lt. Cmdr. Mikiko Flynn, MD, who showed me first-hand just how much is possible for an immigrant if they get to move to a country like ours that embraces education, hard work, and sensible economic policies that lift all boats. I love you, Mom.

Author's Acknowledgments

I'd like to thank the many great economists who managed to get things into my head despite my very thick skull.

Among my teachers, I can't help but thank Caroline Betts, Tim Cason, Richard Ciccetti, Michael DePrano, Richard Easterlin, Robert Kalaba, Timur Kuran, Jeffrey Nugent, and Morton Shapiro for the excellent education I received as an undergraduate at the University of Southern California.

I was equally blessed at UC Berkeley, where I got to complete a doctorate under the tutelage of some true intellectual giants, including George Akerlof, David Card, J. Bradford DeLong, Jan deVries, Barry Eichengreen, Richard Gilbert, Daniel McFadden, Maurey Obstfeld, Matthew Rabin, David Romer, Christina Romer, and Janet Yellen. It was especially fun when Professors McFadden and Akerlof won their respective Nobel Prizes during my last two years at Cal.

However, my fellow economics students often did more than my professors to explain things to me when I wasn't getting them, and they continue to educate me even now. So a very heartfelt thank you to Corinne Alexander, Lorenzo Blanco, Mark Carlson, Carlos Dobkin, Tim Doede, Mike Enriquez, Fabio Ghironi, Petra Geraats, Aaron Green, Galina Hale, Alan Marco, Carolina Marquez, Marcelo Moreira, Petra Moser, Marc Muendler, Stefan Palmqvist, Doug Park, Raj Patel, Steve Puller, Desiree Schaan, Doug Schwalm, Mark Stehr, Sam Thompson, Carla Tully, Jeff Weinstein, and Marta Wosinska.

I've also got to thank my current and former students at Scripps College, Vassar College, and UC Berkeley. Having to answer your many insightful questions made me a much better economist.

A big thank you to my literary agent Linda Roghaar and my old friend Mike Jones for getting me this book deal. They heard *Dummies* and immediately thought of me.

Elizabeth Stilwell, Thomas Hill, and the entire production team at Wiley also deserve huge praise. All their edits, suggestions, and formatting have turned out a book that's far better than anything I could have come up with on my own.

I also have to deeply thank Dr. Robert B. Harris, the technical editor of this book. His comments and suggestions have made it far better than it would have been otherwise.

Finally, many thanks to my parents for always making me do my homework.

Publisher's Acknowledgments

Senior Acquisitions Editor: Elizabeth Stilwell

Editor: Thomas Hill

Copyeditor: Karen Gill

Technical Editor: Robert B. Harris, PhD

Production Editor: Saikarthick Kumarasamy

Cover Photo: © Sittipong Phokawattana/
Shutterstock

WILEY END USER LICENSE AGREEMENT

Go to www.wiley.com/go/eula to access Wiley's ebook EULA.