

The
Applied Theory
of Price

DONALD N. McCLOSKEY

DEPARTMENT OF ECONOMICS

UNIVERSITY OF IOWA

The
Applied Theory
of Price

1985

SECOND EDITION

MACMILLAN PUBLISHING COMPANY NEW YORK
COLLIER MACMILLAN PUBLISHERS LONDON

Copyright © 1985, Macmillan Publishing Company, a division of Macmillan, Inc.

PRINTED IN THE UNITED STATES OF AMERICA

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the Publisher.

Earlier edition
copyright © 1982
by Macmillan Publishing Co., Inc.

Macmillan Publishing Company
866 Third Avenue, New York, New York 10022

Collier Macmillan Canada, Inc.

Library of Congress Cataloging in Publication Data

McCloskey, Donald N.
The applied theory of price.

Includes index.

1. Microeconomics. 2. Prices. 3. Microeconomics—
Problems, exercises, etc. 4. Prices—Problems, exer-
cises, etc. I. Title.
HB172.M39 1985 338.5'21 84-5621
ISBN 0-02-378520-9 (Hardcover Edition)
ISBN 0-02-946400-5 (International Edition)

Printing: 1 2 3 4 5 6 7 8 Year: 5 6 7 8 9 0 1 2 3

ISBN 0-02-378520-9

How to Use This Book

You learned to ride a bicycle by trying it out. You tried getting on, starting, turning, stopping. Each of these was a problem. By solving the problems one by one (and crash by crash) you developed the skill. This book teaches economic thinking as a skill like bicycle riding, and it teaches it the way skills are learned: problem solving. Most books in mathematics or chemistry teach the same way. They show you *what* the subject is all about, but their main purpose is to show you *how* to do it.

Most books in history or geography or economics do *not* use the problem-solving approach. It will seem harder to you at first. In a way it is harder: You can't rely on memorizing a bunch of definitions and dates; you have to get inside the skill and make it your own, like shooting a basket or knitting a sweater. In another way it's easier: Once you've caught on, that's it. You don't need refresher courses in riding a bicycle.

The doing is in the problems. In the early chapters the book gives you some advice on how to handle them. Your instructor can give you more. Try to give yourself advice, too, noticing when and how you cracked a problem. Being your own coach in this way is good practice for mastering other sorts of mental skills.

The chapters themselves have worked problems in the text, which will teach you "how." The "what" points that you have to know about microeconomics are explained in the chapter and highlighted in the What to Read For paragraphs at the beginning of each section.

The graphs play a big role in learning to think economics. Don't skip them. You should try to become fluent in the graphs and in the words treating each idea. The two are like two languages. It's sometimes easier to grasp an idea in one than in the other (and the easiness varies from one person to another: Some people have visual intelligence, others verbal).

Each graph has a full caption, which repeats the idea in different words than those in the text. It's a second chance to grasp it.

The diagrams have been carefully drawn with certain conventions to make

them easier. Lines and points are usually named instead of lettered, so you have another chance to get the idea but do not have to remember things like "locus QWerTy." Thick lines and heavy points are the really important ones; the thin and light ones have supporting roles. Curved labeling lines refer to lines, straight labeling lines to points. When the names of lines and points are referred to in the text they appear in a different typeface, such as this: The Budget Line. Imagine that your instructor is pointing to these terms on a diagram when you encounter them in the text.

At the end of each section are Exercises. These are limited and direct, meant to review your understanding of the material in the section. There are plenty of them, and the odd numbers are answered at the end of the book. You're not coaching yourself very intelligently, of course, if you look at the answers before trying to get them on your own.

After the Exercises are Problems, which will strike you at first as more vague and difficult. The world is vague and difficult, and the point is to ride the bicycle eventually in the world, not always on the practice track. Selected problems are also answered at the back of the book. These problems are identified in the text by a small circle beside the problem. Use these selected answers as models of how to think about the next problem.

The Study Guide, written by my collaborator Joel Scheraga, is an invaluable aid. It gives answers to the What to Read For questions, gives objective review questions, and provides additional help in problem solving. It contains hundreds of additional problems, many out of today's newspaper.

Preface

The main feature of this text on the theory of price is its numerous examples and its approximately 1000 worked problems. While showing the student the form of economics, the examples and problems stress throughout the way an economist uses form to think about substance.

The revisions in the second edition make the approach still more accessible to students. Though it is essentially the same book, the prose has been simplified. Exercises of a straightforward sort have been added and the problems simplified. There are more step-by-step numerical problems, and each section now begins with a piece on what to read for (questions of a general nature that the section answers). Like clean air, space is scarce, so these additions required some cuts, but the basic emphasis on problem solving remains.

The motivation for this approach is clear. A college graduate in engineering can predict that a badly designed bridge will fall down, and why; a college graduate in chemistry can predict that a badly designed compound will blow up, and why. College graduates in economics should be able to predict that a badly designed tax on gasoline will hurt society, and why, but too often they cannot. Our students understand the derivation of demand curves from choice-theoretic axioms, the symmetry of consumption and production theory, the role of prices in a market economy, and many other things, but too often they do not know how to apply them and have no idea how to find out. They know the formalities but not the substance of economics.

Whatever the purpose of a course in price theory—whether it is meant to produce informed citizens, clever economic actors, educated graduates, useful social engineers, or creative economic scientists—it should give the student more than the formalities. Problem solving does this. A lesson on externalities can take the form of an abstract essay: “Define ‘externality’ and translate it into concise mathematical form.” Alternatively, it can take the form of a concrete problem: “Is it true or false that because California is beautiful and has many magnificent public parks it is likely from the economic point of view to be overpopulated?” Students (and, I can say for myself, teachers) do not understand

economics until they have faced and answered problems applying it. In fact, without this practice they hardly recognize the economics: a problem is a way of saying "Note well." Textbooks in economics so far have told *about* economics. This one tells about but also tells *how*.

The "about" book is good in its place, toward the end of an economic education rather than toward its beginning. After one knows the flesh and blood of economics, the exposure of the logical skeleton comes as a wonderful discovery. Someone who thinks of Stefano Fenoaltea the consumer, of the Ford Motor Company, and of the southern textile industry as bits of marginal analysis and then is presented with one of the many books translating and extending Samuelson's *Foundations of Economic Analysis* for the common reader has an experience of intellectual delight. The generation of Samuelson himself, which is responsible for the style of the current texts, was educated originally in an older, more applied economics and had the delightful experience of discovering its general principles. "Bliss was it in that dawn to be alive,/But to be young [and trained in engineering mathematics] was very heaven!"

The attempt to have our students leap directly to heaven, however, has been a mistake. These days a book in microeconomics cannot contain a single derivative, or even very many equations, yet send the message that the form of economics is its scientific substance. The students learn economic calculus before they learn to reason economically, and their capacity for reasoning is permanently damaged. The point is not to banish formal training from economics but to place it at the right stage of the educational process.

The idea of using hundreds of practical problems to transmit a skill is a commonplace in other fields. My college textbook in calculus, and no doubt yours, was jammed with problems, more than half of them answered; my chemistry text was similarly structured. Reading about applications of a theory is a step in the right direction; the completed voyage is to apply the theory oneself. An economist can hardly maintain—with due respect to Leontief—that economic understanding is produced by a unique recipe calling for large amounts of problem solving in fixed proportions. The production function is doubtless neoclassical to the extent of permitting substitutions. The hypothesis suggested by the experience of other fields, however, is that the ratio of problem solving to other techniques in teaching economics is at present inefficiently low.

The other features of the book are pedagogic implications of this hypothesis. The test of utility in problem solving alters the standard list of topics a little, chiefly in order and emphasis. The treatment of supply and demand is unusually extensive because supply and demand is unusually useful: the book returns to it repeatedly, each time with new sophistication in using it. The analysis of two-factor production functions has been deemphasized in the theory of the firm and the industry, where it would only slow down the story, and emphasized in the theory of marginal productivity and the demand for factors of production, where it *is* the story. The Condorcet-Arrow paradox and similar issues in political economy are treated not as a puzzling addendum but are in the middle of a chapter on welfare economics, itself in the middle of the book, as a step in the development of national income as a measure of welfare. Giving measurable utility and risk a full treatment and locating it with ordinal indifference curves early in the book, instead of back in a section of special topics, recognizes its growing importance in economics and its many applications. Consumers' and producers' surpluses get an uncommonly full treatment as well, because they

link the welfare economics of national income, index numbers, and general equilibrium to supply and demand. Applied welfare economics is emphasized throughout the book because it motivates the behavioral theories, such as marginal productivity. The novelties of mathematical coverage are purposely few; these are included for their uses rather than their beauty.

When a piece of analysis is not useful, it is not included, whatever the tradition of textbooks has been. Viner's analysis of cost curves, for example, is well beloved but in its traditional form is not worth the fat chapter usually devoted to it. It appears in this text in a nontraditional form that highlights its usefulness. A younger tradition favors a section on linear programming, but the payoff in economic insight is too small, given the level of mathematical and economic sophistication of advanced undergraduates or first-year graduates, to warrant the large investment of pages necessary to do it well. Monopolistic competition is treated thoroughly but only in the locational context in which it has proven its utility. The kinky oligopoly demand curve is a poor example of discontinuities in marginal revenue and appears to be wrong besides. Every topic must meet such tests of cost and benefit in making the student into a problem solver.

Certain mechanical features of the book contribute to the teaching by problem and answer. The questions in the text are answered fully, as models for answering the questions that conclude many of the sections. These answers are given in the *Instructor's Manual*, which also contains a large number of fresh problems. The summaries at the end of each section will help the student to keep sight of the essential formal skeleton to which the problems are attached.

Each point is explained in as many ways as possible—verbally, mathematically, and diagrammatically. The book is particularly full of verbal explanation of the most important points and has double the usual number of diagrams. The diagrams have self-explanatory titles, and points and lines within them are named rather than symbolized. Each diagram has a full explanatory caption (prepared for the most part by John Martin) written to follow the argument but not the wording of the text. The knowledge most worth having in economics is elusive and bears repetition.

The novelty of a problem-solving approach to price theory requires little adjustment in the teacher's routine. The big change will be in what happens in the student's study, not in the teacher's class. The *Instructor's Manual* contains suggestions for the classroom. The teacher may wish to allocate less time to the less useful pieces of price theory and more to the more useful or lecture a little more on formal principle and a little less on applications, although the many applications in the book may well suggest still more. Above all, the teacher may wish to assign problem sets.

The Applied Theory of Price fits the standard courses in intermediate price theory. The student is supposed to have had an introductory course in economics, most of which has been forgotten over the summer, and no mathematics beyond high school. Since students of intermediate economics can be expected to be taking a calculus course (in which they will acquire some mathematical sophistication as well as the techniques of differentiation), the mathematical level in the book rises gradually, though calculus is limited to the chapter appendixes. For a single-term course for college juniors who have had the usual basic economics course, and that at a fairly elementary level, the core chapters will be ample, constituting a short book:

- Chapter 1 The Budget Line
- Chapter 2 The Consumer's Choice
- Chapter 5 Trade
- Chapter 6 Using Market Supply and Demand (Sections 6.1 and 6.2)
- Chapter 7 Measuring Supply and Demand (Section 7.1)
- Chapter 8 Production Possibilities (Sections 8.1 and 8.3)
- Chapter 10 Consumers' Surplus (Section 10.1)
- Chapter 11 The Firm (Sections 11.1, 11.2, and 11.3)
- Chapter 12 Cost Curves of the Firm
- Chapter 13 Competitive Industry (Section 13.1)
- Chapter 14 The Long-Run Supply Curve and the Principle of Entry
- Chapter 16 Competition for Property Rights (Section 16.1)
- Chapter 17 The Behavior of Monopoly
- Chapter 19 The Welfare Economics of Monopoly
- Chapter 22 Marginal Productivity and the Demand for Labor: The Fundamentals
- Chapter 25 The Supply of Labor (Section 25.1)
- Chapter 26 Capital's Supply and Demand (Section 26.1)

A still shorter course, aiming only at a thorough grasp of supply and demand (a noble and sufficient aim, it should be said), might thin out the later chapters in the list. For the year-long course in microeconomics for undergraduates that ought to become the standard, the whole book can be worked through methodically.

For very well prepared college juniors and MBA students, the whole book can be swallowed in a term without indigestion. I have used it on such audiences for many years with success, finding that in a very selective college even sophomores in their first economics course can digest most of it. We underestimate how much students can learn in three or four months if they are simply required to do so and face up to the requirement: look what the first course in college calculus or chemistry demands and gets from them.

For first-year students of graduate economics the book is useful as a refresher and foundation. I would like to see this book (or the others like it that will follow, if the principle of entry is true!) put in the hands of every one of them for their first month or so of graduate school. It is my experience, and probably yours, that even in highly selective graduate programs the students are weak in the bread and butter of economic thinking. Let them eat cake in December.

The book is tough, but so is life, and economics. We do our students a disservice if we pretend that economics is easy, to be learned by the rote methods with which so many bright and hardworking students conquered high school. Economics is not a list of terms but a way of thinking. If we can get this across to the students we will have at least given them the first part of knowledge, the knowledge that there is a great deal to know. And if the book is harder than most in some ways, it is easier in others. The students must try to identify with the economic actors in the problems, but they are burdened less with abstractions about the marginal rate of indifferent substitution. We must ask what the usual micro course is for, and how we would measure its success. Success is economic thoughtfulness; problem solving is how we all achieved it. It's time we let our students in on the secret. The payoff will be students who see the social world

as a thing of opportunity cost, marginal benefit, competition, collusion, equilibrium, search, ownership, maximization, entry, and scarcity; that is, students who think like economists. And that, we can agree, is a fine thing.

Acknowledgments

In the summer of 1975 Gilbert Ghez of Roosevelt University and I decided to write this book. After putting in much work, he decided that the opportunity cost of text writing was too high, but his contribution was important. My other debts, specific and general, run as follows. At various times Alyce Monroe, Linda Freeman, Tricia Pate, Marye Allen, and Marguerite Knoedel have typed order out of chaos. Anthony English, Charles Place, Chip Price, Eileen Schlesinger, and Robert Hunter, my editors at Macmillan, have been patient and encouraging far beyond the call of profit. Together they illustrate the paradoxical economic theme that capitalism can be altruistic, or perhaps that altruists can be capitalists. My students at Chicago and Iowa in courses on price theory since 1968 have forced me to think clearly and have laughed at my jokes. John Komlos, Moonie Lavi, Bruce Lehmann, Fred Lindahl, Shumeet Banerji, Gundar Kaupins, and David Arens favored me with written comments, but I am uneasily aware that I have lost track of many others whose comments, written and verbal, mattered, too. I want them to tell me so that I can make amends later. Bart Taub, Kevin O'Meara, and especially John Martin made unusual contributions to the book as teaching assistants in courses—all three contributed greatly to the stock of problems, and Martin did the captions for many of the diagrams and numerous other editorial tasks with high intelligence and good taste. But again I fear that my faulty memory is not recalling every name it should. I shall not forget Gary Hawke for his detailed and encouraging comments. Many users have commented on the first edition, among them Ronald Johnson, Byron Boulier, David Vrooman, Fred Carstensen, Lin Lindert, Eric Gustafson, Frederick Harris, and Joel Mokyr. Joel Scheraga of Rutgers did an astonishingly good job in helping me revise the second edition, and in writing the Student Workbook.

I cannot thank all those by name who labored anonymously over commentaries to publishers, a task that pays poorly but that defines our little community. The ones whose identities I do know—Jean-Pierre Benoit, Columbia University; George Bittlingmayer, Washington University; Jack Boan, University of Regina; Richard V. Burkhauser, Vanderbilt University; Gerald Flueckiger, Miami University; Philip Graves, University of Colorado; Frederick H. Harris, University of Texas at Arlington; Lawrence W. Kenny, University of Florida; Etsuske Masuda, State University of New York, Buffalo; Richard P. McLean, University of Pennsylvania; Kenneth B. Moberg, Purdue University, Calumet; Jon Nelson, Pennsylvania State University; Alain Sheer, Texas A&M University; Richard Sutch, University of California, Berkeley; John Vernon, Duke University—gave Macmillan remarkable value for money and me much good counsel. I thank especially the friends and commentators and editors who have told me I was doing something good. Someone who will exaggerate, even flat-out lie, to raise a man's spirit is friend to a higher truth.

D. N. M.

About the Author

Donald N. McCloskey is Professor of Economics and History at the University of Iowa. He received his B.A. and Ph.D. from Harvard. Previous to joining the University of Iowa in 1981, he was a member of the faculty at the University of Chicago for 12 years, teaching introductory economics, graduate price theory, and British economic history. He has been a visiting professor at Stanford University, the University of London, York University, and the Australian National University, and has held senior fellowships from the National Science Foundation, the National Endowment for the Humanities, the Guggenheim Foundation, and the Institute for Advanced Study. He has written three books on economic history and economic method, edited or compiled three more, and contributed dozens of articles to professional journals. Presently he is editor of the *Journal of Economic History*, and his current research interests include the rhetoric of economics and the history of the market.

Contents

*N = 65 sections
in 26 sessions: $\frac{1}{2}$ per session
exact # of sessions = chps!
(2 exams; Thanksgiving; one AWOL)*

How to Use This Book $x_2 = 52$, back 13

v

Preface

vii

Introduction

1

I DEMAND

5

CHAPTER 1

* The Budget Line

6

- * 1.1 Scarcity and the Budget Line 6
- * 1.2 Income in the Budget Line 14
- * 1.3 Prices in the Budget Line: The Law of Demand 19

CHAPTER 2

* The Consumer's Choice

28

- * 2.1 The Pursuit of Happiness 28
- * 2.2 The Shape of Indifference Curves 34
- * 2.3 The Uses of Indifference Contours 40
- * 2.4 Measurable Utility 44

CHAPTER 3

The Measurement of Utility and the Economics of Risk

51

- 3.1 Paradise Lost: Nonmeasurable Utility 51
- 3.2 Paradise Regained: Measurable Utility 54

3.3 Imperfections in Paradise? 59	
3.4 Living in Paradise: The Uses of Measurable Utility 62	

CHAPTER 4	
Indifference Curves and Demand	71

II EXCHANGE **83**

CHAPTER 5	
Trade	84

- * 5.1 Supply and Demand 84
- * 5.2 Exchange Between Two People or Nations 89
- * 5.3 Trade Among Many People: Behavior 96
- * 5.4 Trade Among Many People: Happiness 101

CHAPTER 6	
Using Market Supply and Demand	107

- * 6.1 The Uses of Equilibrium 107
- * 6.2 Adding Up Supply and Demand: Equilibrium 111
- 6.3 Adding Up Supply and Demand: The Diagrams for Analyses 117
- 6.4 Extensions of Supply and Demand 122

CHAPTER 7	
Measuring Supply and Demand	131

- * 7.1 Elasticities of Supply and Demand: The Essential Ideas 131
- 7.2 Constant Elasticities 140

III PRODUCTION AND WELFARE **151**

CHAPTER 8	
Production Possibilities	152

- * 8.1 Production and Specialization 152
- 8.2 The Production Function Also Leads to the Production Possibility Curve 158
- * 8.3 How an Economy Works in the Large 165

CHAPTER 9	
The Economics of Welfare and Politics	171

- 9.1 The Economics of Ethics 171
- * 9.2 National Income and Its Ambiguities 179
- * 9.3 The Economics of Politics 187

CHAPTER 10 Consumers' Surplus • 10.1 Consumers' Surplus: The Elements 196 10.2 Further Uses of Consumers' Surplus: Middlemen and Other Exchanges 207 10.3 A Change in Consumers' and Producers' Surplus Is the Same as a Change in National Income 217	196
<hr/>	
IV PRODUCTION AND MARKETS The Firm * 11.1 Whether and Why Firms Exist 224 * 11.2 The Profit Motive 229 * 11.3 Marginal Cost: Why the Firm Produces What It Produces 234 11.4 Equimarginality: How the Firm Produces What It Produces at Minimum Cost 241	223
<hr/>	
CHAPTER 12 Cost Curves of the Firm * 12.1 Production and Input Supply: Why Marginal Cost Is As It Is 247 * 12.2 Cost Curves in Use: The Long Run and the Short Run 257 * 12.3 Cost Curves in Use: Fixed and Variable Costs 265	247
<hr/>	
CHAPTER 13 Competitive Industry * 13.1 Industry Supply with a Fixed Number of Firms and Costs Independent: Optimality and Upward Slope 276 13.2 Industry Supply with a Fixed Number of Firms but Costs Interdependent: Externalities 283	276
<hr/>	
CHAPTER 14 The Long-Run Supply Curve and the Principle of Entry * 14.1 The Long-Run Supply Curve 290 * 14.2 The Uses of Entry and Exit 298	290
<hr/>	
CHAPTER 15 Taxes	307

CHAPTER 16	
Competition for Property Rights	320
• 16.1 Competition for Supernormal Profits	320
• 16.2 Unassigned Property Rights and External Effects	330

V MONOPOLY 343

CHAPTER 17	
The Behavior of Monopoly	344

- 17.1 Monopoly: The Elements 344
- 17.2 Advanced Applications of Simple Monopoly 360

CHAPTER 18	
Measuring Monopoly	371

CHAPTER 19	
The Welfare Economics of Monopoly	388

CHAPTER 20	
Monopolistic Competition and the Economics of Location	402

- 20.1 Monopolistic Competition as Competition Among Local Monopolies 402
- 20.2 The Theory of Location 411

CHAPTER 21	
Competition Among the Few	420

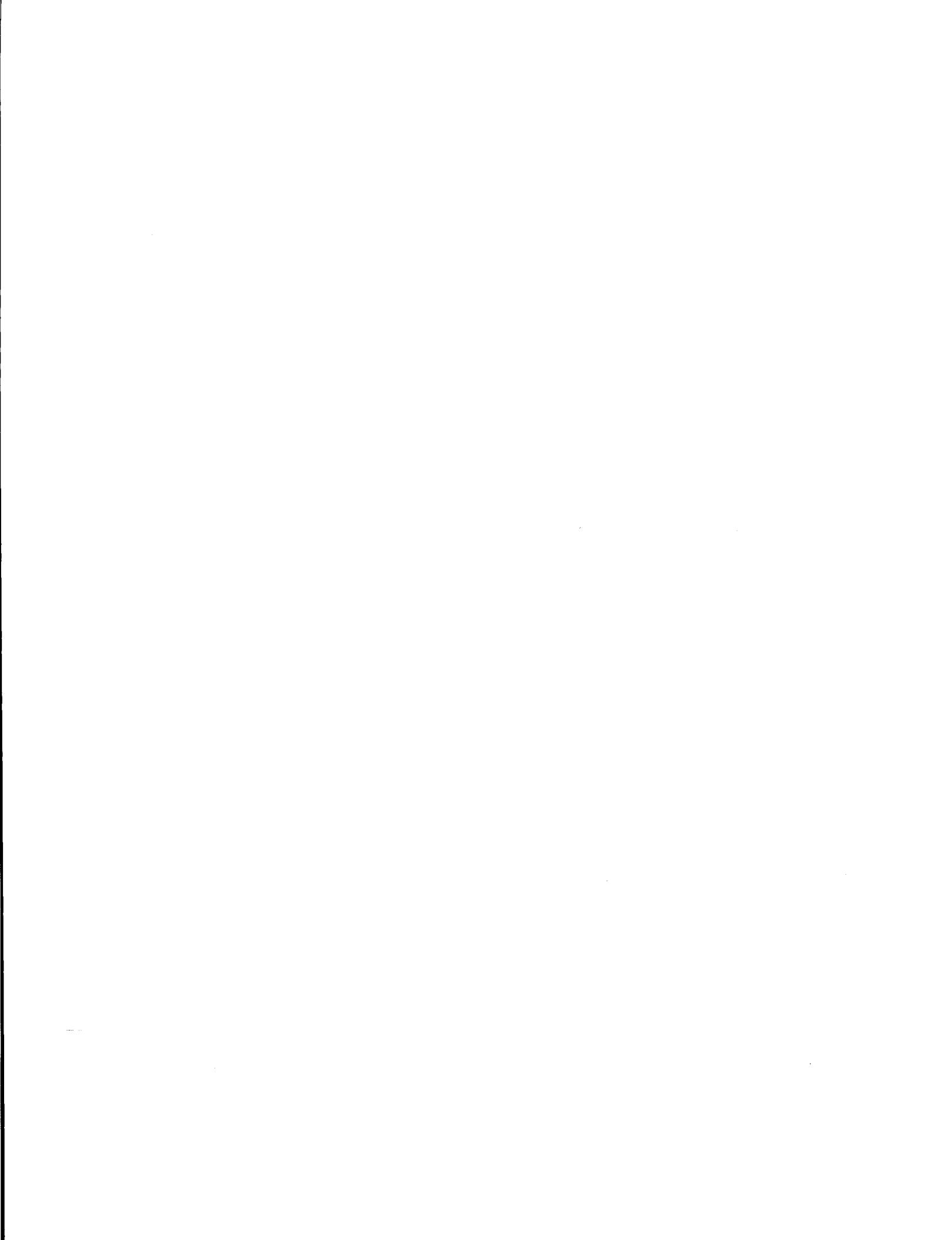
- 21.1 Simple Solutions: Bertrand and Cournot 420
- 21.2 The Irrationality of Simple Solutions to the Problem of Fewness 428
- 21.3 Cartels and Game Theory as Solutions to the Oligopoly Pattern 434

VI LABOR, CAPITAL, AND DISTRIBUTION 447

CHAPTER 22	
Marginal Productivity and the Demand for Labor: The Fundamentals	448

- 22.1 Labor as a Commodity 448
- 22.2 Marginal Productivity as the Demand for Labor by the Firm 457

CHAPTER 23		
Marginal Productivity in Theory and in Use		467
23.1	Many Inputs, Constant Returns to Scale, and the Fundamental Theorem	467
23.2	Changes in the Production Function	476
CHAPTER 24		
Misallocation and Monopoly in Factor Markets		487
24.1	Good and Bad Allocation	487
24.2	Monopoly in Factor Markets	498
CHAPTER 25		
The Supply of Labor		508
25.1	The Leading Idea in Labor Supply: Compensating Differentials	508
25.2	The Choice Between Work and Leisure	513
25.3	The Choice of Schemes of Payment	521
CHAPTER 26		
Capital's Supply and Demand		528
26.1	The Interest Rate	528
26.2	Supply and Demand Curves	535
<hr/>		
APPENDIX		
Answers to Odd-Numbered Exercises and Selected Problems		542
Index		619



Introduction

You have just embarked on the study of the theory of price, known also as *microeconomics*, the science of markets. Although its Greek meaning is "small housekeeping," microeconomics is not the little or trivial portion of economics. On the contrary, it comes close to being the whole. It is essential for an understanding beyond the first course. Not all fields of economics are based on microeconomics, but all strive to be: Most of the lasting advances in economic thinking over the past century or so have consisted of reducing one or another piece of economic behavior to microeconomics. What, then, is this craft or sullen art? Put briefly, it is the understanding of maximization and markets. It is one of the great products of the human mind.

You wouldn't know it from the image of economists as a confused mob of social forecasters. You may have heard of disagreements among economists—that if all economists in the world were placed end to end they wouldn't reach a conclusion; that if ten economists went into a conference room they would come out with eleven different opinions. Ha, ha. Very funny. The truth is that when economists disagree it is commonly about *macroeconomics*—the study of inflation and unemployment—not about the subject of this book, *microeconomics*—the study of markets. Similarly, astronomers disagree about the age of the universe or the way galaxies are formed, yet agree about why stars burn or why the moon's orbit is elliptical. As you might expect, a sample survey of economists found "more consensus about micro issues than macro issues."¹

The astronomers are spared a further source of disagreement: Economic questions are often moral as well as factual. Should we support milk prices? Should the income tax be made strictly proportional to income? Should we restrict immigration from Mexico? These involve issues of morality as well as of fact. A moral issue is whether it is good to keep poor Mexicans from bettering their lot by coming to the United States; a factual issue is how immigration affects

¹ See J. R. Kearn, Clayne L. Pope, Gordon T. Whiting, and Larry T. Wimmer, "A Confusion of Economists?" *American Economic Review* 69 (May 1979): 33.

American workers. Both issues figure in the decision, and economists cannot evade the need to make a moral judgment. This book will talk much of the morality. But no astronomer need ask whether the sun *ought* to burn on hydrogen fusion. What matters is that it does, or can. She need not dispute with her colleagues about morality, since morality is not the issue. The survey found in fact that among economists “‘can’ propositions generate more consensus than ‘should’ ones.”

Considering the obstacles, economists agree about a surprisingly large number of things. Their agreements, in fact, are often about things that noneconomists would think silly or wrong or even evil. That is, economists are in surprising agreement about surprising statements. To shock you out of your complacency, and keep you awake, this book will put things in surprising ways. The survey asked the economists to state whether they “generally agreed” with the following statements, “agreed with provisions,” or “generally disagreed”:

1. Tariffs and import quotas reduce general economic welfare.
2. A minimum wage increases unemployment among young and unskilled workers.
3. A ceiling on rents reduces the quantity and quality of housing available.
4. Effluent taxes [taxes on chemical wastes, for instance] represent a better approach to pollution control than imposition of pollution ceilings.
5. Reducing the regulatory power of the Interstate Commerce Commission [a government group regulating trucking and railroads] and the Civil Aeronautics Board [regulating airlines] would improve the efficiency of the United States economy.
6. The ceiling on interest paid on time deposits should be removed [it has been since the survey was taken].

Having had an introductory course in economics you can probably guess how economists would respond—“generally agree” or “agree with provisions” to every one, by majorities from 78 to 98%. But coming fresh from reading the newspaper, if you had as little economic education as newspaper reporters usually have, you would probably respond exactly the other way. Tariffs “save jobs” in the auto industry. The minimum wage “protects working people.” A rent ceiling “stops rent gouging.” Taxes on waste would never stop polluters, who “need” to pollute. The regulatory agencies protect us from “ruinous competition” and “chaos” in the market for bus rides, as did once the ceiling on interest in the market for banking services.

The list of surprising agreements is a long one. Most of the 20,000 or so members of the American Economic Association would answer yes to questions such as:

1. If gasoline is taxed to conserve energy, will the quantity consumed go down by a nontrivial amount, despite the protestations of drivers that they cannot do with less than the amount they are now consuming?
2. Was the rise in the standard of living of the American worker over the last 50 years chiefly a result of better knowledge and more machines rather than of activity by trade unions?
3. Is the American Medical Association, far from being a benevolent organization set on improving medical care, in fact a monopolistic trade union like the plumbers, longshoremen, and electricians?
4. Does the resting place of the burden of the social security tax depend exclusively

on how workers and employers react to a change in wages, and not at all on the legal division of the tax (paid half by workers, half by employers)?

5. Is there an optimal amount, greater than none, of polluted air and water, noisy streets and airports, and ruined countryside?

And so forth. Notice that the economist's affirmative answers are bipartisan, stepping on everyone's toes. Republican or Democrat, you can find an economist to insult your most cherished belief. Furthermore, the answers are not mere matters of faith. The economist could persuade the open-minded noneconomist that these economic propositions are true by the same method that an astronomer would use to persuade them that astronomical propositions are true: refined common sense, consistent reasoning, and ascertainable fact.

The economist faces the special obstacle that the people being persuaded are themselves economic bodies and have elaborate opinions of their own. The Earth's own opinion about the movement of heavenly bodies would probably be that they all move around the Earth itself in circles. Untutored economic experience is a bad teacher of economics, just as the unaided eye is a bad teacher of astronomy.

Practically everything that you thought you knew about economics before studying it is wrong: Inflation hurts everyone; market competition is chaotic, leaving consumers at the mercy of business; if the penny is eliminated from circulation, merchants will round up to the nearest nickel, contributing to inflation; the draft makes the armed forces cheaper; price controls on natural gas will always help consumers; energy consumption can be predicted on the basis of physical need, without regard to its price; Lake Erie should, of course, be drinkable; a temporary boycott can permanently reduce a high price; a rise in the minimum wage makes workers better off; a tax imposed on employers in Chicago is never a burden on employees; and on and on.

The vocabulary of such ersatz economics, the economics of the man in the street, contributes to the confusion: Unions and corporations have more "bargaining power" than do their opponents and, hence, "exploit" them; a consumer can "afford" medical care, "needs" housing, and finds food a "basic necessity"; business managers maintain their "profit margins," probably "obscene" or "unwarranted," by "passing along" a higher wage, which causes workers to demand still higher wages, in a "vicious circle"; the protection of the American worker's "living wage" from "unfair competition" by "cheap foreign labor" should be high on the nation's list of "priorities." To understand price theory you must clear your mind of such jargon, just as to understand astronomy you must stop thinking of the sun's "rising." This book will attack ersatz economics again and again, with ever-increasing violence. Price theory uses words such as "scarcity," "opportunity cost," "rationality," "competition," "monopoly," "equilibrium," "incidence," "arbitrage," "production functions," "marginal product," and "supply and demand." A widespread but childish presumption to the contrary, however, learning the vocabulary is not learning the science. Nor do you learn it by memorizing its axioms or formulas, tables of numbers, and facts. You learn the science by practicing it, the way you learned to play baseball or the tuba. The basic notion of this book is a variant on the old joke: "How do I get to Carnegie Hall?" "Practice, practice." How do you get to the other side of a course in price theory or to the Council of Economic Advisors? Practice, practice. This book gives you the practice, lots of it.

The practice comes as exercises and problems, similar to the word problems that made your life miserable in chemistry, physics, or math. ("If a bee going 10 mph is chasing a man running away at 5 mph, and starts 10 feet away, when will the man get stung?") You will find that economics is less quantitative than the physical sciences, and the problems often require little essay answers of the sort you give in history exams. Yet they have the puzzle quality that science problems have. And *never* will a problem in this book be mere regurgitation ("Describe the Battle of Bull Run"; or "What are three simple compounds of carbon and nitrogen?"). You will very seldom be able to answer a problem by flipping back through the text to find The Answer. Sorry, it ain't there. It has to be in your brain. That is, you have to have learned by the end of each section how to *think* like an economist about the subject at hand.

The book gives you as much help in achieving this noble end as anyone could reasonably demand. The first few chapters contain lots of help in solving the problems. With the book and your instructor's examples and explanations, you can teach yourself to think like an economist. It's about as hard as elementary Latin or Spanish: If you don't put in the hours learning the regular conjugations, the vocabulary, the irregular verbs—by practice, practice, practice—you will learn nothing at all: a few grunts and how to order a sandwich in Mexico City. But if you work at it, it will come. The results will be amazing. Your thinking about an astonishing range of matters will change, for the better.

The theory of price is one among the larger intellectual achievements of the nineteenth century, such as the theory of heat engines, the decipherment of hieroglyphics, the professionalization of history, the invention of abstract algebras, and the theory of evolution. Its market theory is the characteristic gift of economics to the study of society. It applies to the enormous range of ordinary markets for wheat, ditch diggers, insurance, haircuts, heroin, police, guns, Bibles, typewriters, telephones, professors, and preachers. But it also applies to extraordinary markets for education, property rights, political favors, pollution, marriage, discrimination, business skill, charity, public housing, and crime. Price theory explains much human behavior.

A grasp of it distinguishes an economist from a sociologist, historian, engineer, or applied mathematician interested in economic matters. Confident skill in price theory defines an economist just as confident skill in diagnosis defines a medical doctor or confident skill in leading others in battle defines a military officer. Other skills are subordinate. When economists say scornfully that so-and-so, a reputed expert on economics, is "not an economist," they do not mean that he is ignorant of the institutions of economic life, or of its history, or of its statistics, or of its mathematical representations. They mean that he is ignorant of the theory of price.

A major achievement of Western thought, the jewel of the social sciences, a widely applicable tool for understanding society, a necessary condition for economic expertise—price theory is worth a little effort.

I

DEMAND

CHAPTER

1

The Budget Line

1.1 Scarcity and the Budget Line

What to Read For¹ What do economists mean by *scarcity*? Are most things *scarce*? What are *fungible* goods, and how is this concept related to scarcity? Which point won the game? What unit sets the value of a pile of goods? What is a *budget line*, and how is it related to scarcity? What is *opportunity cost*, and what is its relation to the price of one thing in terms of another? How do you calculate *relative price*, given two money prices? In a two-good diagram, how is a rise in the price of one of the goods (and a fall in the price of the other) represented using a budget line?

Scarcity in One Dimension A thing is scarce when it is desirable but limited in quantity. Oxygen to breath is desirable, but it is not limited and it is therefore not scarce. Garbage is unlimited, but it is not desirable, and it is therefore not scarce. The radioactive waste emitted by a nuclear power plant in a year is limited, but it is not desirable, therefore it is not scarce. Food, housing, education, military protection, justice, freedom, books, candy, friendship, children, drinkable water, travel, clothing, and good health are all desirable and limited and therefore scarce.

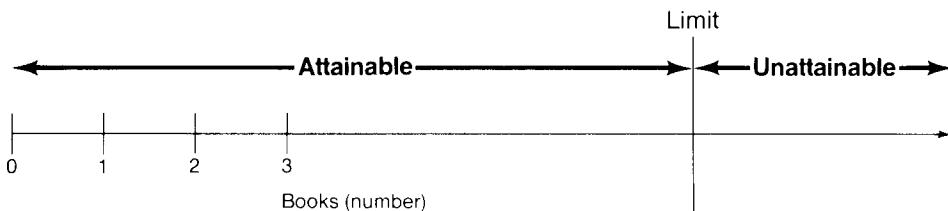
The truth from which economics begins is that *scarcity* is very common. We cannot have much more of the things we want merely by wishing. The fact is unspeakably sad. It is sad that our poverty—our inability to have all the automobiles or ice cream sundaes or beautiful poems or true loves that we would like—confines us so. It is sad that there is no free lunch.

The definition of scarcity can be put in a diagram by saying that, along the scale of amounts of desirable things, such as books in your library or food in

¹ A little pile of questions called What to Read For will begin every section. If you can't answer all the questions by the end of the section of text, you have not read carefully enough, or thought enough about what you have read. Keep looking back at them to see if you're getting the main points.

Figure 1.1
A Scarce Thing Is Limited

The Limit is the most books you can consume because you are not infinitely wealthy.



the world, there comes a stop. You and the world cannot have unlimited amounts of books or food. In Figure 1.1 the area marked Unattainable is just that.

Fungibility

The diagrammatic way of putting it makes clear that each book is scarce, though it is not the "last" book that uses up the allotment of books. The reason is that the scale treats all books alike. They can be substituted one for the other. They are, to use a strange word popular among economists, *fungible*. The notion that scarce goods are fungible is surprisingly powerful in applications. The University of Iowa beats the University of Michigan for the Big Ten basketball championship by two points in double overtime, 70 to 68. Whose two points won the game? One's first thought is to look to the last points made. But points are fungible. Any two points can be viewed as the crucial points that make the difference between a score of 68 to 68 and 70 to 68. The shot that Weisskoff made in the first 5 minutes of play counts as much as the last. For the purpose of making the score what it is, there is effectively no last, no crucial, point. Consider the following.

Q: Your mother gives you \$500 toward a new car. If you were going to buy the car anyway, in what sense is her gift crucial "for" the car?

A: Not at all, or entirely, or partly. The answer is arbitrary because the question is meaningless. You can, if you wish, take the very dollar bills from her gift and pay them to the car dealer. That would be to use them "for" the car. But if you used those dollar bills to pay

your grocery bills for a while, using other dollar bills to buy the car, would her gift have been any less "for" the car? No. You are made \$500 richer and use the money to buy a car. On the other hand, one could say that some other \$500, gotten "earlier," was used to buy the car. Since money is fungible, there is no sensible way in which to draw a line between your mother's money and other money.

COMMENT

You have just been given an example of problem solving in economics. The best way to use the problems in the text is to cover the answer with your hand and think about what the answer might be. You must be self-critical in thinking up answers. Suppose you answered so: "The \$500 is crucial 'for' the car, because you can pay it to the car dealer." The remark is true so far as it goes, but it does not go far. You aren't getting into the spirit of thinking seriously about fungibility. Give yourself one point out of five possible points and try to think tougher and longer the next time.

Notice that the best "answer" here in fact says that the question is

meaningless. Be on the lookout for meaningless questions: The world tosses up a lot of them, and to think economically you must develop the sense not to try to answer them.

The applications of the idea of fungibility are limitless. In general, when you hear someone speaking of certain points being the reason for a victory, or certain motivations being the reason for a decision, you should suspect that they are ignoring the fungibility of things.

Scarcity in Two Dimensions

Because of fungibility, the value of the last increment to a pile of things governs the value of the whole. All the things measured along a single axis (tons of wheat, number of books, dollars of income) are taken to be just like any other. Wheat is distinguishable from books, of course. The world is not composed of all-purpose little animals that can be eaten, drunk, used for roofing, woven into cloth, and so forth. The natural extension of the single axis measuring amounts of, say, Books is a second axis measuring amounts of All Other Goods. The expression of scarcity in such a diagram is that the person or the society must stay inside some Attainable area (see Figure 1.2).

To capture the essence of scarcity, the curve bounding the Attainable area need not be smooth. Its shape depends on the methods and resources available for making Books or All Other Goods. For present purposes the only thing important about its shape is that it eventually slopes downward. That is, one cannot in a world of scarce resources consume unlimited amounts of both Books and All Other Goods. For simplicity the next few chapters use straight lines. Economists call these all sorts of things: the *transformation curve*, the *production possibility curve*, the Books–All Other Goods *trade-off*, the *production frontier*, or simply the *budget line*.

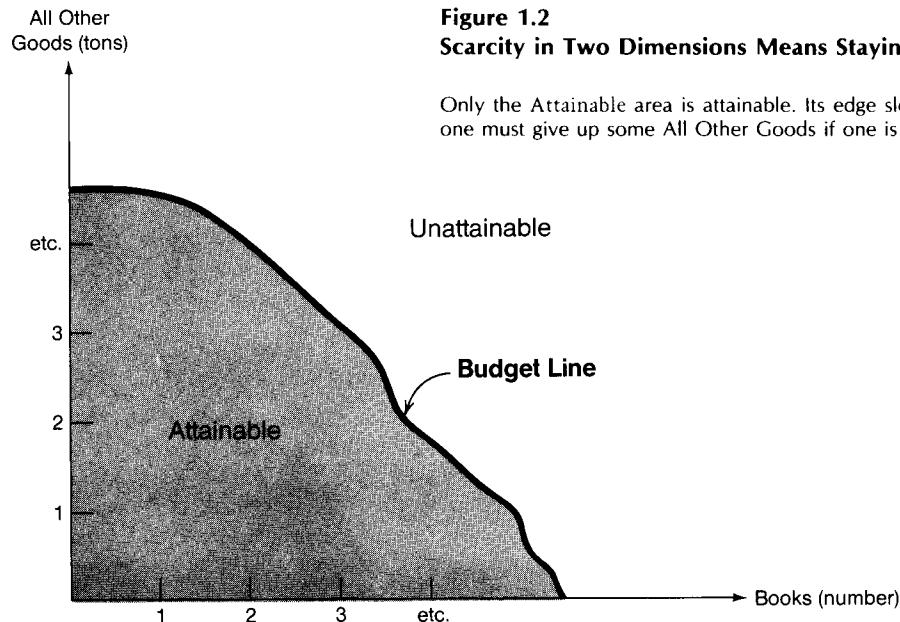


Figure 1.2
Scarcity in Two Dimensions Means Staying Inside One's Budget Line

Only the Attainable area is attainable. Its edge slopes down. That is, to get more Books, one must give up some All Other Goods if one is on the Budget Line.

Opportunity Cost Is the Slope of the Budget Line

Economists spend a great deal of time thinking about the budget lines facing people and societies. The first thing to notice about a budget line between, say, education and other goods is its slope. The slope, or "the rise over the run" (as the phrase in high school algebra puts it), is equal to the decrease in education divided by the increase in other goods. It measures the cost of increasing the amount of one commodity in terms of the consumption of the other goods that is sacrificed. It is the cost of one item in terms of the other. In other words, it measures the *opportunity cost* of one in terms of the other. Read that all again, slowly: It's important.

The economist's way of measuring the cost of, for example, studying economics one more year is *not* the pain and suffering involved in the year. It is the maximum amount of other scarce things sacrificed by choosing the additional year. The cost of assembling another automobile is not the sum of physical toil and mental anguish suffered by workers or the sum of their hours of work, or even the sum of their paychecks. *It is the output of other things producible by the workers in some alternative employment of their energies.* The alternative employment can be in another job that pays a wage, in housework, or in leisure.

T or F: The cost to a student of a year of college is the cost of books, tuition, room, and board.

A: The student sacrifices the highest wage that she could earn if she were employed full time at another job. Books, tuition, room, and board, therefore, are not the

only sacrifice. Therefore, false. Notice that to gain a year of education older students typically sacrifice more than do younger students, because the older students can usually earn higher wages. This is one reason (not a very important one) that the years of education are concentrated in youth.

COMMENT Another problem answered. Like many problems in the book, this one is in the True or False form. It could have been put as a question: "What is the cost to a student of a year of college?" Putting it in the T or F way gives you a clearer target to shoot at. And many false statements about economic events come in a form you can easily translate into a T or F question. A rise in tuition proposed for the University of Iowa might be attacked as raising by 10% "the cost" of a year of college from \$2000 to \$2200. You should try to think of all such remarks as: "T or F: A rise in tuition is a rise in the whole cost of a year of college."

The way to answer the question is to put yourself into it: What would you sacrifice by going to college for a fifth year? What would enter into your decision? Merely the tuition? Wouldn't you also consider the lost opportunity to work at a salary and to use what the salary could buy? Repeat: Put yourself into the situation. Economics is refined common sense. Use your imagination to see into the situation.

Opportunity cost, then, is a result of scarcity. This is represented by the downward slope of the budget line. One must choose, and in choosing you move down the line. Though unpleasant, it is a fact of life. You cannot go to school full time and also work full time. You cannot study economics and also revel to the fullest in the latest album. You cannot go home for Christmas and also go to Fort Lauderdale. To live is to choose, and to choose is to come under the eye of the economist, or of the poet:

Two roads diverged in a yellow wood,
 And sorry I could not travel both
 And be one traveler, long I stood
 And looked down one as far as I could
 To where it bent in the undergrowth.²

The Slope Is the Relative Price

At a more mundane level, the opportunity cost of a thing is simply its price along the budget line, namely, the amount of other things you have to give up to acquire a unit of the thing you are buying.

Q: You go to a bar serving beer at \$0.50 a glass and wine at \$1.00 a glass. What is the price of wine in terms of beer?

A: Clearly, it is two glasses of beer per glass of wine. One dollar spent on a glass of wine could buy two glasses

COMMENT

Numerical problems with real numbers are often easier to answer than more abstract problems like the one about the cost of college. Make use of this fact. Use real numbers as examples in an abstract problem. Make them up.

Notice the important point here. The talk has been about how many physical units of one good you give up in order to get a physical unit of another good. You have to realize in this problem that the question asks for physical units—so many glasses of beer per glass of wine. It does not ask for dollars, which is the usual way of speaking of the price. This is elementary, but mildly confusing at the beginning. The currency used does not matter. It can be hours instead of dollars, for instance.

Q: A Victorian novel takes 20 hours to read; the *Classic* comic book recounting the same story takes 15 minutes. Considering the time spent, what is the price of novels in terms of comics? (See Figure 1.3.)

A: Again,

$$\frac{\text{Comics}}{\text{Novel}} = \frac{\text{hours/novel}}{\text{hours/comic}} = \frac{20}{1/4} = 80 \text{ comics per novel}$$

COMMENT

You can use the units to check your answer. The “price” of a novel must be some number of the other thing—comics—per novel. So it is.

Recognize that different slopes of budget lines represent different relative prices. Budget line 1984 in Figure 1.4 represents a higher price of books relative to clothes than does 1985. To identify which price has changed, imagine sliding

² From “The Road Not Taken” from *The Poetry of Robert Frost*, edited by Edward Connery Lathem. Copyright 1916, © 1969 by Holt, Rinehart and Winston. Copyright 1944 by Robert Frost. Reprinted by permission of Holt, Rinehart and Winston, Publishers. British rights to reprint these lines by permission of Jonathan Cape Limited, London.

of beer. Notice that, because the question asks for quantities of beer per glass of wine, one divides the given money price of wine by the money price of beer:

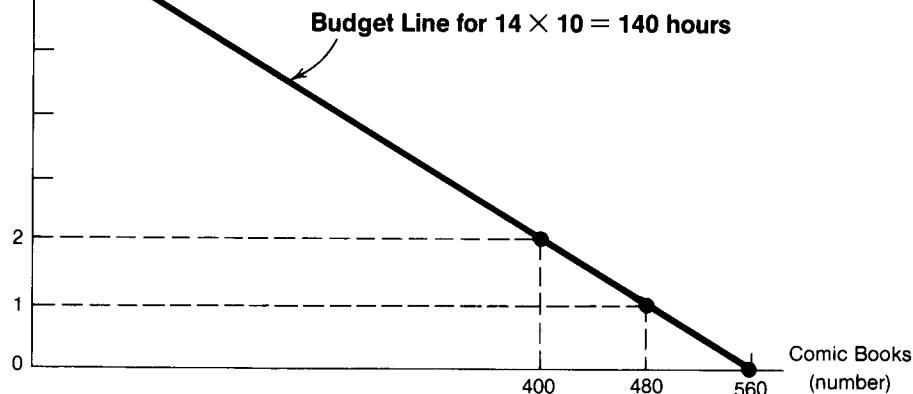
$$\text{Beers per wine} = \frac{\text{glasses of beer}}{\text{glasses of wine}} = \frac{\text{dollars/glass of wine}}{\text{dollars/glass of beer}}$$

or 0.0125 novel per comic. With two weeks of leisure in which to read 10 hours a day, you could consume 7 Victorian novels or 560 comics; alternatively, you could trade off 80 of the 560 comics in order to read 1 novel and $560 - 80 = 480$ comics, or 2 novels and $560 - 80(2) = 400$ comics. In other words, you can consume in two weeks any straight-line combination between 7 novels and 560 comics.

Victorian Novels
(number)

Figure 1.3
The Budget Line of Literary Consumption

The Budget Line, drawn for a given number of hours available for reading, shows by its slope the rate at which the consumer forgoes reading Novels to read Comic Books.

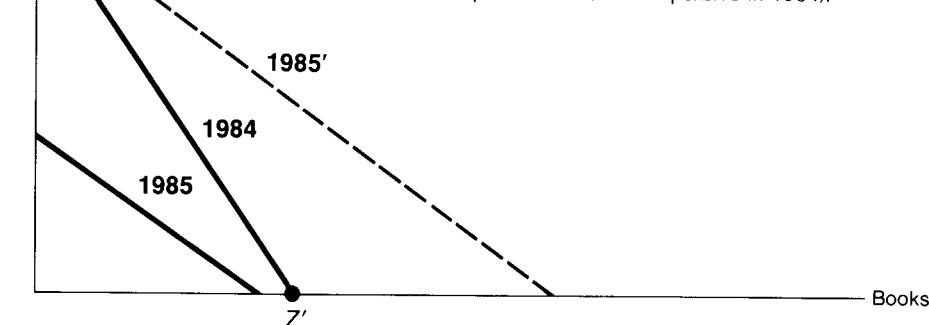


1985 over to 1985' (read it as "1985 prime"), parallel to 1985 (parallel budget lines have, of course, the same slope *and represent, therefore, the same relative price*). Starting at the same amount of clothes, Z , you could get more books along 1985' than along 1984. Evidently from 1984 to 1985 books became relatively cheaper (and clothes, of course, relatively more expensive).

Clothes

Figure 1.4
Which Relative Price Has Fallen?

In 1985 more books can be acquired per unit of clothing than in 1984. This is easily seen by pretending (through a shift in the budget line) that the maximum quantity of clothing is the same in both years, but the maximum quantity of books is larger: Books are relatively cheaper in 1985 (more expensive in 1984).



Q: Instead of sliding the 1985 line over to point Z , slide it over to Z' . Are books still relatively cheaper in 1985 than in 1984? Explain.

A: Yes, they are relatively cheaper. The same amount of books (namely, the amount at Z') buys fewer clothes

in 1985 than it did the year before. To say that books buy fewer clothes is to say that books have cheapened (clothes have become more expensive).

Summary³ “Scarce” means “desirable but limited in amount.” It is natural for economics, the science of scarcity, to assert that scarcity is pervasive. But it is also true. Desirable things are scarce either because they are limited (like sunlight in Sweden in December) or because to get them one must use up resources that could be used to produce other things (like rubber products made from crude oil). Scarcity is expressed as a budget line along which the consumer operates. The slope of a budget line measures the relative price of one item in terms of another. It measures the rate at which one item can be substituted for another. In other words, it measures the opportunity cost of one item, that is, the amount of the other item sacrificed. As the slope changes, the opportunity cost or relative price changes.

EXERCISES FOR SECTION 1.1⁴

1. *True or false:* Friendship is scarce because it is costly in time forgone making friends.
2. A sportswriter describes a certain play at second base as “the turning point” of a game. Does the description survive the idea of fungibility?
3. Notice on an elevator: “Occupancy by more than 15 people is dangerous and unlawful.” An elevator with 15 people stops at the third floor on the way up. Carl Mosk gets on and no one else gets off. Who is the criminal? Really? Couldn’t someone else have gotten off?
4. Opportunity cost is related to choice. Suppose that the world were made up of only one all-purpose good called a shmoo (that can be eaten, drunk, used for roofing, sewn into cloth, and so forth). *True or false:* In such a world shmoos could be scarce, but opportunity cost would mean nothing.
5. *True or false:* The cost of a week of vacation is simply the money cost of the hotel, the plane, food, and so forth.
6. What is the price of the item in the first column relative to the corresponding price in the second column? Be sure to express the prices in the correct units. The prices are historically accurate. You might find it interesting to reflect on what the relative prices would be today.

³ The Summary paragraphs put the gist of the section in slightly different words than the text used. Later you should read all the old Summaries in sequence. It’s a good way to review the larger structure of the book. A reminder: A good way to review the section is to answer the What to Read For questions at the beginning.

⁴ The Exercises are pretty easy problems that require you only to apply the reasoning of the sections, in the same order as it was developed. The Problems go further, and are usually more difficult and less directly linked to the section. At first you will miss even some of the Exercises, and the Problems will be impossibly difficult. Stick with it. The Exercises and Problems may seem at first “vague and unfair” to students who are accustomed to multiple-choice questions. They will find that life is vague and unfair. Might as well get used to it: The world, alas, does not come in multiple-choice questions.

Prices in England in the 1400s

First Column	Second Column
a. Bread, $\frac{1}{2}$ pence per 2-pound loaf	Bricklayer, 6 pence per day
b. Wine, $1\frac{1}{2}$ pence per quart	Ale, $\frac{3}{4}$ pence per quart
c. Bailiff, 344 pence per year	Shepherd, 300 pence per year
d. Female servant, 168 pence per year	Agricultural worker, 300 pence per year
e. Fat sheep, 12 pence	Bullock, 60 pence
f. One acre of good plowing land, for a year, 3 pence	Plumber, 6 pence per day

PROBLEMS FOR SECTION 1.1

- 1. Suppose that bread and housing must be paid for in three ways: money, ration coupons, and time spent waiting. To get, for example, a loaf of bread one must pay 10 pence, hand over 3 ration coupons, and stand in line 2 hours. To buy a house, likewise, one must spend a specified amount of money, hand over ration coupons, and spend time looking for the house.
 - a. Draw the three budget lines for bread and housing facing Steve Jones. What is the relevant budget constraint along which he can operate? (*Hint:* Steve Jones is constrained by the fixed amount of time, income, and ration coupons he has to spend. Can he use up more money than he has? Coupons? Time? What portion of the three budget lines represents this fact?)
 - b. *True or false:* As more bread and less housing are bought, the price of bread relative to housing that Jones faces rises. (*Hint:* What is the shape of the attainable area? What happens to its slope?)
- 2. William Parker buys books at a fixed price per book but buys paper at a lower price per ream the more he buys (that is, on his purchases of paper he gets a quantity discount). Draw the Books–Paper budget line that Parker faces.
- 3. Draw your budget line between Consumption Now and Consumption Later (later stands for all future years lumped together). What does the slope of this budget line represent? (*Hint:* What would you do with Consumption Now if you didn't take it now? Is it possible to get more than the present consumption in the future? Suppose you put it in a bank?)

True or False

- 4. When it is desirable, sunlight is scarce.
- 5. Another way in which to characterize scarce things is to say that they cost something.
- 6. Children are scarce only because raising them requires money expenditure by their parents on food, shelter, and school education.
- 7. A tax of \$3 per fifth of liquor will raise the price of Cheapo White Lightning relative to a very expensive liquor such as Drambuie, Grand Marnier, and Napoleon brandy.
- 8. A sales tax of 10% on the before-tax price of automobiles and groceries will leave the price (after taxes) of automobiles relative to groceries unchanged. (*Hint:* Do the arithmetic with P_A and P_G .)
- 9. As more guns are acquired at the sacrifice of butter, the opportunity cost of one additional gun (in terms of butter) falls. (*Hint:* Draw a bulging budget line. Think about what would happen in the real world to the cost of guns as more were bought.)

1.2 Income in the Budget Line

What to Read For How is income represented in a budget line? Is income in the form of particular goods (housing, food) as useful to the consumer as the same value in money? Do people with higher incomes face lower prices? In a two-good diagram, how does one represent a rise in income using a budget line? What is the income effect? What is a normal good? An inferior good? How do you express budget lines in algebra?

Income Is the Height of a Budget Line

The budget line facing a consumer has a slope that measures the price of an additional unit of one item relative to the necessary sacrifice of other items. The budget line also has a height, which measures the income of the consumer.

It is obviously desirable to have a budget line that is higher everywhere, that is, to be richer. You are richer if you can have more things: A higher budget line is desirable because it gives you more things. It is also desirable to have access to the whole of a budget line rather than to only part of it.

T or F: A family receiving charity is equally well off if it receives \$100 in cash each month as if it receives free housing costing \$100 a month.

A: False. That is, false unless the family would have consumed more than \$100 worth of housing when it had the additional \$100 cash, or unless the housing provided can be resold. The family's budget line is moved from the line marked Very Poor to Poor by the \$100 (see Figure 1.5). The entire line Poor (including the

crossed segment) is available to the family. By the gift of housing costing \$100, however, the budget line is moved from Very Poor to the uncrossed segment of Poor. The crossed segment is not available. Think about that. Something is not available that would be if the gift were in cash. The area under the budget line—the area available to be chosen—is smaller than it would be if the \$100 check were given to the family.

For this reason most economists agree that charity given directly in goods ("in kind") is worse than charity

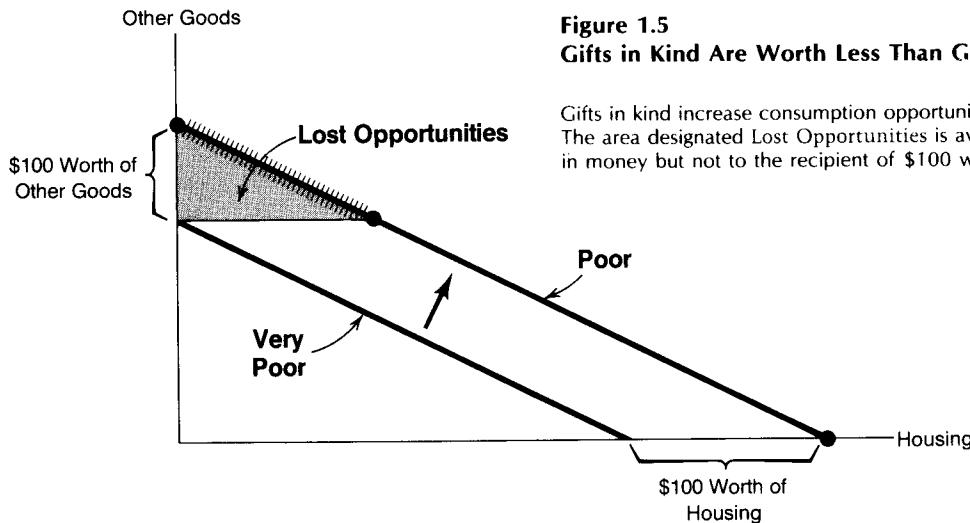


Figure 1.5
Gifts in Kind Are Worth Less Than Gifts in Money

Gifts in kind increase consumption opportunities less than do gifts in money. The area designated Lost Opportunities is available to the recipient of \$100 in money but not to the recipient of \$100 worth of Housing.

in money, supposing that the purpose of charity is to increase the happiness of the poor. Of course, public housing, public food stamps, and public medical care are not in fact meant merely to increase the happiness

of the poor. The purpose may be in fact to increase the happiness not of the poor but of the rich, by assuring the rich that poor people live in housing that is decent, or at any rate distant.

COMMENT The key to understanding most economics is to put yourself into the situation. Imagine yourself being presented with a choice between \$100 worth of an apartment or \$100 in cash. Which you would choose is a matter of common sense. Use it.

Do not confuse slopes and positions.

T or F: Cadillacs are cheaper for a professor than for a poor student.

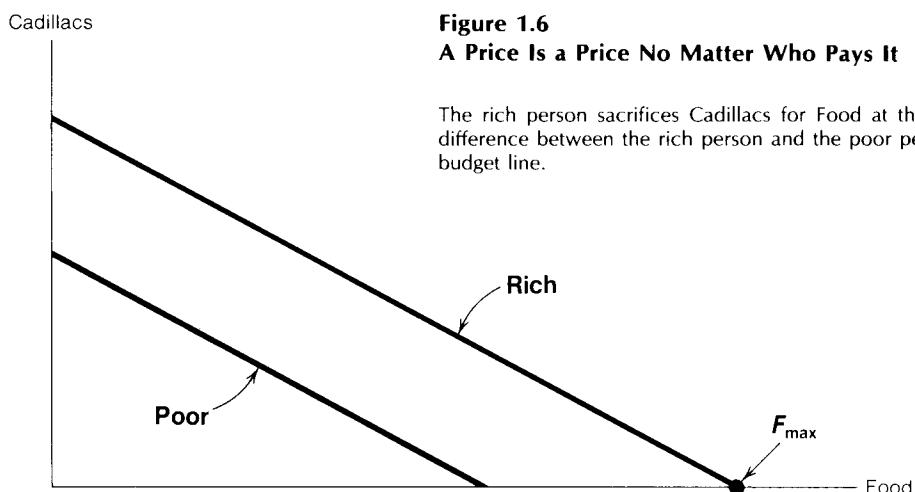
A: A professor's budget line between Cadillacs and food is higher than that of most students, but both face the same price of Cadillacs relative to food; that is, the slopes of their budget lines are the same, as shown in Figure 1.6. So, false. No one can have unlimited

amounts of everything, but a professor with \$20,000 can have more of everything than can a student with \$2000. When the professor has used fully his large but finite income he, too, faces a trade-off of one thing for another, the opportunity cost of Cadillacs in terms of food. If the two face the same prices, the professor's budget line is parallel to the student's.

COMMENT The key to answering this problem is seeing that "cheaper" means "cheaper relative to other goods, such as food." If the question gave all this explicitly it would be too easy. Too easy is bad, because real-world problems don't come in nice, easy forms.

A very clever argument might be able to show that the price *did* differ. For instance, Cadillacs take time to enjoy, and a rich man's time is more expensive. But you have to choose a reasonable level of cleverness. The problem isn't worth two days of close thinking.

The Income Effect The amounts of a good that consumers buy will change when their incomes and the prices they face change. Between 1947 and 1957 in the United States,



visits to doctors, operations performed, medicine taken, and other medical care increased 24%. Why? The theory of demand says that events such as these are to be explained by changes in incomes and prices. From 1947 to 1957 the per capita income of consumers increased in real terms (that is, after accounting for inflation) by 29%; the price of medical care relative to all other items of consumption rose 16%. The budget line changed in position and in slope, as a result of which, the theory says, the quantity of medical care changed.

The effect of an increase in income alone on the amount of a good consumed is called the *income effect*. It can be positive or negative. If it is positive for, say, medical care, an increase in income causes an increase in the amount of medical care consumed. Medical care in such a case is called a *normal good*. Normally, one expects consumers to consume more of a good when they become richer. For example, from the evidence for 1947–1957 just given, it appears that medical care is a normal good, since the quantity consumed of medical care increased as income did even though its relative price rose.

If the income effect is negative the good is called an *inferior good*. "Inferior" here is merely a technical word: It does not mean that the good in question is a bad. It means simply that as a consumer's income rises the consumer substitutes other goods. Consider Cynthia Morris, a poor meat eater who consumes only hamburger and the best steak at given prices. In Figure 1.7 she is initially at point *A*. She then receives a large inheritance from a rich aunt. That is, her budget line moves out parallel to the old one (parallel, of course, because the prices do not change). She will move from *A* to some point on her new budget line. If she moves to point *B* after the rise in her income, she has reduced her consumption of hamburger and has increased her consumption of steak. In other words, hamburger is for her an inferior good and steak a normal good.

Notice that the hamburger-steak example involves two varieties of the same commodity, namely, meat. As the consumer gets richer, she consumes more of the better variety of meat and less of the worse. A bad variety of a thing is often inferior, technically speaking, even when the thing as a whole is normal. It would not be surprising to find that, within food, hominy grits and bacon ends are inferior (sirloin steak being normal); or that, within transportation, bus and subway rides are inferior (auto trips being normal—a difficulty that enthusiasts for public transportation face); or that, within clothing, homemade cotton underwear and wooden shoes are inferior (factory-made shirts being normal). Broadly defined goods (food, transportation, clothing) tend to be normal.

T or F: Not all goods can be inferior.

A: If all goods were inferior, the amounts purchased of all of them would fall as income rose. But wait a minute. Rising income means that more of all goods could be purchased, yet less are. The consumer will evidently not be using all of her income. Not using all one's

income is unreasonable (you must define savings as a way of using income, using it for a rainy day). Since the theory of consumption is a theory about reasonableness it cannot have such odd behavior. A point such as *D* in the diagram, for which not all of income is used, is ruled out. So at least one good must increase, to use up the larger income.

The Simple Mathematics of Budget Lines

All this can be summarized with a little easy and important high school algebra. Suppose that a consumer is allocating money between candy and all other goods. Call the income of the consumer *Y* dollars. Signify prices by *P*'s and quantities by *Q*'s, with subscript *C* attached for candy and *A* for all other goods. The consumer spends all his or her income by definition, for savings is one of the

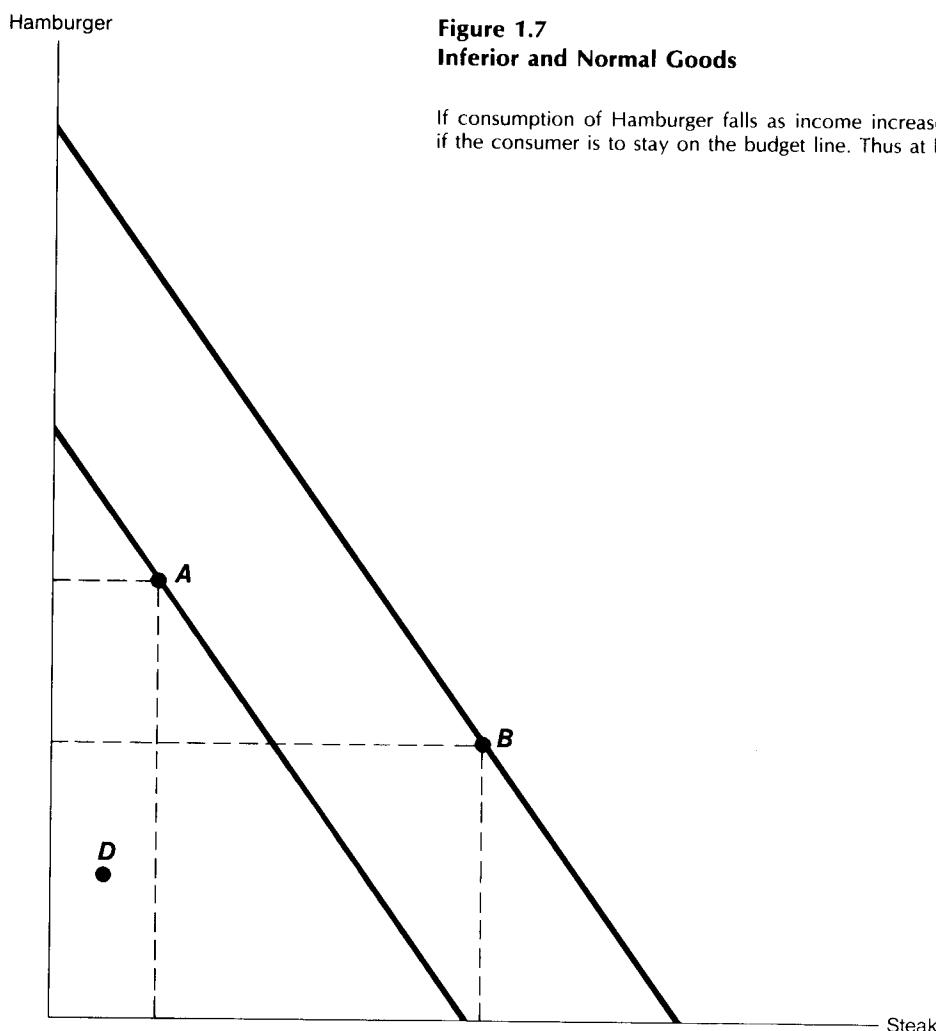


Figure 1.7
Inferior and Normal Goods

If consumption of Hamburger falls as income increases, consumption of Steak must rise if the consumer is to stay on the budget line. Thus at least one good must be normal.

other goods. Evidently, then, the sum of all the prices times quantities of the goods (that is, the sum of what the consumer spends on each good) will equal income: $P_C Q_C + P_A Q_A = Y$. This is one equation for the budget line. Alternatively, if you wanted to express it as a relationship between Q_C on the vertical axis and Q_A on the horizontal, you could solve it for the variable Q_C , to get in a couple of steps the equation

$$Q_C = \frac{Y}{P_C} - \frac{P_A}{P_C} (Q_A)$$

This is the equation for a straight line drawn on a pair of axes measuring Q_C and Q_A . The quantity Y/P_C is an intercept, which is the point where the budget line cuts the vertical (Q_C) axis. When Q_A is zero, Q_C takes on the value of the intercept. It shows the amount of candy the consumer can buy if he spends

all of his money on candy and none of it on other goods. When Q_A is something greater than zero the amount Q_C is reduced to some degree, the degree depending on the slope of the budget line, P_A/P_C . The slope is the relative price, or opportunity cost.

The theory of consumption asks how the consumer reacts to changes in the budget line, that is, to changes in the intercept and slope of the budget line. In other words, it asks what is the income effect, discussed here, and what is the relative price effect, discussed in Section 1.3.

Summary

The position of the budget line measures income. The more limited the income, the tighter the budget constraint and the lower the budget line. Algebraically, for two commodities bought in quantities Q_C and Q_A at money prices P_C and P_A with money income Y , the budget line is all points for which $Y = P_C Q_C + P_A Q_A$. When the consumer purchases a combination of the two commodities at any point on the budget line, his income is completely exhausted.

The theory of consumption asserts that changes in the quantities of a good consumed are caused by changes in incomes and prices facing consumers. An increase in income, prices held constant, is represented by an outward, parallel shift of the budget line. The effect of an increase in income is called the *income effect*. If it is positive the good in question is *normal*; if it is negative the good is *inferior*.

EXERCISES FOR SECTION 1.2

1. Your mother buys you a hideous-looking lamp for \$80 and gives it to you. In her opinion it was worth even more than \$80. You value it at only \$5 and can sell it to a junk dealer for \$20. Would you rather have had the \$80 in cash? How much richer does the gift make you?
2. If the housing given to the poor can be resold easily by them at the \$100 cost, does the gift make them as well off as a gift of \$100 in cash?
3. *True or false:* In the theory of consumption the income of consumers varies from one consumer to another; the market prices they all face do not.
4. The urban transport problem is supposed to be that people have increasingly moved from buses and subways to private cars. *True or false:* The theory of the income effect suggests that it will get worse, since average income is rising.
5. Write down and explain each step from the equation

$$P_C Q_C + P_A Q_A = Y \quad \text{to} \quad Q_C = \frac{Y}{P_C} - \frac{P_A}{P_C} (Q_A)$$

PROBLEMS FOR SECTION 1.2

1. From 1947 to 1957, real per capita income in the United States increased 29%, yet per capita consumption of clothing did not change. Because of the introduction of synthetic fabrics, the price of clothing fell 14% relative to other goods. *True or false:* On this evidence, clothing is an inferior good.
- o 2. Suppose that food and all other goods sell for money and for ration coupons. Each coupon equals one point and the two goods sell for different numbers of points. The consumer faces two budget lines, unless she can trade ration points for money. If she can trade ration points for money, she faces only one budget line. Prove this last assertion and interpret

the budget line you get verbally, graphically, and algebraically. Use the following notation: F = amount of food, A = amount of all other good (notice that Q is not used: it is often convenient to use the letter of the item for its quantity), P_A = price of all other goods, P_F = price of food, Y = money income actually spent, Y^* = money income with which a consumer is endowed, c_A = coupon points per unit of all other goods, c_F = coupon points per unit of food, C = coupon point income actually spent, C^* = income of coupon points with which the consumer is endowed. What is the full price of A in terms of F ? What is the full income of the consumer?

1.3 Prices in the Budget Line: The Law of Demand

What to Read For

What happens to the amount of a good demanded when its price changes? What is the *Law of Demand*? Do consumers do what they do out of habit? Is gold more expensive than a haircut? How do money prices and relative prices figure in the *Law of Demand*? Does the ratio of prices or the difference between prices matter to how much is consumed? Does an equiproportional increase in all prices and money income have any effect on amounts consumed? Why should you believe the *Law of Demand*? What is the *Weak Axiom of Revealed Preference*? What is a *demand curve*? How is it related to budget lines?

What the Law Says and Does Not Say

The income effect deals with changes in income, holding relative prices constant. The *price effect* deals with changes in relative prices, holding income constant. The price effect is embodied in the *Law of Demand*: More of a commodity is demanded when its relative price falls. Because the price of a hand calculator falls from \$300 to \$10, the number sold increases. Because the price of gasoline rises, the quantity demanded falls, even though drivers claim that they “need” so many gallons per week. Because the price of divorce falls, the number increases. Because the price of children rises, fewer are conceived. All this happens despite assertions that the amounts of goods consumed are determined not by price but by fashion, social pressure, advertising, impulse, whim, habit, or need.

The *Law of Demand* does not apply to comparisons between different goods. To say that more toothpaste is demanded than perfume because toothpaste is cheaper is to say nothing, certainly not the *Law of Demand*. “More” does not mean anything in such a context. Is a tube of toothpaste more or less than a bottle of perfume, or more or less than a motorcycle or a pin?

Nor does the *Law of Demand* apply to “absolute” or “money” prices, that is, dollars per quart. It applies to *relative* prices, which come in physical units, that is, so many quarts of X per ton of Y . Nowadays a quart of wine costs about twice what a quart of beer does. Relative to beer, however, wine cost about the same in 1492 as it does nowadays. It is irrelevant that a quart of wine sold for $1\frac{1}{2}$ pence in England then and about 400 pence now. What is important is that the price of wine relative to beer has not risen at all. “The price” of which the *Law of Demand* speaks is the relative price of wine to beer.

Consider the following.

T or F: A consumer buys housing and all other goods. A doubling of all the prices the consumer faces, including the price of her labor (wage), will have no effect on the amounts of housing and all other goods that she consumes.

A: True, by the logic of budget lines. Look at the picture of the consumer's initial position in Figure 1.8.

Only if the budget line changes in some way will the consumer have any interest in moving from the point chosen initially, Z. Does the budget line change

if all prices and income double? No. The slope of the budget line (P_H/P_A) does not change, because both P_H and P_A have doubled. Does it matter if they have trebled or risen 10%? No: Clearly the proportion in which both rise does not matter, as long as it is the same for both. Does this position of the budget line change? No. H_{\max} , for example, is still the point at which the budget line intersects the H axis, because both Y (income) and P_H have doubled. H_{\max} is the ratio of Y and P_H , that is, the amount of H he can buy with income Y . The ratio doesn't change if both double.

COMMENT You learned in the preceding section that a budget line such as the one between housing and all other goods, $P_H Q_H + P_A Q_A = Y$, can be written as the equation of a straight line with intercept Y/P_H and slope P_A/P_H :

$$Q_H = \frac{Y}{P_H} + \frac{P_A}{P_H} (Q_A)$$

If you understand the equation the question is simple. Since Y , P_H , and P_A have all doubled, their ratios are unchanged. So nothing happens to the intercept and slope. So the graph does not move. So no change has taken place in the consumer's situation.

Why the Law Is True: Revealed Preference

The law of demand is difficult to disbelieve. Consider a consumer of food and all other goods consuming initially bundle 0 along the solid budget line in Figure 1.9. The relative price of food rises, and the consumer, Morris, is to have constant real income: Her new budget line is the dashed one, pivoting

All Other Goods (A)

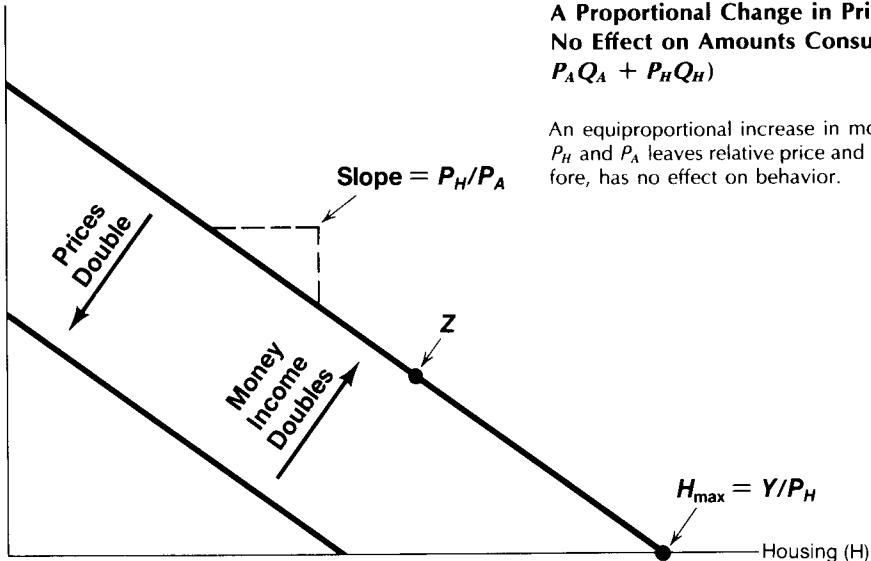


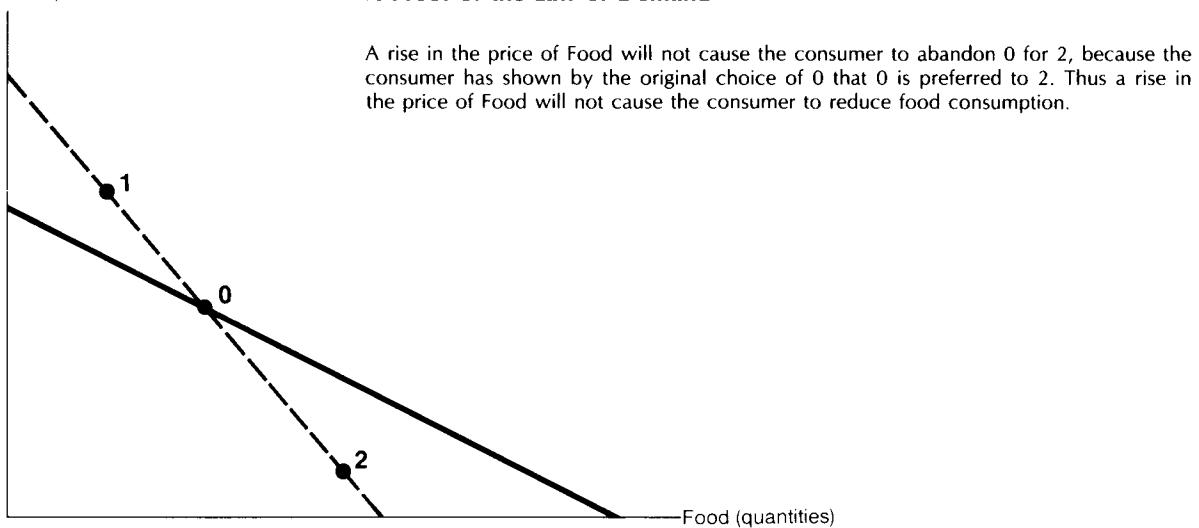
Figure 1.8

A Proportional Change in Prices and Money Income Has No Effect on Amounts Consumed (for a budget line $Y = P_A Q_A + P_H Q_H$)

An equiproportional increase in money income Y and money prices P_H and P_A leaves relative price and real income unchanged and, therefore, has no effect on behavior.

All Other Goods
(quantities)

Figure 1.9
A Proof of the Law of Demand



around 0. The question is, to which point will she move on the new dashed budget line—point 1 or point 2?

At first it is not obvious. At 1 she has more of all other goods and less of food; at 2 less of all other goods and more of food. But the introduction of a simple postulate that is consistent with the Law of Demand implies that she will always move to point 1, not 2. That is, if the relative price of food increases, real income held constant, the amount of food consumed will decrease (and the amount of all other goods will increase). The postulate, first formulated in 1938 by Paul Samuelson, is that the consumer's behavior is always consistent, in the following sense. If the consumer is presented with an opportunity to buy some bundle, call it 2, yet chooses in fact to buy another bundle, call it 0, then the consumer will never choose to buy 2 in circumstances in which 0 can also be purchased. This is called the *Weak Axiom of Revealed Preference* (WARP). It simply says that if a consumer reveals by her behavior that bundle 0 is preferred to bundle 2, she will never act later as though in fact 2 was preferred to 0. If consumers' preferences did change in this way, it would clearly be difficult to frame a theory of consumer behavior, and it would be impossible if one had no insight into why the behavior changed.

Look back at Figure 1.9. When Morris faced the solid budget line, before the rise in the price of food, she could have chosen to consume bundle 2, for bundle 2 is inside the line: Bundle 2 is in the attainable area. But she in fact chose 0. That is, bundle 0 is "revealed preferred" to bundle 2. When the consumer faces the dashed budget line, she is again able to consume bundles 0 and 2. If she in fact chose to consume 2, she is violating the WARP, for in that case she reveals by her behavior that she in fact prefers 2 to 0. Indeed, if she chose any point along the dashed budget line to the right of 0, she would be violating it. Given that the WARP holds, if the relative price of food rises (real income held constant), Morris will assume a position along the *other* part of the dashed line, such as 1. Points such as 2 cannot violate the WARP because

they were not available initially. The solid budget line indicates that bundle 1 is not affordable. Therefore, Morris's choice to consume bundle 0 does not mean that it is revealed preferred to 1 because with that budget line Morris is not able to consume 1. A consumer satisfying the WARP, in short, will always move to the left of 0 when the relative price of food rises. She will consume less food. This is the Law of Demand.

The WARP Applied

If you could have bought a bundle *A* of goods but instead bought bundle *B*, *B* is revealed preferred to *A*. You would be inconsistent if you later bought *A* in a situation where you could have bought *B*. You would have changed your mind.

Now the meaning of "could have bought a bundle *A*" is that *with the same or less expenditure as bundle B* you could have bought *A*. The idea of same or less expenditure makes revealed preference applicable:

Q: When in Italy, Lou Cain spends a monthly income of \$100 in the following way:

Price of spaghetti	\$2/kilogram
How much spaghetti Cain buys	20 kilograms
Therefore (multiplying), how much he spends on spaghetti	\$40
Price of hamburger	\$5/pound
How much Cain buys	12 pounds
How much he spends on hamburger	\$60

Now he moves to Britain, where he earns and spends \$110. But he spends it this way:

Price of spaghetti	\$3
How much he buys	22 kilograms
How much he spends	\$66
Price of hamburger	\$4
How much he buys	11 pounds
How much he spends	\$44

1. When in Italy, earning \$100 and facing the Italian prices, could he have bought the British bundle (that is, 22 kilograms of spaghetti and 11 pounds of hamburger)? That is, would the British bundle have cost his \$100 income or less?

2. Apparently, then, which bundle does he prefer?

3. When in Britain, could he have bought the Italian bundle?

4. According to the evidence of part 3, which bundle does he prefer?

5. Is Cain consistent between Italy and Britain?

A. 1. To buy the British bundle at Italian prices he would have to have spent $(\$2/\text{kilogram})(22 \text{ kilograms}) + (\$5/\text{pound})(11 \text{ pounds})$, or $\$44 + \$55 = \$99$, which is \$1 less than his \$100 income. So he could have done it but did not.

2. That he did not implies that the Italian bundle is revealed preferred to the British. He apparently likes a bundle of 20 kilograms of spaghetti and 12 pounds of hamburger better than a bundle of 22 kilograms and 11 pounds. If he had to choose between the two bundles sitting in grocery carts, he would choose the Italian bundle.

3. But wait. When in Britain he could have bought the Italian bundle, because $(\$3)(20) + (\$4)(12)$ is \$60 + 48 = \$108, which is \$2 less than his actual British expenditure. But he did not.

4. So Cain *also* prefers the British bundle to the Italian.

5. Cain is *not* consistent. He violates the WARP. At one time he prefers the Italian bundle to the British, at another the British to the Italian. If you will, he is "irrational," or at best a changed man when in sunny Italy.

COMMENT

This is the first of many long problems in the book that guide you through an extended piece of economic reasoning. They show that not all economic reasoning can be accomplished in three crisp sentences. The WARP provides a test of consistency. To put it another way, it tells you whether or not you can treat a set of data as coming from the behavior of one consistent

mind. Since a part of the theory of consumption presupposes one consistent mind, the test even has some use.

FURTHER COMMENT

Cain's inconsistency is evident in a simpler way. The price of spaghetti relative to meat rises from Italy to Britain, from $\frac{2}{3}$ pound of meat per kilogram to $\frac{3}{4}$ pound per kilogram. But Cain buys relatively more spaghetti in Britain, 22 kilograms as against 20. The price of spaghetti goes up but its quantity purchased goes up, not down. Cain violates the Law of Demand. The violation is easy to spot with two goods. Since it is not easy with many goods the elaborate arithmetic of revealed preference still has a place. But the basic point is the same: The WARP is the Law of Demand, and if one is violated so is the other.

The Law of Demand Applied

A philosopher once remarked that "not for nothing do we call the laws of nature 'laws': The more they prohibit, the more they say." What the Law of Demand prohibits—a rise rather than a fall in the quantity demanded of some thing when its relative price goes up—does not seem at first glance to be much of an insight into the laws of social behavior. At second glance, however, it is the hammer in the economist's box of tools. It reveals its full power only when applied to groups of consumers, that is, to markets, the subject of most of the rest of this book.

T or F: Simultaneous increases in the first-class postage rate from 20 to 30 cents and the overnight delivery rate from \$1.00 to \$1.10 will leave unaltered the proportion in which the two types of mail are consumed.

A: One's first instinct might be to suppose that, because both types have increased in price by the same amount, the consumption of both types would be reduced by the same amount. But the Law of Demand concerns

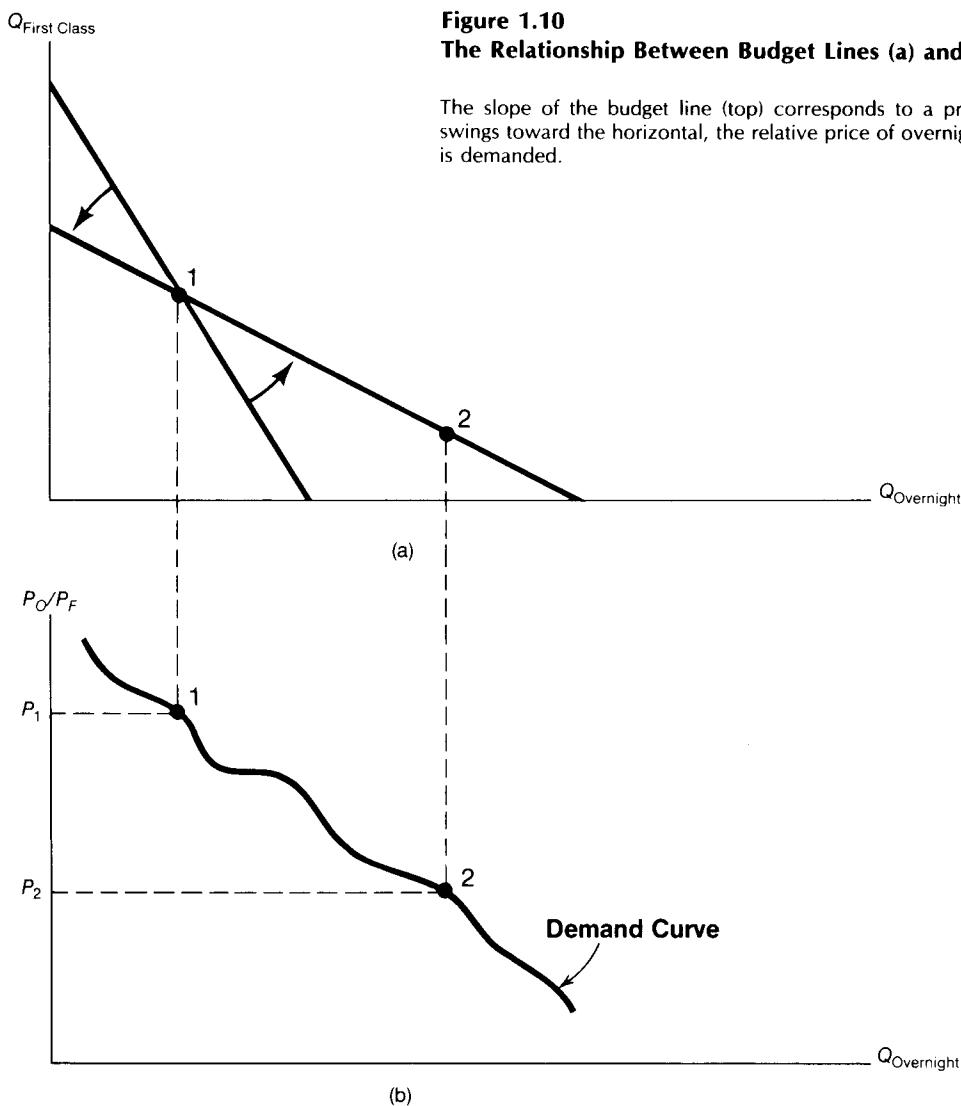
relative prices, not absolute prices, and the relative price of overnight delivery has fallen. Specifically, the relative price of overnight delivery has fallen from $(\$1.00/\$0.20) = 5$ to $(\$1.10/\$0.30) = 3.7$. More, therefore, will be consumed relative to first-class mail, even though—again by the Law of Demand—fewer mail services considered as one commodity will be consumed. Therefore, false. The proportion will change.

COMMENT

With all the hammering on the significance of relative prices in this section you should have seen that the *ratio* of prices, not their difference, is what matters to how much is consumed. The difference changes when all prices double; the ratio does not. Eventually you will learn to spot such points without the assistance of hints in the section. The way to use such a problem to learn economics is to think up a half-dozen other examples. How about the price of a movie at the university film society as against a movie downtown? Or the price of cheap as against expensive beer with the same tax per bottle?

There is another way to represent the Law of Demand that will be important later. The top panel in Figure 1.10 is the same as in earlier diagrams, plotting consumption of first-class mail against consumption of overnight mail. The bottom panel plots the *demand curve*, that is, the relationship between the price of overnight relative to first-class mail (measured on the vertical axis) and the quantity of overnight mail consumed. Notice that the horizontal axes of the two diagrams *both* measure the quantity of overnight mail.

For every diagram of the top type (in Figure 1.10), there is a diagram of the bottom type. For each budget line (having a certain slope = P_o/P_F), there corre-



sponds a price (the same P_O/P_F) along the demand curve. The Law of Demand in the top panel is that a shift in the budget line causing overnight mail to cheapen will cause consumers to consume more overnight mail. The Law of Demand in the bottom panel is that the demand curve for overnight mail slopes downward.

Now consider the following problem.

Q: During Prohibition in the United States, much to the dismay of some, alcoholic beverages continued to be produced and sold, despite the threat of confiscation,

fine, or jail. *True or false:* One would expect the average quality of alcoholic beverages to be lower than before Prohibition.

A: False. One might well expect it to be higher. If the fines and the probability of being apprehended were the same per bottle whether the stuff is Sterno or Napoleon brandy, the relative price of Napoleon brandy would have fallen during Prohibition. If the only pen-

alty were confiscation, there would be no effect, assuming again that the probability of apprehension were the same, since the penalty would be proportional to the value of the good consumed.

COMMENT This is a tricky problem, the first that has made you use economic reasoning without much guidance. Keep in mind that using it without much guidance is the goal: You'll not normally have an economist standing beside you to tell you how to think. In the present problem you have to do your own thinking, making the reasonable assumption that better booze costs more. It always helps, again, to put the abstract words ("alcoholic beverages") into concrete form ("Sterno"; "Napoleon brandy") so that you can imagine yourself in the situation of buying.

The applications of the Law of Demand are not confined to postage stamps and the other minutiae of life.

T or F: One would expect voter participation on election day to be higher in densely populated urban areas than in sparsely populated rural areas.

cost of going to the polling place therefore higher. By the Law of Demand, at the higher price of voting less would be demanded.

A: True, for polling places in sparsely populated areas would be on average farther from the voters, and the

COMMENT The operative phrase in the question is "One would expect." The Law of Demand provides in this case merely a working hypothesis, the violation of which would occasion mild surprise and a search for an explanation. It is perfectly possible, for example, that the value of time is lower in the countryside or that rural citizens are markedly more public spirited than are their urban compatriots and march to the polls in droves. The experiment is not controlled perfectly in the way that is imagined in the Law of Demand. Yet this fact does not make the Law useless. In the case of higher public spirit, for example, if rural voters turn out more heavily on election day, their public spirit (or whatever else it is that moves them) is so strong that it is offsetting the price effect. If the sensitivity of participation to distance from the polling place were known in other contexts, the higher participation of rural citizens would measure the strength of their public spirit.

Or again:

T or F: The price of higher education relative to most other goods has increased in the last 40 years. But the percentage of college-age people going to college has also increased. These observations contradict the Law of Demand.

A: False. Many other things have changed in the last 40 years, most notably the average real incomes of po-

tential buyers of higher education. As their incomes have risen, they have bought more of it (education is a normal good), even though the relative price of a college degree has risen. Once again, the Law of Demand is a statement about what will happen in the absence of countervailing forces. Without the rise in the relative price of a college degree, still more would have been demanded.

Summary

The Law of Demand says that when the relative price of something goes up the quantity demanded of the something will go down. It does not say that "the cheaper good will be demanded" (whatever "cheaper" could mean). Nor does it say that changes in dollar prices change what is demanded. *Relative* prices are what matter.

The Law can be proven by using the postulate that a consumer does not change her mind in an inconsistent way. The postulate, called the WARP, is really a definition of consistency, and can be used in its own right to think about consistency. The Law is powerful. A number of applications, for instance, make use of its key feature—that the important number is the ratio of prices (not the absolute—money—difference). The Law can be put so: The demand curve is downward sloping.

EXERCISES FOR SECTION 1.3

1. *True or false:* Gold is more expensive than a haircut.
2. Show that the budget line does not change if the prices of each of *three* goods and income increase by 1%. Think of a concrete example, with one of the goods being All Other Goods.
3. Suppose that oranges cost the same to ship to the Northeast whether they are high quality or low quality. *True or false:* The price of high-quality oranges *relative* to low will be greater in the Northeast than in Florida or California, and the Northeast will demand worse oranges.

PROBLEMS FOR SECTION 1.3

1. Roderick Floud consumes only food and clothing, initially in these monthly amounts and at these prices: food: 60 pounds at \$1 per pound; clothing: 20 yards at \$2 per yard. First, some trivial preparations:
 - a. What is Floud's money income in dollars?
 - b. What is the share of food in his total money expenditure (= income)?
 - c. What is the share in (b) multiplied by -10% , a 10% fall?
 Second, the substantive question:
 - d. Floud's real income in terms of his initial bundle is Y , his money income, divided by $P_F \cdot (60 \text{ pounds}) + P_C \cdot (20 \text{ yards})$, a price index with weights on the prices P_F and P_C equal to the initial amounts of the two goods (60 pounds and 20 yards). When P_F and P_C are at their initial values (namely, $P_F = \$1$ per pound and $P_C = \$2$ per yard), this quotient is 1.0 (satisfy yourself that this is true). Suppose that P_F alone changes, rising by 10% from \$1.00 to \$1.10. If money income is held constant, what happens to the quotient? Calculate it with the new value of P_F . What is the percentage change in the quotient?
 - e. Notice that the answer to (d) is the same as the answer to (c). What general rule do you suppose holds for small price changes among the percentage change in a price, the share of the item in total expenditure, and the percentage change in real income?
 2. A student of economics went one night to Jimmy's Bar where beer and whiskies both cost \$1.00 and drank six beers and four whiskies. The next night, on the eve of the final examination, the student went to the Eagle Bar where beers cost \$0.50 and whiskies \$1.50 and drank two beers and six whiskies. *True or false:* The student was too irrational to pass the final examination.

- 3. In 1958 underground coal miners in Italy spent 803,000 lire on their purchases and underground coal miners in Germany spent 7330 marks. The Italian bundle would have cost 6270 marks in Germany; the German bundle would have cost 1,139,000 lire in Italy. *True or false:* With these data one cannot reject the hypothesis that Italian and German coal miners have the same tastes. (*Hint:* Is one bundle revealed preferred to the other? Now is the other, too?)

4. Consider the following data on purchases of coal miners in France and the Netherlands:

**Value of Purchases by
Coal Miners at Local and
Foreign Prices (local currency
units, in thousands)**

At Prices in:	Bundles (quantities) in:	
	France	Netherlands
France	919	990
Netherlands	6.84	6.56

True or false: According to these data the French bundle is superior.

5. What are the rankings of bundles in the following data? Is the WARP ever violated? What do you conclude about appealing to differences in national tastes to explain differences in consumption bundles? (*Hint:* It is possible that some of the bundles will be unranked relative to others: equal, as it were.)

**Value of Consumption at
Local and Foreign Prices
for Underground Coal
Miners, Married with Two
Children, for Six Localities
in the European Coal and
Steel Community, in 1958
(local currency, in thou-
sands)**

At Prices in:	Value of Bundles Bought in:					
	Germany	Belgium	France	Italy	Netherlands	Saar
Germany	7.33	9.85	8.99	6.27	9.19	8.43
Belgium	88	107.8	101.8	70.6	104.1	100.4
France	828	1031	980	625	990	925
Italy	1139	1462	1246	803	1387	1278
Netherlands	5.59	7.14	6.84	4.87	6.56	6.35
Saar	835	1082	959	649	1023	935

6. Bloomingdale's is an expensive and fashionable department store in New York; Woolworth's is an inexpensive and unfashionable one. *True or false:* A rise in the cost of parking in New York would favor Bloomingdale's.

The Consumer's Choice

2.1 The Pursuit of Happiness

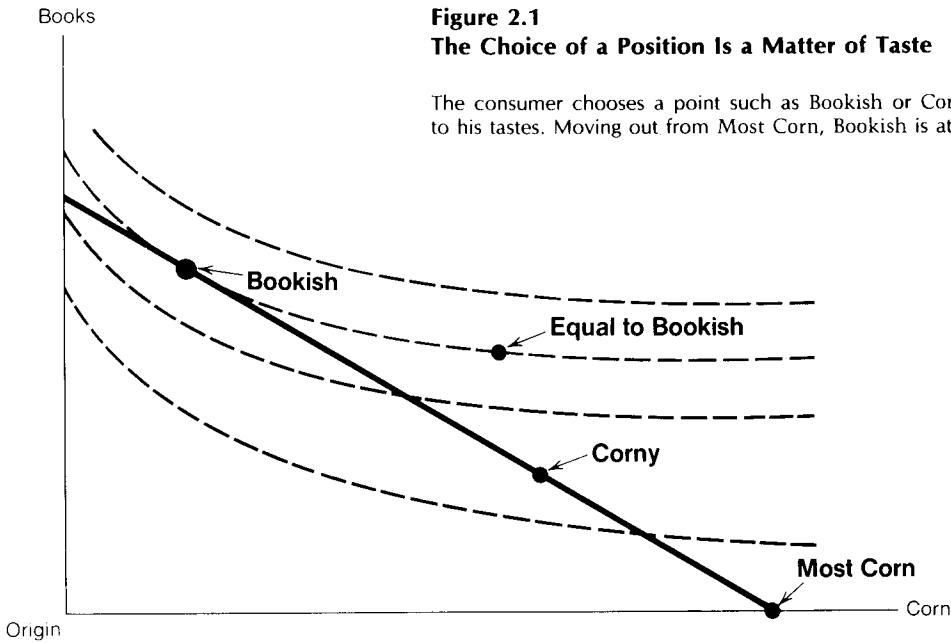
What to Read For How does a consumer choose between different combinations of goods? What is the *hill of utility?* Indifference curves? How do you represent people with preferences for one good over another? How do the indifference curves look between a good and a "bad"? How are indifference curves useful for thinking about the nation's choices?

A consumer, then, has a budget line. The budget line constrains him to certain possibilities—such is the world of scarcity we inhabit. But he still has choices to make along the budget line. He could choose much of one good (and then of course he would be choosing little of the other). Or he could choose much of the other.

The Consumer Has Tastes and Tastes Are Contour Lines on a Hill

Obviously, a consumer chooses a particular point on his budget line because he likes it better than other points available. Consider a studious farmer, Richard Zecher, who consumes only corn and books. Figure 2.1 shows Zecher's budget line between the two (ignore the dashed lines for a moment). If he consumed only corn he could get at most the point Most Corn. If he likes books a lot he will choose the point Bookish; if he likes corn a lot he will choose the point Corny.

So much is pretty obvious. The notion of "choosing what he likes" can be sharpened by thinking of Zecher getting happiness, or *utility*, from various combinations of books and corn. The more of both goods he gets, clearly, the better off he is. His choices, or *preferences*, about books and corn and the amount of happiness he gets from any bundle of books and corn can be represented as a *hill of utility*. The higher up the hill, the better. Think about that. The hill has no peak if more books and corn are always good. Think about that, too. The dashed lines in the diagram are contour lines on the hill, like contour lines on a map of a real hill. Along a contour line combinations



such as the points Bookish and Equal to Bookish have equal utilities—equal “heights” up the hill as it were. That is, the consumer would be equally happy with the combination of books and corn at Equal to Bookish as with the combination at Bookish. He doesn’t care which he gets. He is indifferent between them. For this reason the dashed contour lines are usually called *indifference curves*.

The point of all this is that the indifference curves show everything one needs to know about a consumer’s taste. As the curves are drawn, and with the budget line he has, Zecher chooses the point Bookish. Bookish is the highest point on his hill of utility that he can reach, considering that he has to stay on or below his budget line. Bookish is his optimal, best, equilibrium point. Great.

An important feature of the point Bookish is that at this point an indifference curve *just touches* (“is tangent to”) the budget line. Think of the budget line as a fence built across the side of the hill beyond which he cannot afford to go. Imagine Zecher setting out at the point Most Corn to find the highest point on the hill that he can attain. He walks up along the fence, crossing successively higher contour lines until he reaches Bookish. At that point if he kept going he would start moving back down the hill and cross successively *lower* lines. So he stops, looking longingly through the fence at the still better but unattainable points beyond. Yet he can comfort himself with the thought that he has reached the best point he can. He has made himself as well off as he can possibly afford. And the contour line at the best point just touches the budget line.

Tastes can vary from one person to another, so that different people may well have different indifference curves. Therefore, people with the same budget line might well pick out different points of equilibrium. If Zecher himself were less of a book lover his indifference curves would not be crowded so close to

the books axis, and he might choose some less bookish point as the best he could do, such as *Corny*.

Zecher would also choose a different point if his budget line changed. For instance:

Q: If the price of books rises, Zecher will choose a point somewhere to the right of the point *Bookish*.

A: His budget line will rotate in a counterclockwise direction, because for any given amount of corn given up he will get fewer books. Use your pencil as the budget

line and try out different positions. Zecher will choose an equilibrium somewhere to the right of *Bookish*. That is, the pencil will just touch (be tangent to) an indifference curve at a point off to the right. You will see that the amount of books bought will fall: the Law of Demand once again.

COMMENT

To be successful with diagrammatic problems you have to get right inside them, using your pencil and your hands to represent the curves, talking to yourself as you try out different positions. At first this will be hard to do. Eventually, though, you will grow familiar with the diagrams typical of economics. At that stage you will be able to use the diagrams for thinking. Thinking is easier with diagrams because the diagrams silently limit the possibilities you have to think about.

The framework of budget lines and indifference curves, then, splits the consumer's decision into opportunities and tastes, into *Can* and *Want*. Differences over how to spend the family's money between a husband and wife, for instance, are pure differences in tastes. Presumably both agree on the budget line they face—that is, what they can afford to buy. The only reason he wants to spend more on the car and she on the house is the difference in their indifference curves. Knowing this, they can settle down to persuading the other to change his or her tastes, leaving arguments about affordability aside.

Consider, on the other hand, differences over how to spend the nation's resources to do as well as possible in providing Domestic Prosperity and Security from Our Enemies. It is perfectly possible that Democrats and Republicans, say, agree entirely on tastes, yet disagree on the opportunities that face the nation. They might have the same indifference curves between the axes of Prosperity and Security, as in Figure 2.2, but different ideas about what the budget line is. Democrats, as shown, may think that in fact the nation gives up a lot of prosperity by spending more on bombs and soldiers; the Republicans may think that it gives up little. The upshot will be contrasting policies. The Policy of Democrats will have more prosperity than security relative to the Policy of Republicans.

The merit of thinking in this way is that it makes clear what things exactly we agree and disagree about. If Republicans and Democrats realized that their underlying tastes are similar, they could focus their attempts to persuade each other about the actual shape of the budget line. The persuasion would be factual. If they agreed on the facts but still disagreed on the policy, the differences would apparently be matters of taste, that is, issues of what is good. The persuasion could then be moral.

When the two items being chosen are both *goods* the indifference curves slope *down* across the diagram. As he moves along a particular utility contour, the consumer maintains the same level of happiness as he gives up some of

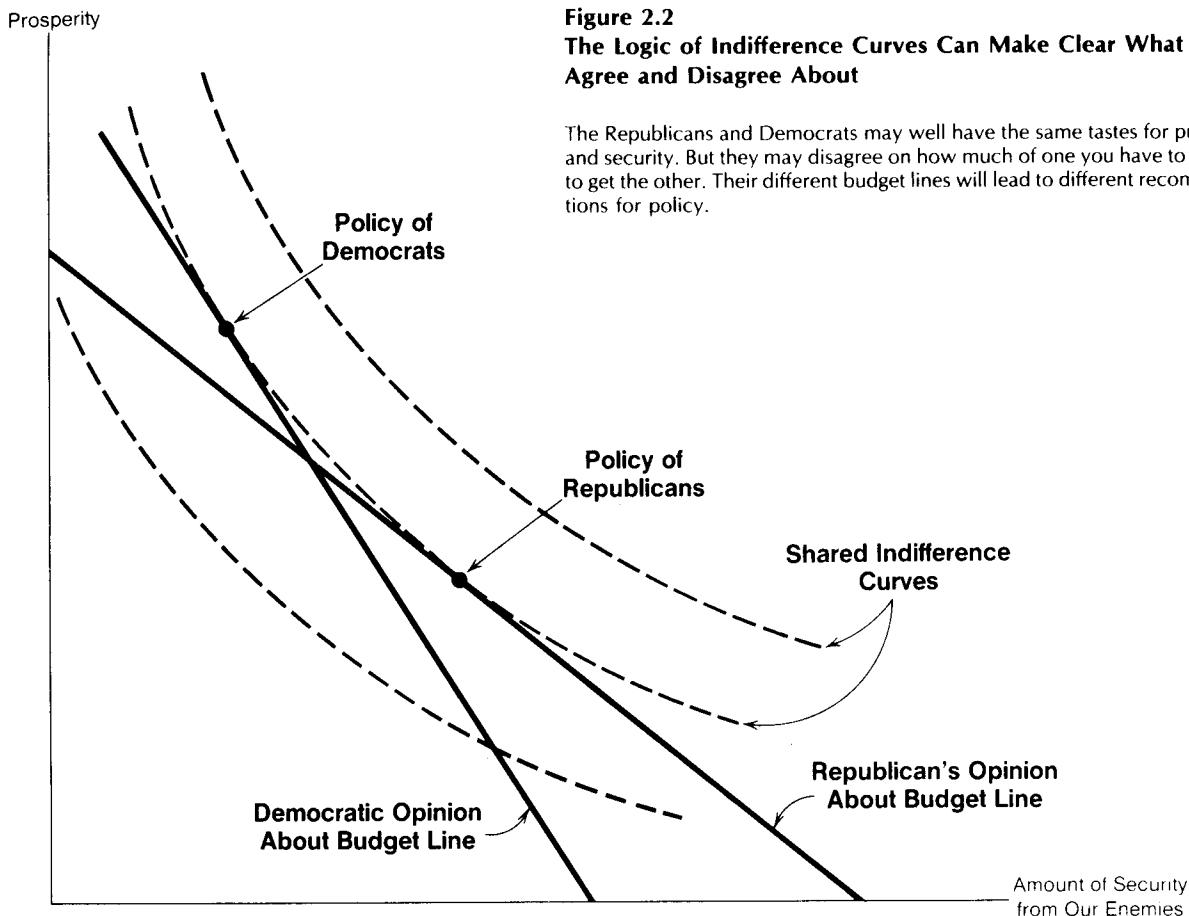


Figure 2.2
The Logic of Indifference Curves Can Make Clear What We Agree and Disagree About

The Republicans and Democrats may well have the same tastes for prosperity and security. But they may disagree on how much of one you have to sacrifice to get the other. Their different budget lines will lead to different recommendations for policy.

one good but acquires some of the other good in its place. One gets happier—one gets to higher contours—by moving upward and to the right. But the analysis is easily adapted to considering the choice between a good and a *bad* connected by scarcity. For instance, suppose that more cars produce more pollution. Removing pollution is desirable, but can only be done by sacrificing some cars. Look at Figure 2.3. The shaded area is the set of opportunities the society faces: It can't have the point Nice but Not Available. The indifference curves in such a case slope upward, which says that we are better off with little pollution and lots of cars. The hill climbs upward and to the left, not to the right. The point Optimal is therefore not the best combination of cars and pollution conceivable, but it is the best we can do considering that having more cars causes more pollution.

The idea of an “optimal amount of pollution” strikes noneconomists as silly. It is not. Less pollution is desirable; but if having less pollution is possible only with fewer cars, then at some point less pollution is *not* desirable.

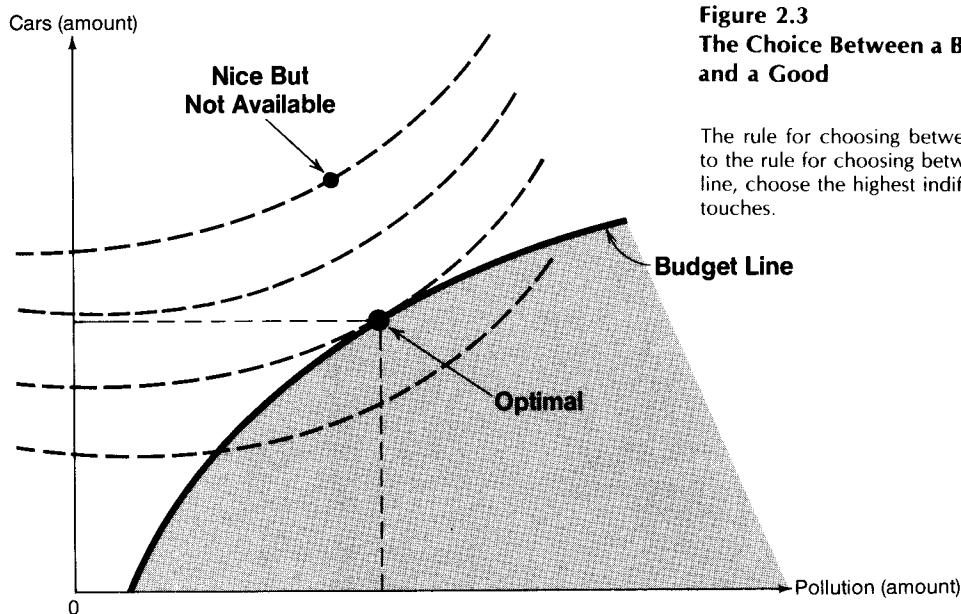


Figure 2.3
The Choice Between a Bad That Is Costly to Remove and a Good

The rule for choosing between a good and a bad is identical to the rule for choosing between two goods. Given the budget line, choose the highest indifference curve that the budget line touches.

Summary The income and prices that consumers face limit their choices, but within these limits the exact amounts of goods (or bads) they choose are a matter of taste. A consumer's taste for two goods such as guitar lessons and beer can be described as a hill of utility. The consumer chooses the highest contour in the beer-lessons plane that he can reach on his budget line, that is, the point at which a contour line is tangent to the budget line. This framework is applicable to any choice, such as the public choice between schools and highways, defense and courts, or automobile rides and air pollution.

APPENDIX TO SECTION 2.1 The Calculus Solution of the Problem¹

Mathematically speaking, the farmer's problem, as formalized by the economist observing him, is to choose the amounts of books, B , and corn, C , to maximize his *utility function* (a function of B and C) subject to the constraint that the combinations of B and C chosen satisfy (that is, lie on) his budget constraint (money income $Y = P_B B + P_C C$, supposing that his expenditures on books and corn exhaust his income). To solve this problem one needs to know the utility function. A simple case is $U = BC$; that is, the level of utility (happiness) of the farmer is the product of the amount of books and the amount of corn he consumes. With 10 books and 3 bushels of corn, his utility would be 30 units; with 2 books and 20 bushels of corn, 40 units; and so forth.

¹ There will be occasional Appendixes to the sections giving mathematical treatments of the material. If you have not had a year of calculus you will not be able to follow them. Don't fret; they are not essential. Incidentally, anyone serious about economics (or about being an educated person) should take calculus. For advanced study it is essential.

Brute Force Solutions For any particular values for Y , P_B , and P_C in the budget constraint, the problem can be solved graphically. Plot the budget constraint on a graph of B and C ; plot a utility contour on the graph for one chosen value of U (say, $U = 10$), then for another, and another, and another; keep plotting the utility function until you find one that is just tangent to the budget line; then read off the equilibrium values of B and C . An alternative algebraic method of trial and error would be to choose a pair of B 's and C 's that is on the budget line by choosing at random a value of C and inserting it into $(Y/P_B) - (P_C/P_B)(C) = B$ to get the corresponding value of B ; from the utility function $BC = U$ calculate the utility for this pair; then increase C a little, recalculate the B from the budget line, recalculate U , and see whether U rises or falls; if it rises, continue increasing C ; if it falls, decrease C in the next trial. And so on, until either you have discovered the maximum attainable U or the battery in your calculator has gone dead.

An Elegant Solution with Calculus

For such a simple utility function, however, first-term calculus solves the problem quicker and gives more insight into the solution. The budget constraint must be satisfied always. To ensure that it is, solve it for one of the goods, for example, books, and substitute this expression for B into the utility function:

$$U = \left[\frac{Y}{P_B} - \frac{P_C}{P_B} (C) \right] (C) = \frac{Y}{P_B} (C) - \frac{P_C}{P_B} (C^2)$$

Geometrically, this substitution amounts to slicing the three-dimensional surface of utility along the budget line. The equation for U now shows how utility rises and falls as the amount of corn is varied along the budget slice. Because it now embodies the way in which B depends on C along the budget constraint, the utility function has been made into a function of C alone. Y , P_B , and P_C are given numbers, and C alone is variable. To maximize U , set its derivative with respect to C equal to zero:

$$\frac{dU}{dC} = \frac{d}{dC} \left[\frac{Y}{P_B} (C) - \frac{P_C}{P_B} (C^2) \right] = \frac{Y}{P_B} - 2 \left[\frac{P_C}{P_B} (C) \right] = 0$$

The optimal C , given Y , P_B , and P_C , is simply the solution of the last equality:

$$C = \frac{1}{2} \left(\frac{Y}{P_C} \right)$$

This, incidentally, is the demand curve for C , though a very strange one for any actual good. The optimal B can be derived in the same way, and naturally, because the utility function here is symmetrical, the form is similar to the expression for C , $B = \frac{1}{2}(Y/P_B)$; or by substituting the expression for C back into the budget constraint

$$B = \frac{Y}{P_B} - \frac{P_C}{P_B} (C) = \frac{Y}{P_B} - \frac{P_C}{P_B} \left[\frac{1}{2} \left(\frac{Y}{P_C} \right) \right] = \frac{Y}{P_B} - \frac{1}{2} \left(\frac{Y}{P_B} \right) = \frac{1}{2} \left(\frac{Y}{P_B} \right)$$

If income were \$100, the price of books \$2 per book, and the price of corn \$1 per bushel, the equilibrium combination would be

$$B = \frac{1}{2} \left(\frac{\$100}{\$2} \right) = 25 \text{ books}$$

$$C = \frac{1}{2} \left(\frac{\$100}{\$1} \right) = 50 \text{ bushels}$$

To summarize, then, the geometry of choosing the best bundle along the budget line—that is, choosing the one point at which the budget line is tangent to a contour line, can be put mathematically as follows. Choose the bundle to maximize the utility function subject to the constraint that expenditure equals income. First-term calculus suffices to solve the problem—at any rate if the utility function is a simple one and if only two goods are involved.

EXERCISES FOR SECTION 2.1

1. Describe the effect on Zecher's choice of a combination of corn and books of:
 - a. A fall in the price of corn.
 - b. A fall in Zecher's income.
 - c. A decline in his desire to read.
2. Much of what university professors do has nothing directly to do with their teaching. They do research, attempting to expand our knowledge of economics or of the past or of the chemistry of plastics. Show in an indifference curve diagram:
 - a. The indifference curves between teaching and research of the typical professor as contrasted with the typical student.
 - b. The budget line between teaching and research.
 - c. The budget line that would hold if it is true, as professors commonly say, that more research produces better (and therefore more) teaching.

PROBLEMS FOR SECTION 2.1

1. Do animals obey economic laws? Describe an experiment to discover whether rats obey the law of demand and other pieces of consumer behavior, using the following ideas. Rats like food and water; each good can be given a price in right- or left-lever pushes; income can be N pushes a day. Relate the laboratory conditions to the theory of consumption item by item.

True or False

- 2. Because more is preferred to less, the hill of utility has no overhanging cliffs. (*Hint:* Draw the contours for an overhang. You will have to imagine the hill clearly.)
- 3. If consumers are consistent and more is preferred to less, then indifference contours on the hill of utility cannot cross. (*Hint:* Draw two contours that do cross; compare points with more of both goods on the two, showing that the conclusion violates consistency.)
- 4. The demand curves for B and C derived in the appendix to this section satisfy the Law of Demand.
- 5. The utility functions $U = \alpha + \beta(BC)$, where $\alpha + \beta$ are constraints and β is positive, or $U = (BC)^2$ or $U = \ln(BC)$ give exactly the same demand curves as does $U = BC$.

2.2 The Shape of Indifference Curves

What to Read For

What does an indifference curve usually look like? What are the extremes within which the shape of a typical indifference curve must fall? What does an indifference curve between two perfect complements look like? Between two perfect substitutes? When two

goods are complements, does the pricing of either good alone matter to the consumer? Can the prices of substitutes vary relative to each other? What is the marginal rate of indifferent substitution, or the marginal valuation? How does it vary along an indifference curve?

Complements

The shape of indifference curves is sometimes apparent from the types of goods. For example:

Q: Draw the indifference curves between left and right shoes.

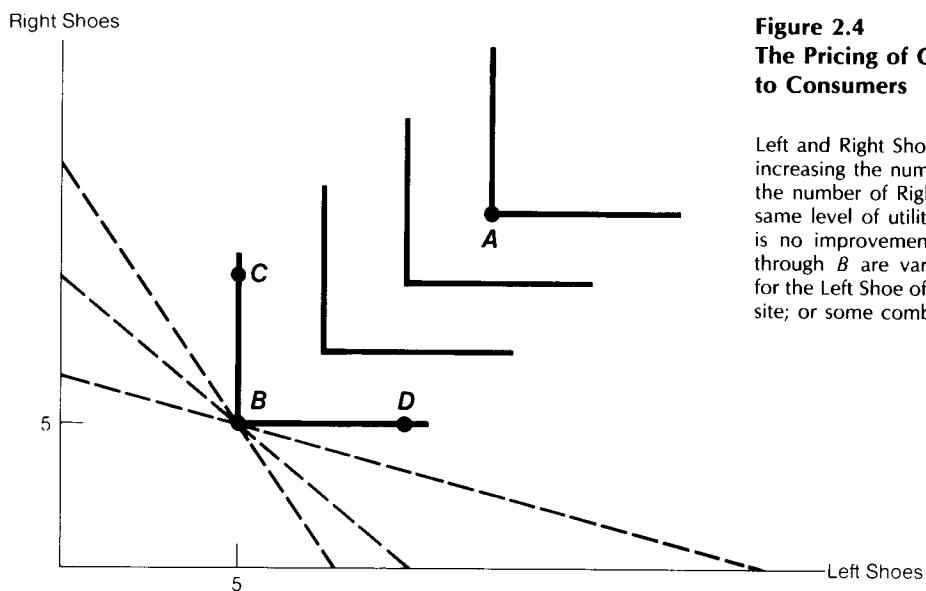
A: Start at some point in a diagram with left shoes measured on the horizontal axis and right shoes on the vertical. A natural starting point would be five pairs of shoes: five left and five right. Now imagine moving away from the point—moving, say, straight up, getting

more right shoes *but no more left shoes*. Will utility increase? Obviously, no; five pairs plus a few odd right shoes is no better than five pairs alone. In other words, the vertical move keeps the person on the same indifference curve. The same is true of an exactly horizontal move, getting more left shoes but no more right. A typical indifference curve between left and right shoes is an L-shape, as in Figure 2.4.

COMMENT Once again, you need to get yourself inside the diagram to understand it. Imagine yourself with five pairs getting some odd left shoes.

The same is true of other goods that come, so to speak, in pairs: bread and butter, housing and furniture, autos and gasoline. In these cases, however, the L-shape is not perfectly square. You do not require exactly one dining room table per house, and even if you have one it can be big or small. There is some flexibility. The L-shape is rounded a little at the edge.

Now consider how sellers would set the price of complements, supposing



**Figure 2.4
The Pricing of Complements Does Not Matter to Consumers**

Left and Right Shoes are perfect complements. That is, increasing the number of Left Shoes without increasing the number of Right Shoes leaves the consumer at the same level of utility. Moving from B to either C or D is no improvement. The various dashed budget lines through B are various different pricing schemes: \$30 for the Left Shoe of a pair, \$1 for the Right; or the opposite; or some combination. All lead to B. None matter.

that they sold both. The case that makes the point clearest is the *perfect complements*, left and right shoes:

Q: Does it matter how much one half of a pair of shoes is priced, as long as the pair sells for \$15.00?

A: As long as the price of a pair of shoes does not change, it clearly does not matter to a consumer how each shoe is priced. At \$15.00 a pair the consumer will think it odd but unobjectionable if a lunatic shoe-seller sells left shoes at \$14.99 and right shoes at \$0.01. The buyer of one pair makes the same sacrifice of income to acquire a pair, which is to say diagrammatically that the budget lines for the various relative prices of left and right shoes (all adding up to \$15.00) run through the same bundle, as in Figure 2.4. Shoe-sellers often display left

shoes in unsupervised street stalls, pricing them, as it were, at the low cost of stealing them. But to step out with any grace you must step in and pay up.

The pricing of complements is not always so innocent. Computer cards used in programming big computers are complements with the services of the computer. It might appear that it does not matter whether IBM takes its just reward in high prices for cards or high prices for the computer itself. As the discussion in Chapter 18 will show, however, a high price for cards is a way of extracting more money from heavier users of the computer.

Substitutes At the other extreme are substitutes. One eats bread *and* butter (complements) but butter *or* margarine (substitutes). Complements are consumed together. Substitutes are consumed in place of one another.

Q: Draw Dudley Baines's indifference curves between bottles of 1967 Château Latour wine with odd serial numbers and bottles with even serial numbers.

A: Start, again, with some particular bundle, say 10 bottles odd-numbered and 10 even-numbered. To find the shape of an indifference curve through the bundle, imagine a move away from the initial bundle that would

leave Dudley Baines just as happy. He would not care whether he had an odd- or even-numbered bottle. So a one-for-one trade would leave him just as happy. And it would always do so, no matter what proportion of odd- and even-numbered bottles he acquired. Therefore, a straight line of unchanging slope 1.0 is the indifference curve through the initial bundle (see Figure 2.5).

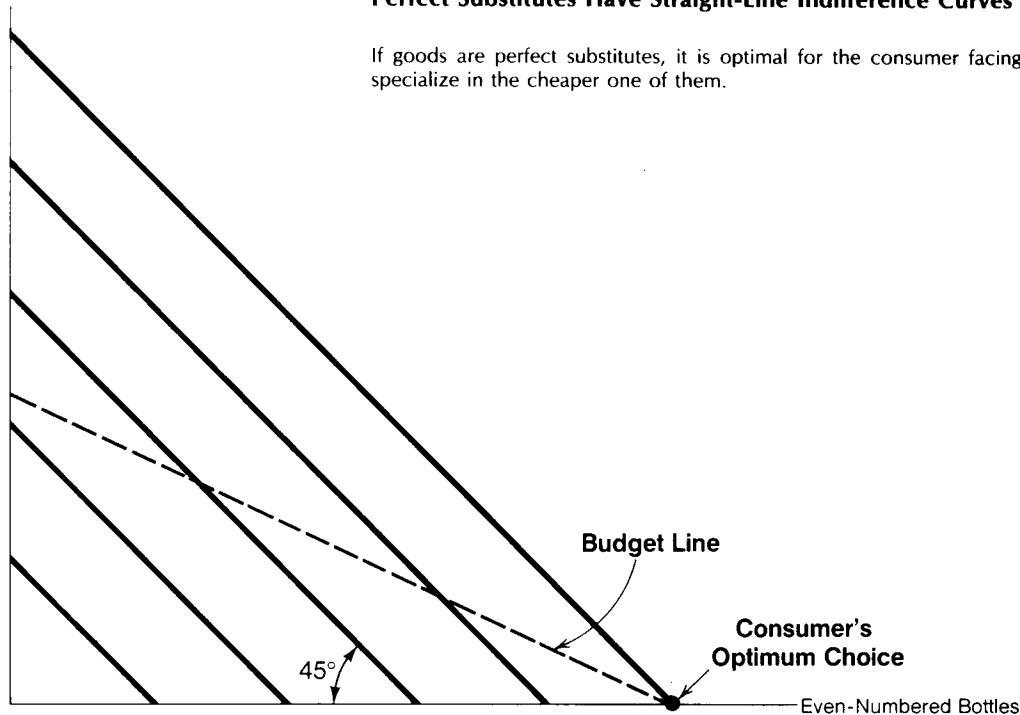
If the relative price of even-numbered bottles in terms of odd falls below 1.0 (as along the dashed budget line), so that the price of even-numbered bottles is less than the odd-numbered bottles, Baines will buy only even-numbered bottles. Because there are many such Baineses who would drive a low price for one type back up, the prices of the two types must stay equal. This holds for less perfect substitutes, too. The prices of domestic and foreign goods, for example, move together, as do those of butter and margarine, felt-tipped pens and ball-points, or one book on price theory and another.

The slope can be different from 1.0 if the differences in absolute prices is due to quality differences. Different grades of iron, for instance, are perfect substitutes at a price reflecting their different qualities. From 1891 to 1896, for example, the price of number 1 foundry pig iron at Philadelphia fell from \$17.52 a long ton to \$12.95. The price of gray forge pig iron made from Lake Superior ore at Pittsburgh was always about 20% cheaper, falling from \$14.06 to \$10.39. At a slope of 1.20, in other words, the two were close enough substitutes to be viewed indifferently. If you sold gray forge pig iron to him for

Odd-Numbered
Bottles

Figure 2.5
Perfect Substitutes Have Straight-Line Indifference Curves

If goods are perfect substitutes, it is optimal for the consumer facing constant prices to specialize in the cheaper one of them.



20% less than number 1 foundry pig iron, an iron or steel maker would not care which he used.²

Deciding whether two goods are complements or substitutes depends on who uses them for what. The carpenter's hammer and the bricklayer's trowel are unrelated—neither substitutes nor complements—for the individual tradespeople. They are, however, complements for a building project, on which so many hammers and so many trowels will be in use on average. Sometimes it will be obvious what the pair is used for, such as different grades of coal. At other times the observer needs to supply the context.

The Usual Shape Is Convexity

The usual shape of indifference contours for consumers as a group, however, is *convex* to the origin, that is, bowed toward the origin. (The opposite of convex is *concave*, which you can remember by noting that curves with a hollow—a cave—opening toward the origin are *concave*.)

² The two were not in fact used for exactly identical purposes, though there was enough overlap to keep their prices linked. The discount for forge pig iron seems to decline after 1896. The prices are from Peter Temin, *Iron and Steel in Nineteenth-Century America: An Economic Inquiry* (Cambridge, Mass.: MIT Press, 1964), pp. 284–285. That prices of two goods move together is necessary if they are close substitutes in consumption. But prices can move together without the goods being close substitutes: Both may be affected by the same weather, for instance, or some other common influence on the cost of producing them.

The convexity of indifference curves can be expressed in several different ways. The slope of an indifference curve between food and clothing is called the *marginal rate of indifferent substitution* between food and clothing or the *marginal valuation* of food in terms of clothing. It is the minimum amount of clothing that a consumer would accept in exchange for giving up a unit of food—the minimum amount being, of course, the amount that just barely keeps the consumer on the same indifference curve and therefore at the same level of utility.

Look at Figure 2.6. If the consumer got more or less than ΔC in exchange for ΔF , the consumer would arrive at a point such as Better or Worse (than the initial point Start). Convexity can be expressed so: Moving along any given indifference curve in the direction of increasing food and decreasing clothing will *decrease* the marginal valuation of food. This way of putting it is plausible for the same reason that the Law of Demand is plausible. In fact, it *is* the Law of Demand. The more food you have relative to clothing, the less you will value the prospect of still more food.

Summary

The usual shape of indifference curves is bending in (convex) to the origin. The limiting cases for two goods are perfect complementarity—right angles—and perfect substitutability—straight lines. For complements such as furniture and housing a change in the relative price at a given real income has no effect on the amounts consumed; for substitutes such as Italian-made shoes and American-made shoes a change in the relative price has a large effect. These limiting

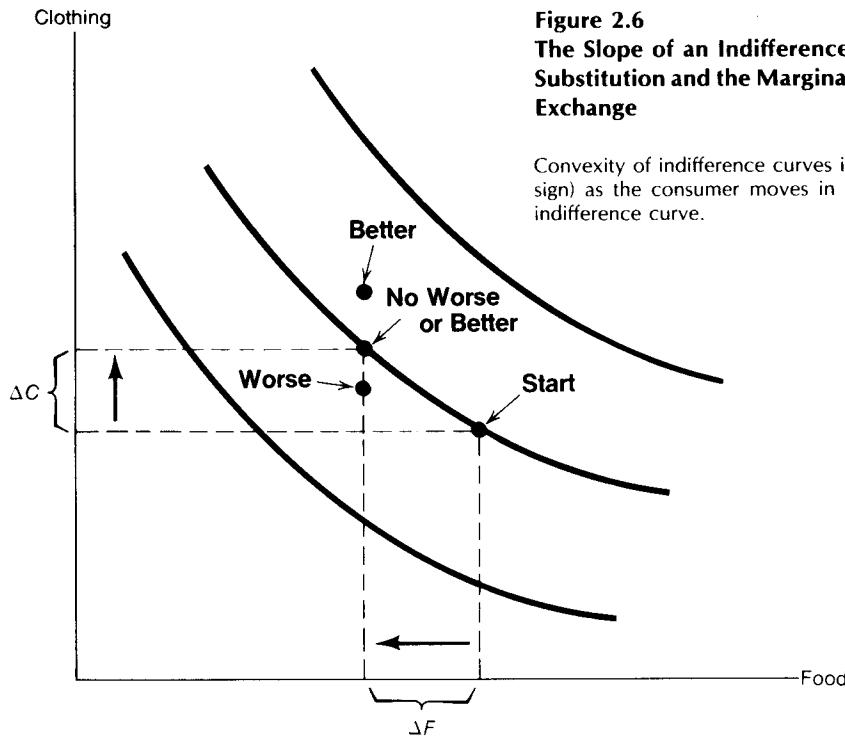


Figure 2.6

The Slope of an Indifference Curve Is the Marginal Rate of Substitution and the Marginal Valuation Is the Minimum Acceptable Exchange

Convexity of indifference curves implies that $\Delta C / \Delta F$ rises (ignoring its negative sign) as the consumer moves in the direction of more clothing along a given indifference curve.

cases do occur. But generally, people neither consume goods in rigid proportions nor specialize in consuming one good. To repeat the usual case in alternative language, the marginal rate of substitution usually declines, and the point of rest is usually *interior*, within and not on the axes.

EXERCISES FOR SECTION 2.2

1. Which of the following are complements and which substitutes? Explain.
 - a. Plate glass and bricks.
 - b. Structural steel and reinforced concrete.
 - c. Wine and beer.
 - d. For making water, molecules of hydrogen and oxygen.
 - e. Domestic autos and imported autos.
 - f. Persian Gulf crude oil and southern Illinois coal.
 - g. Paper and pencils.
2. Suppose that Carl Mosk is shipping wheat from Scott City, Kansas, to Philadelphia over two railroads, the first leg to St. Louis and the second from there to Philadelphia. As long as the total freight cost is the same, does he care how it is divided up between the two railroads? (*Hint:* Are the two legs complements?)
3. At what price, apparently, did British consumers of grain 1771–1775 treat barley and oats as good substitutes? Explain.

Year	Price of Barley (shillings per 8 bushels)	Price of Oats (shillings per 8 bushels)
1771	26.4	17.2
1772	26.1	16.7
1773	29.2	17.7
1774	29.3	18.3
1775	26.8	17.0

PROBLEMS FOR SECTION 2.2

True or False

1. If Hank Gemery's utility contours between clocks and peanut butter are all straight lines and he faces fixed relative prices, he will never change the commodity (clocks or peanut butter) in which he specializes as his income increases.
- o 2. The corners of indifference contours for complements need not lie on a straight ray through the origin.
- o 3. A consumer with concave indifference contours between heroin and bread will specialize in consuming heroin or bread.
4. If the bread price of heroin increases instead of staying constant as more heroin is consumed (a situation that is, in fact, said to characterize the market for heroin), the corner solution is not inevitable even with concave indifference curves.
5. Corner solutions are possible with convex indifference curves.
6. Portions of indifference contours may well be shaped perversely (rising marginal valuations), but for a consumer with a straight budget line, such portions will never be points of rest.

2.3 The Uses of Indifference Contours

What to Read For

Do economists actually have precise measurements of the shapes of indifference curves? If not, why bother with the idea? Is it possible to answer interesting social questions using indifference curves, even though we don't precisely know their shapes? Is it better to give small towns cheap mail or direct cash subsidies? Can an offer of free, public education sometimes lower the amount of education consumed? How do indifference curves answer such questions?

Like crossword puzzles and mathematical games, the study of the shapes of indifference curves has a certain arid charm. The next chapter will continue the fun and games. But it is as well to show immediately here that indifference curves are not mere toys. On the contrary, they show real behavior.

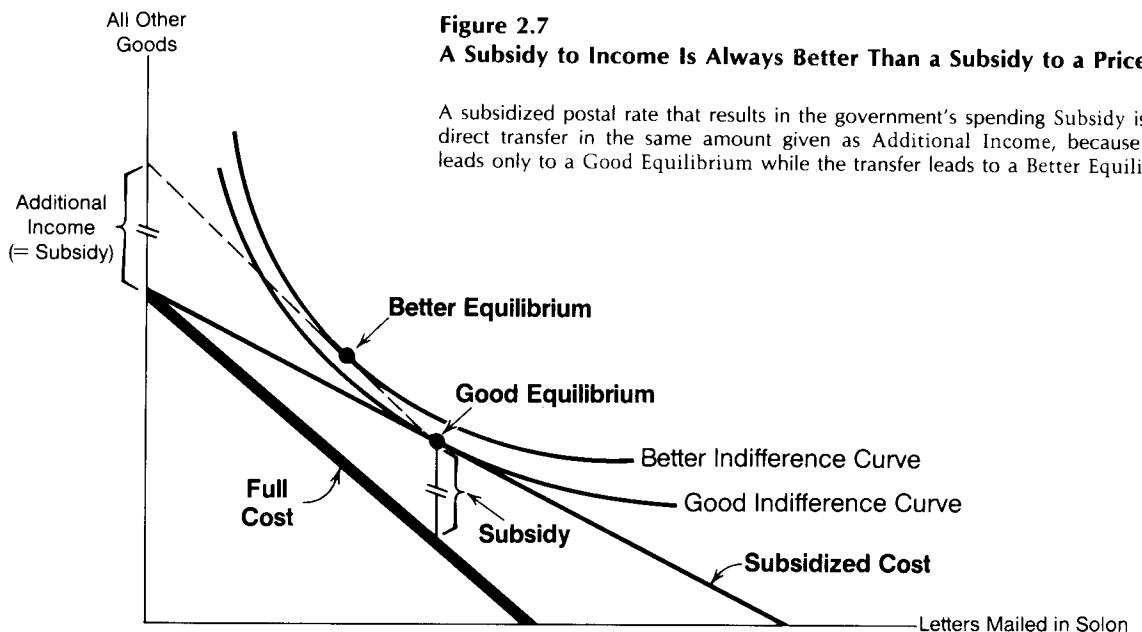
This would be obvious if economists had actually measured indifference curves. If they could look up in a "Handbook of Hills of Utility" the shapes of curves between food and all other goods, as engineers can look up the breaking point of steel under stress, then they would have no trouble predicting exactly how consumers would react to a rise in the price of food or to a fall in income. There is no such handbook. Economics, like most other social sciences and unlike physics and chemistry, cannot very easily fill in the blanks in its theories with laboratory experiments. Like astronomy or meteorology, it must take its experiments as the world gives them. And it has the additional handicap that, unlike astronomy and meteorology, which are applied physics and chemistry, it cannot draw on a rich array of results from the laboratory. Statistical methods in economics—*econometrics*—have since the 1930s filled in some of the blanks. Yet a tendency to apply these methods only to recent experience has limited our understanding of economic behavior. The next chapter offers some further, theoretical, reasons why the handbook of facts does not exist.

Subsidies of Income and of Price

Still, economists can go a long way with a crude notion of the shapes of indifference curves. The crude notion is that they are convex. For example, it is the mere convexity of indifference curves which makes a subsidy in cash more valuable than a subsidy in kind (see Problem 1 at the end of this section, and the discussion on p. 14).

A similar point applies to subsidies that reduce prices. For instance, customers of the U.S. Postal Service in Baroda, Michigan, or Solon, Iowa, get rates lower than the cost of mail there. These are paid for by rates above costs in New York City, or by government subsidies. The heavy budget line in Figure 2.7 is the budget line representing the full cost of sending and receiving letters in Solon, the ordinary line is the budget line after the subsidy. The price of letters falls.

The subsidy is in the amount marked Subsidy if the citizens of Solon buy the mail they want at the low price and achieve the indifference curve Good. They would be even better off, however, if the Subsidy extracted from New Yorkers was simply given to the Solonians directly as Additional Income (in the form of all other goods) without the artificial reduction in the relative



price of mail. The Solonians could move out the dashed line to the point Better Equilibrium, tangent to the Better indifference curve. The dashed line constructed in this way (parallel to the unsubsidized price line since the relative price of mail is now kept unchanged) intersects the Good indifference curve at the Good Equilibrium, which means that there is always a Better Equilibrium. In short, for the same hurt to New Yorkers, Solonians would be made better off by paying the full cost of mails and receiving the subsidy as additional income. It is better to give people additional income than to fool with the prices they face.³

Gifts That Foreclose Other Opportunities

It is not always true that subsidies to a price or gifts of goods increase the amounts consumed. This also can be demonstrated with indifference curves. Consider a good provided free that makes it impossible to buy still more of the good, such as education. If Joseph Reid goes full time to a free public school or college, he cannot simultaneously go full time to an expensive private school or college.

- Q:** 1. Draw the budget line between education and all other goods facing Reid. (*Hint:* It is not a line alone, but a line and a point outside the line. How much of all other goods could he consume if he accepted the free public education?) Assume that public education is a Standardized Amount and that the alternative, private education, can be purchased in varying amounts.
2. Draw an indifference curve that would imply that

Reid would not accept the offer of free public education in the Standardized Amount. In other words, draw an indifference curve such that a point on the ordinary, market budget line is on a higher indifference curve than is the point Standardized Amount.

3. Draw an indifference curve implying that Reid would accept free public education and would consume more education than he would without the offer.

³ But see the Warning at the end of the section. It applies here.

4. Draw an indifference curve implying that Reid would accept free public education but would consume *less* education than he would without the offer.

5. When would you expect an offer of a Standardized Amount of public education to reduce rather than raise the educational attainment of the population?

A: Reid can either accept the offer of free public education, getting the Standardized Amount of education and devoting all his income to all other goods, or he can refuse the offer, giving up some all other goods to get some private education. In other words, his budget line—the collection of attainable points—is the line Private Education or (not *and*) the point Public Education in Figure 2.8. At the point Public Education he can devote all his income to all other goods, since the government has spared him any expense of education.

If the indifference curve tangent to private education resembles U_2 , then the offer of the alternative public education would have no effect, since the point Public Education is on a lower indifference curve. This is the case of education-loving students, for whom the Standardized Amount of public education is too low to be tempting. If the indifference curve resembles U_0 , then

the offer increases the amount of education consumed, since Public Education is now on a higher indifference curve, and in accepting the offer the students consume more (the Standardized Amount is more than E_0). This is the case of all-other-goods-loving students, for whom the Standardized Amount of public education exceeds the amount they would consume privately.

If the indifference curve resembles U_1 , however, the offer *decreases* the amount of education consumed. The point Public Education is on a higher indifference curve, but in accepting the offer students consume less than they would without the offer (the Standardized Amount is less than E_1). This is the case of education-loving-but-not-loving students for whom the Standardized Amount of public education is literally acceptable, but less than they would consume privately. Evidently, if much of the population consumes larger amounts of private education per student than is to be provided free, then the offer can *reduce* the educational attainment of the population. The population is happier with the offer, but happily ignorant. The government may in this way fail if it is attempting to increase the amount of education consumed.⁴

COMMENT

The key to the problem is to set up the diagram correctly. Once the details of the diagram are correct—a budget line between the two goods and a point mutually exclusive to the budget line—the problem solves itself. The diagram is a machine for thinking. You can make mistakes and still get the essential argument correct. For instance, you might not realize that the point Public Education will be exactly horizontal to the maximum attainable of all other goods. But if you recognize that the deal being offered is a choice between the point (wherever located exactly) and the budget line, you have most of it.

Warning: The Society Must Stay on Its Budget Line

A warning is in order. The arguments about how a subsidy of one or another sort affects consumers are fine if one is interested only in how the particular group of consumers is affected and if the group is not a large one in society as a whole. The cross subsidies from New Yorkers to Fairfieldians fit the mold; the subsidy to all public education does not. To provide the Public Education in the diagram shown in Figure 2.8, the economy must tax itself. That is, the

⁴ The logic of the argument is not controversial, but the facts are. Two people who believe the facts warrant taking the argument seriously are Sam Peltzman, "The Effect of Subsidies-in-Kind on Private Expenditures: The Case of Higher Education," *Journal of Political Economy* 81 (January 1973): 1–27, who applies it to college education in the United States; and E. G. West, "Educational Slowdown and Public Intervention in 19th-Century England," *Explorations in Economic History* 12 (January 1975): 61–87, who applies it to primary schooling in England after the introduction of state education in the 1870s.

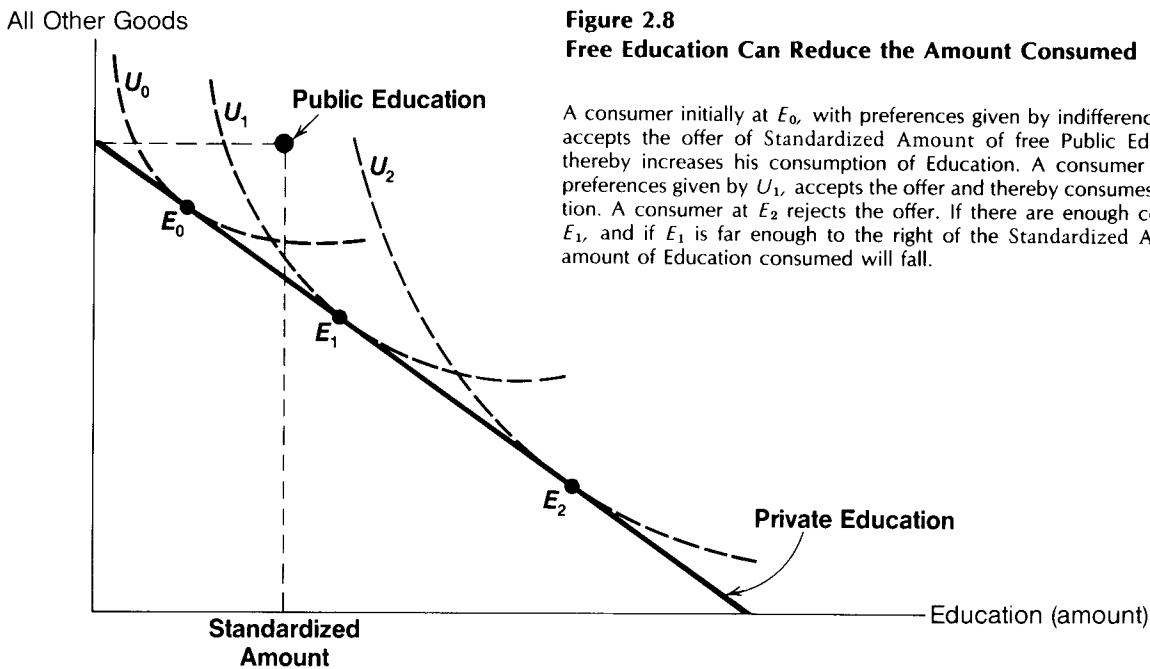


Figure 2.8
Free Education Can Reduce the Amount Consumed

A consumer initially at E_0 , with preferences given by indifference curve U_0 , accepts the offer of Standardized Amount of free Public Education and thereby increases his consumption of Education. A consumer at E_1 , with preferences given by U_1 , accepts the offer and thereby consumes less Education. A consumer at E_2 rejects the offer. If there are enough consumers at E_1 , and if E_1 is far enough to the right of the Standardized Amount, the amount of Education consumed will fall.

society as a whole cannot in reality provide itself with a point such as Public Education outside its original budget line. The budget line typical of the society as a whole must be the budget line in the diagram. Unless a tiny group is being subsidized disproportionately by the rest of society (as is the case, for example, with state-supported higher education), an analysis that simply assumes that the point Public Education is available is incomplete and may be radically misleading.⁵

Summary

Economists have usually only a vague idea of the shape of indifference curves between bread and meat or education and all other goods. Vague though it is, the idea of convexity is enough to cast light on many social questions. The key to answering such questions is close thinking about the budget lines involved. A nonordinary budget line dropped onto ordinary indifference curves may yield definite results. Definite though they are, however, the results must not violate the true budget line for the whole society.

EXERCISES FOR SECTION 2.3

1. The other side of the subsidy to Solon, Iowa, in mail service is the extra cost—virtually the tax—on New York City. Draw the Full Cost line for New York, in the style of Figure 2.8, and draw the line of Taxed Cost.

⁵ The point is made in Milton Friedman, *Price Theory* (Chicago: Aldine, 1976), pp. 65–75, and works cited there. It was first made by Friedman and others in the late 1940s, in response to widespread misuse of the newly popular method of indifference curves.

2. What would be the equivalent for the New Yorkers and their "tax" of the statement in the text: "In short, for the same hurt to New Yorkers, Solonians would be made better off by paying the full cost of mails and receiving the subsidy as additional income"? (*Hint:* Go through the sentence word by word replacing each item with its equivalent.)
3. Suppose that the purchase of private education is *not* foreclosed by accepting in Figure 2.8 the point Public Education. To some degree this is true: You can buy music lessons or Sunday school for your child even if she goes to public school. If it were entirely true, what would be the budget line facing the consumer? (*Hint:* Think it through. What would you do if you were given a gift of free public education that you could supplement with private?)
4. If the United States taxes its own consumption of food, does the national budget line move in? (*Hint:* Examine the Warning a little while back.)

PROBLEMS FOR SECTION 2.3

True or False

1. To keep the poor at the same level of happiness, a scheme of direct money payments to the poor (the so-called negative income tax, for example) would have to give money payments to the poor *below* the money value of the food, housing, medical care, and so forth given to them directly under present welfare programs.
2. If education and all other goods are perfect complements, then an offer of free public education will never reduce the amount consumed.

2.4 Measurable Utility

What to Read For

What is marginal utility? How is it related to total utility? What is the Rule of Rational Life? What is marginal cost? How does marginal utility guide the consumer to a correct utility-maximizing choice? How does one show that the theory of marginal utilities is the same as the theory of indifference curves?

The Relation Between Total and Marginal Utility

There is an old-fashioned way of looking at indifference curves and the consumer's choice that has been declared dead many times but refuses to stay in its coffin. It is called *marginal utility*. Suppose that Richard Zecher's hill of utility lying between the books and corn axes has altitudes measured in joys (a unit of happiness). One could slice the hill at, say, 25 books and look at the profile of the slice. The profile would be Zecher's total happiness or utility in joys achieved from 25 books and varying amounts of corn (the middle panel of Figure 2.9).

The lowest panel is Zecher's marginal utility of one more bushel of corn given various amounts of corn already in hand. In other words, it is a plot of the little increases in his total utility when the amount of corn he has increases a bushel at a time. In still other words, it is the slope of his total utility curve (look at the cases for 6 and 13 bushels). It is, by the way, the first example

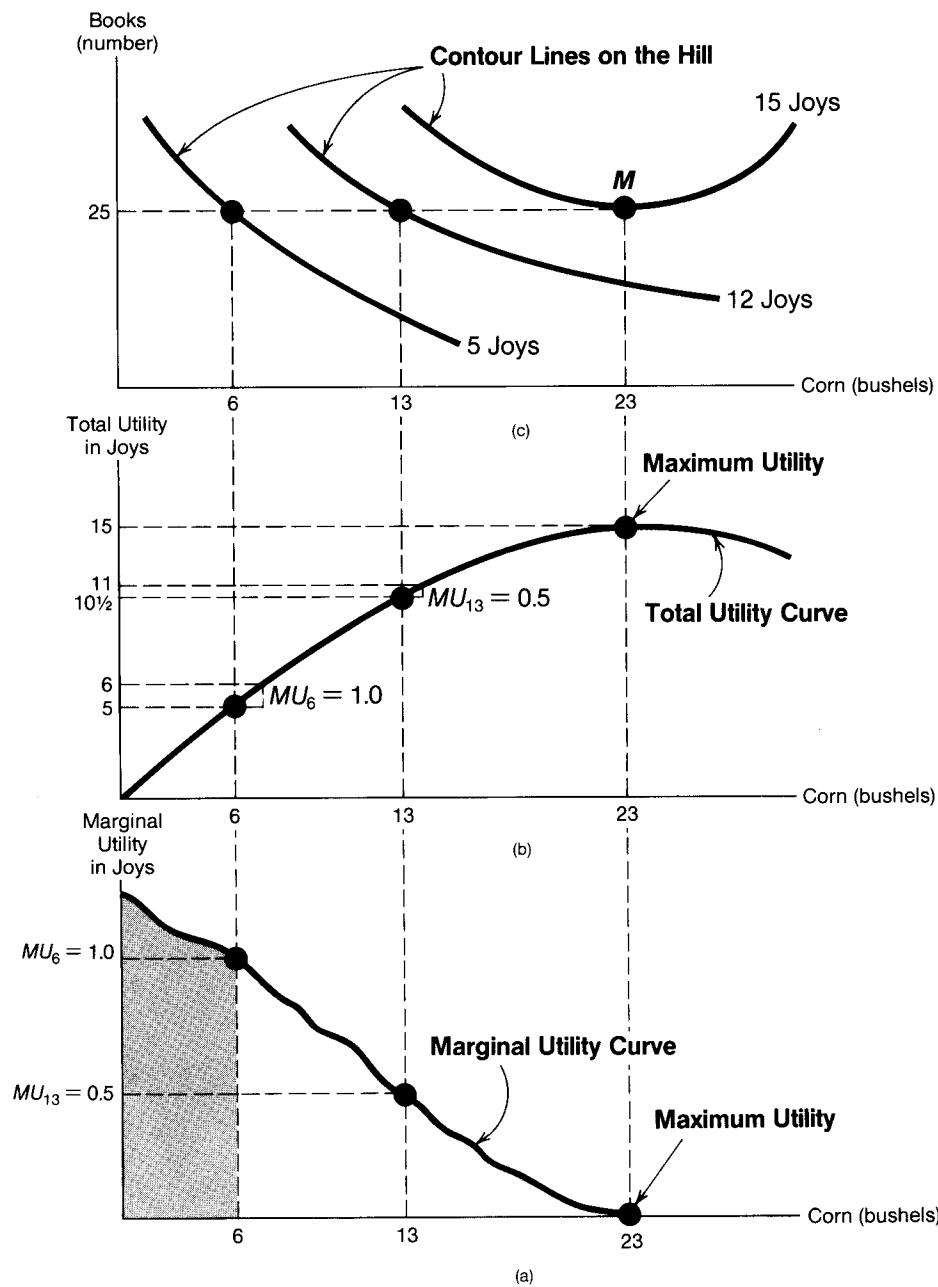


Figure 2.9
The Relations Among
Indifference Curves (a),
Total Utility Curves (b),
and Marginal Utility
Curves (c)

The top panel shows a slice through the hill of utility at a level of 25 books. This slice is then redrawn in the middle panel. The slice is the Total Utility Curve. It rises to the point of Maximum Utility, then falls; beyond Maximum Utility more Corn is a bad, not a good. The bottom panel shows the Marginal Utility Curve. It shows the addition to total utility when the amount of corn increases a bushel at a time. It is also a graphing of the total curve's slope—the rise over the run. At 6 bushels of Corn, for instance, another bushel gives 1 additional Joy, that is, Total Utility rises from 5 Joys to 6 Joys. At 13 bushels another bushel gives an addition to utility of only $\frac{1}{2}$ a Joy. This makes sense: As more Corn is consumed, additional bushels have a lower value.

of dozens in this book of drawing *marginal curves*, that is, curves that are slopes of other curves.⁶

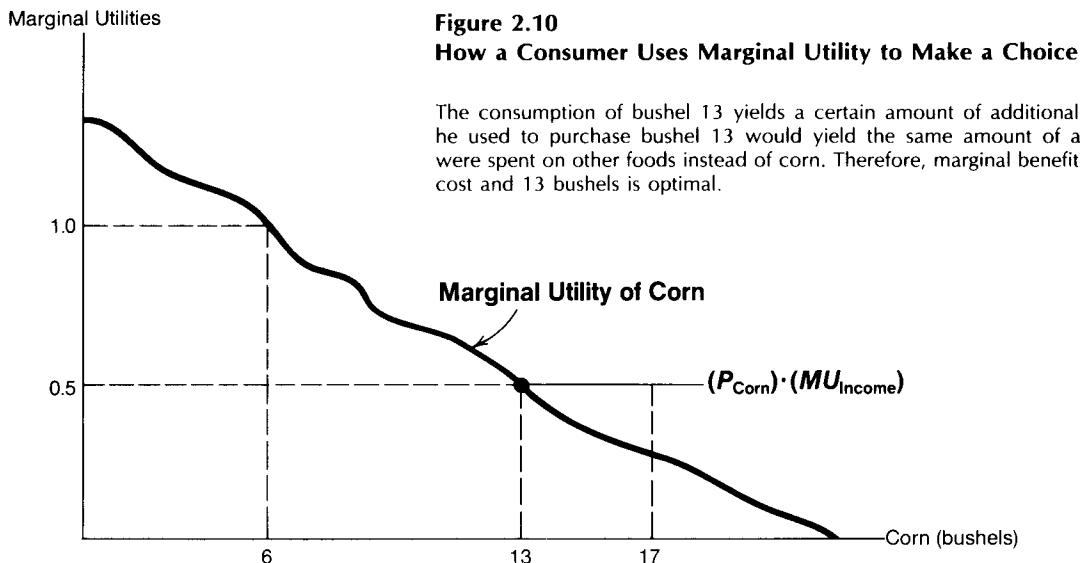
The Rule of Rational Life

The concept of marginal utility gives a description of the consumer's choice that is an alternative to tangencies of indifference curves. In fact, marginal utility was the first description, the leading idea in the "marginalist revolution" in economics during the 1870s. Since that time economists have never stopped talking about "margins." The Rule of Rational Life is: *Pursue an activity until the marginal benefit is equal to the marginal cost of pursuing it further.* In other words, pursue an activity until the additional benefit gained is equal to the additional cost that is incurred by pursuing it one step further. The marginal benefit of talking this way is to this day still above its marginal cost. If the marginal benefit of sleeping or going to college or eating corn is still above its marginal cost, you will get a net benefit from doing more. The benefit on the marginal increment to sleeping—the extra 15 minutes—is higher than the cost. On the other hand, if the marginal benefit is now below the marginal cost, you will get a net benefit from doing *less*. The benefit of cutting back your sleep by 15 minutes will exceed its cost. Finally, when you have arranged matters so that the two are equal, you stop. You have then chosen the optimal amount of sleep—or of college going, corn eating, economics studying, walk taking, advertisement placing, iron pumping, date arranging, steel making, shoe repairing, grass cutting, or whatever you're doing. The Rule shows how to be sensible. It is a definition of being sensible. Economists believe that consumers are on the whole sensible, and apply the Rule to what they do.

Consider Zecher, who with 6 bushels of corn, is deciding whether to buy another bushel. The Rule says that he should continue to pursue his corn consuming a little further as long as the marginal benefit from another bushel is greater than the marginal cost. The marginal benefit from another bushel is the marginal utility of bushel 7, symbolized by $MU_{C=7}$. The marginal cost—that is, what he has to sacrifice to get the bushel—clearly has something to do with the money price of the bushel, P_{CORN} . Dollars and joys, however, are not directly comparable. To measure the marginal cost in units comparable with the marginal utility, one must multiply the dollar price of corn by the utility sacrificed elsewhere (in consuming other things, such as books) per dollar spent on corn. This multiplier can be called the *marginal utility of income*, MU_{INCOME} . It is the amount utility falls when a dollar of income is taken away (by spending it on corn). Figure 2.10 shows Zecher's marginal utility of corn and the price times the marginal utility of income.

The marginal utility of corn declines as Zecher consumes more. Eventually it declines to be equal to the marginal *cost* of corn (which is the utility from other goods Zecher gives up by using his income to buy another bushel of corn). The marginal utility of corn is equal to the marginal cost of corn at 13 bushels. There he stops.

⁶ You might, incidentally, be able to persuade yourself of the true assertion that the shaded area under the marginal utility curve from 0 to 13 bushels is equal to the total utility at 13 bushels: the sum of the increments to utility from the bushel 1, 2, . . . , 13 (which is the area) is clearly the total utility achieved up to bushel 13. If you can do so, congratulations, for you have grasped the fundamental theorem of the calculus. If you cannot, do not despair: The point will come up again in many other contexts until it is clear.



As Tom Nall ate his seventy-fourth tortilla in a sitting at Marciano's Mexican Restaurant in Dallas on October 16, 1973, he undoubtedly reflected that the first tasted better. Had his purpose been pleasure in the food rather than winning the second world championship in tortilla eating, he would have stopped eating sooner. Likewise, Zecher does as well as he can when he arranges his affairs to set the marginal utility of an extra bushel of corn equal to the marginal cost of the bushel. That is, he buys 13 bushels. At 6 bushels he would do well to buy more, because what he gains in utility from more corn is greater than what he loses elsewhere; at 17 bushels, for similar reasons, he would do well to buy less.

The Equations of Marginal Utility

The form of the rule of rational life is in this case $MU_{\text{CORN}} = P_{\text{CORN}} \times MU_{\text{INCOME}}$. Unless it is a very peculiar good, the marginal utility of corn (and of other things, such as tortillas) diminishes after some point as more is purchased and consumed, enabling Zecher to achieve equilibrium at 13 bushels. In other words, the marginal utility of corn is a function of (depends on) the amount of corn consumed and is in fact after some point a diminishing function.

The marginal utility of corn is also in general a function of the amount of other goods consumed, such as books: A choice of 10 rather than 25 books in the first panel of Figure 2.10 would have resulted in a different slice of total utility and a different curve of marginal utility. Therefore, the theory is not complete unless it also tells how the number of *books* is determined. The number is determined as corn is, by the marginal utility equation; that is, $MU_{\text{BOOKS}} = P_{\text{BOOKS}} \times MU_{\text{INCOME}}$. Nor is the theory complete unless it keeps Zecher on his budget line—he cannot spend more than his money income, Y , and will not spend less, that is, $Y = P_C C + P_B B$. The marginal utility theory of the consumer's choice is therefore three equations:

1. $MU_C(C, B) = P_C \times MU_Y$ Best choice of C
2. $MU_B(C, B) = P_B \times MU_Y$ Best choice of B
3. $Y = P_C C + P_B B$ All income spent on C and B

The C and B in parentheses after the MU 's say that the MU 's are functions of C and B . These are to be solved for three unknowns, the amounts of corn and books purchased and the marginal utility of income. All three will usually vary as income, Y , and prices, P_C and P_B , vary. Any particular choice of income and price determines the three.

Marginal Utility Theory Is the Same as Indifference Curve Theory

At first glance all this appears to be radically different from the earlier theory of consumers' choices, which has them seeking tangencies between their budget lines and one of their indifference contours. At second glance, however, the two theories are the same. Look at the three equations just given. Eliminating MU_Y by substitution leads to one equation in place of the first two:

$$\frac{MU_C}{P_C} = \frac{MU_B}{P_B}$$

It is illuminating to express this equation in a different notation, by writing out MU_C as $\Delta U / \Delta C$ —the change in utility per unit of change in corn consumed:

$$\frac{\Delta U}{P_C(\Delta C)} = \frac{\Delta U}{P_B(\Delta B)}$$

In other words, to be satisfied with his choice of corn and books, the consumer must get equal increments to utility per additional dollar spent on each [since $P_C(\Delta C)$, for example, is the number of dollars spent on ΔC]. Reread that last sentence. You will agree that it is a persuasive way of describing the behavior of an optimizing consumer. Now begin thinking again of budget lines and tangencies. Think of Zecher as moving along his budget line, reducing his books in exchange for increasing his corn. From this perspective the equation just given says that, when the consumer has achieved equilibrium, the utility lost from fewer books is exactly balanced by the utility gained from more corn. That is, total utility is now constant. That is, the two ΔU 's are equal. Therefore, one can rearrange the equation just given and cancel the ΔU 's:

$$-\frac{\Delta B}{\Delta C} \Big|_{U \text{ CONSTANT}} = \frac{P_C}{P_B}$$

The vertical bar signifies a side condition and the negative sign appears because more corn implies (along the budget line) fewer books.

The equation now says that Zecher sets his marginal valuation of corn equal to the slope of the budget line ($-P_C/P_B$). As was just demonstrated, the marginal valuation is related to the slope of an indifference curve, *an* indifference curve because utility is held constant. But arranging matters so that these slopes are equal *is* the tangency condition for an optimum in the more usual theory. Adding the condition that consumers spend all their income will complete the theory. It is at bottom the same as the marginal utility theory.

Once the sameness of the two theories is recognized, the marginal utility variant, with its cumbersome expression, loses much of its appeal. It is more direct to say "Zecher chooses the highest indifference curve he can reach" than to say "Zecher equalizes the marginal utilities per dollar of expenditure on each good, subject to his budget constraint," much less to say "Zecher, subject

to his budget constraint, sets the marginal utility from a good equal to the product of the good's price and the marginal utility of income." Life is too short for pointless mental gymnastics.

T or F: If Max Hartwell consumes only bread and water and if the marginal utilities of each depend only on the ratios in which the two are consumed, then neither is inferior to him, technically speaking.

A: True. Along any ray through the origin in Figure 2.11, the marginal utilities of bread and water are constant, because along a given ray the ratio in which the two are consumed is constant. Since the marginal utilities along the ray are constant, so too are their ratios. But the ratio of marginal utilities is the slope of an indifference curve, as was just shown.

COMMENT Don't worry if your eyes glaze over when you study such a problem. They should: It is a dull problem, aimed at a detail of the logic, not of an important feature of economic life. Never mistake this kind of exercise for real economics. On the other hand, you need to know how to exercise before you can play the game.

Summary The theory of marginal utility is the oldest theory of the role of tastes in consumption. It amounts to an application of the rule of rational life: Pursue an activity (such as consuming corn, producing steel, seeking public office, waging war, or learning economics) until the marginal benefit of pursuing it a little further has fallen to the point of being equal to its marginal cost. To do less or more is to forgo some happiness. Attractive as this formulation is, it is clumsier in expression than the theory of indifference curves, and the same in content.

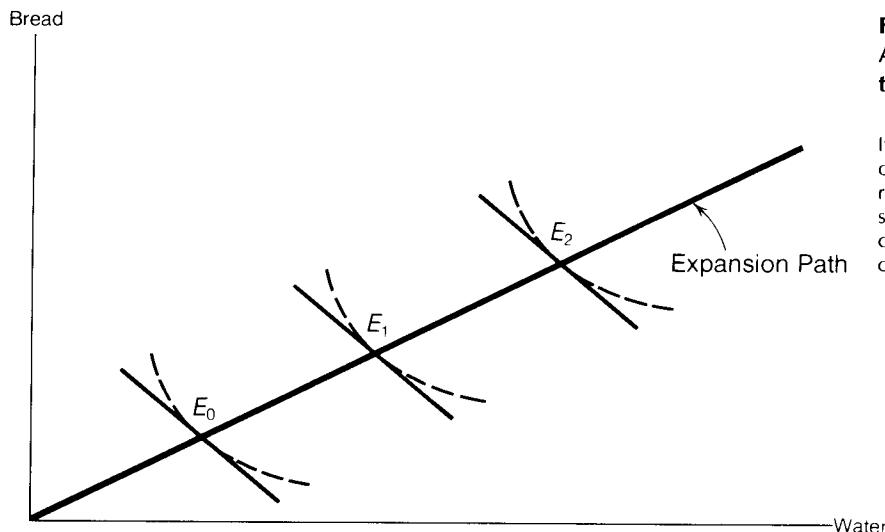


Figure 2.11
A Condition for the Expansion Path to Be a Ray

If marginal utilities, and hence the slopes of indifference curves, depend only on the ratio of the two goods consumed, the consumer does not alter the ratio of goods consumed when his or her incomes increases.

EXERCISES FOR SECTION 2.4

1. Identify the marginal benefit, marginal cost, and the optimal point in these activities:
 - a. Information gathering.
 - b. Lowest price searching.
 - c. Lawn weeding.
 - d. Line-for-Star-Wars-Returns joining.
 - e. Examination question creating.
2. *True or false:* The Rule of Rational Life is to pursue an activity until its *total* benefit is worth its *total* cost.

PROBLEMS FOR SECTION 2.4**True or False**

1. The marginal utility of garbage is negative.
2. If the marginal utility of heroin increases as more is consumed, an addict would set the marginal utility of a shot of heroin equal to its price multiplied by the marginal utility of income.
3. In taking an exam, Atack, a rational student, allocates his time to the various questions so as to equalize his marginal point utility per minute on all questions.
4. The marginal utility of food to Zecher depends only on the amount of food (and not on the amount of housing) and the marginal utility declines as more food is consumed; likewise for housing. Therefore, both food and housing are normal goods. (*Hint:* Express the first two equations of the marginal utility theory in the alternative form $MU_F/MU_H = P_F/P_H$. Notice that P_F/P_H is fixed. If all of an increase in income is spent on F , can the equality be maintained?)
5. Margaret Comi, a consumer of housing and food with a so-called *Cobb-Douglas utility function*, that is, $U = H^\alpha F^\beta$, will increase her amounts of housing and food (prices held constant) in proportion to increases in her income, Y , and will spend the same shares of her income on housing and food regardless of their prices or her income.

CHAPTER

3

The Measurement of Utility and the Economics of Risk

3.1 Paradise Lost: Nonmeasurable Utility

What to Read For **Can one really measure utility, or marginal utility? What, then, does an indifference curve say? Do indifference curves need to be ordinal or cardinal to describe sensible consumer behavior? Is it possible to draw indifference curves by simply observing the behavior of consumers? How?**

The Measurement of Utility Is Difficult

The talk of “marginal utility” and “contour lines on a hill of utility” began to make economists uncomfortable almost as soon as it started. How is one to measure utility? To put it another way, how could one devise an experiment to determine whether or not Coatsworth gets more utility from the third hot dog than from the fourth? Such questions are answerable for height or temperature. For example, having agreed that a certain platinum-iridium bar at the International Bureau of Weights and Measures near Paris is to be called “one meter,” Coatsworth’s height in meters is measurable; having agreed that the height of a column of mercury at the freezing point of water under certain specified conditions is to be called “32° Fahrenheit” and the boiling point “212°,” his body temperature is measurable. The question is not so easily answered for his utility of hot dogs. The volume of his squeals of delight as successive hot dogs are presented to him is not much of an answer. Neither is his own testimony, although it might be worth listening to (“Ah yes, the third hot dog gives me 10 joys, the fourth only 8 joys”). The theory appears to depend on quantities that are hard or even impossible to measure.

The Measurement of Utility Is Pointless

For nearly fifty years some of the best minds in economics, from Pareto in his *Manuale di economia politica* (1906) to Houthakker in his “Revealed Preference and the Utility Function” (*Economica*, 1950), labored to free economists from

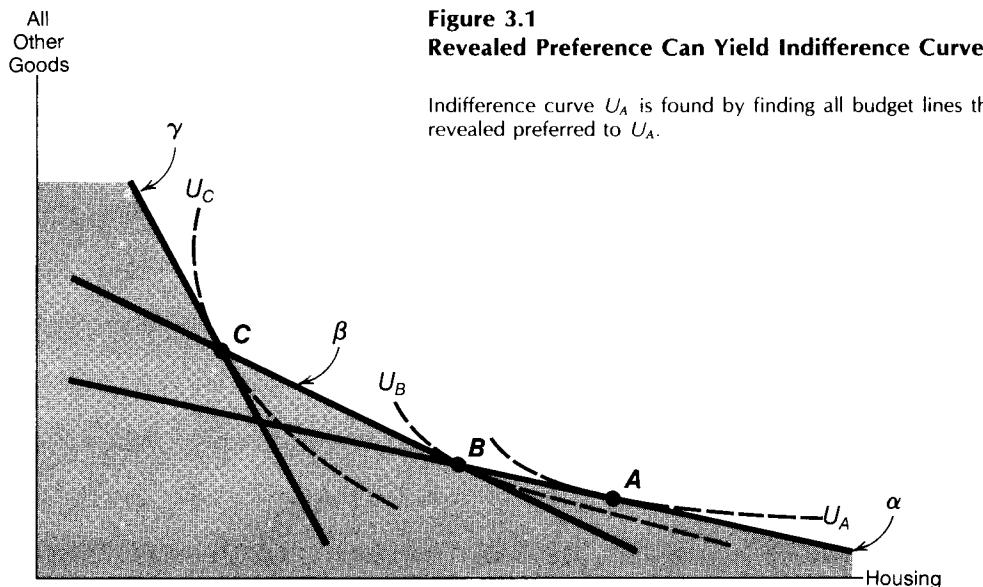
the embarrassment of speaking quantitatively about things they could not quantify. The labor was successful, although it is unclear that the result was worth the opportunity cost of alternative employment for these minds.

The result is easily explained. As long as two hills of utility between housing and food have the same contours dropped onto the housing-food plane, each will give the same equilibrium point for given budget lines. In other words, it does not matter whether three successively higher contour lines are labeled, on the one hand, 10 joys, 20 joys, and 30 joys or, on the other, 65.3 joys, 65.4 joys, and 6,598,135 joys. If the consumer wants to move from the second to the third utility contour, that is all that matters. It does not matter whether the move transports the consumer to ecstasy or merely to a small bit more of happiness. The consumer will in either case make the move. Indifference curves need only be *ordinal* (a certain curve is the first, another the second, and so on), not *cardinal* (a certain curve yields 10 joys, another 20 joys, and so on). The *order* of the curves is all you need to describe the behavior resulting from the consumer's tastes.

The Theory of Consumption Can Get Along Without Measurable Utility

The indifference curve then, speaks of combinations of food and housing among which a consumer is unable to choose, not of the absolute amount of happiness the consumer gets from the combinations. One could map the indifference curves of a consumer by questionnaire, asking the consumer to rank many different bundles of, say, housing and all other goods. Economists, however, are more suspicious of questionnaires than are other social scientists. They are so suspicious that they seldom call on even imaginary questionnaires, preferring to infer preferences from imaginary experiments in consumer behavior. In the most neutral language, indeed, indifference curves are called *behavior lines*.

The experiment from which one could infer behavior lines is an exercise in revealed preference, exhibited in Figure 3.1. Recall that one combination of



housing and all other goods presented to Cynthia Morris is “revealed preferred” to another if the second combination is inside the budget line when she buys the first. She could have bought the second combination rather than the first, but did not. Present Morris with budget line α and watch her choose point A on it. Now find a budget line β with a slope greater than α for which the consumer will choose a point such as B , also on the α budget line. Now find a budget line γ with a slope greater than β for which the consumer will choose a point such as C , also on the β budget line. By revealed preference, A is the best of the three points (see the dashed indifference curves); indeed, it is superior to the entire shaded area below the budget lines for which the three points are in fact chosen.

By parallel reasoning one can construct an upper shaded area of points *better* than A . And by putting B very close to A , C close to B , D close to C , and so forth, one can reduce the unshaded area to a line, at least locally. The line will separate points superior to A from those inferior to A . It will be an indifference curve (the U_A curve). The experiment (which, it must be pointed out, has never been performed on anyone but college students in the grip of professors of economics or psychology) yields the curves without mentioning the dreaded word “utility.”

As the inventor of indifference curves put it in 1881, “We cannot *count* the golden sands of life; we cannot *number* the ‘innumerable smile’ of seas of love; but we seem to be capable of observing that there is here a *greater*, there a *less*, multitude of pleasure-units, mass of happiness; and that is enough.”¹

Summary From the 1870s, on, economists hoped that experimental psychology, with its new techniques for measuring stimuli, would provide them with a way to measure happiness. They gradually lost hope, but made the best of a bad situation by noticing that none of their uses of contour lines on the hill of utility depended on the measurability of their altitudes. The problem remained that the shape of the contour lines themselves appeared to depend on asking consumers to rank bundles. Asking consumers whether they are happier with one bundle than with another seemed to economists in the grip of a naive view of science to be unscientific. In the 1930s and 1940s, however, even this element of subjectivity was erased from the theory, by demonstrating that the observed behavior of consumers (not their testimony) sufficed to draw up their indifference curves.

APPENDIX TO SECTION 3.1: The Calculus Solution of the Problem

The mathematics of the ordinality of indifference curves depends only on the chain (or the function of a function) rule of second-year calculus. Take Morris’s utility function to be any function, $U(H, F)$. The condition for equilibrium, signifying partial derivatives with subscripts, is $U_H/U_F = P_H/P_F$. Now suppose that $U(H, F)$ is transformed by inserting it into a function G , such as $G(U) = U^2$ for positive

¹ Francis Ysidro Edgeworth (1845–1926), *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London, 1881), pp. 8–9, his italics. Edgeworth, fellow of All Souls College, first editor of the *Economic Journal*, the product of a marriage of an Anglo-Irish landowner to a Spanish refugee, invented much of the diagrammatic machinery of economics. To quote him as an ordinalist (although that is what the sentence appears to mean) gives an inaccurate picture of his views. Like most of the leaders of the marginalist revolution, he believed utility to be measurable.

U , that always rises and has no flat places (that is, it is monotonically increasing). The hill is stretched to higher altitudes, and the higher a place on the hill is initially the more it is stretched. The “new” equilibrium for the new utility function will be, applying the chain rule, $(G_U \times U_H)/(G_U \times U_F) = P_H/P_F$. Clearly, if G_U is never zero (that is, if G has no flat places), G_U cancels, and the same condition as before holds. The equilibrium is not new; it is the same. Clearly, therefore, any assignment of utilities to various contours that keeps their order (as does a monotonically increasing reassignment) will do as well as any other. That is, HF , $(HF)^2$, $\log HF$, $\alpha + \beta(HF)$, and so forth all have the same indifference contours in the H, F plane.

EXERCISES FOR SECTION 3.1

1. Which of the following could be the values of Coatsworth’s marginal utilities of the third and fourth hot dogs?
 - a. Third = 50 joys, fourth = 25 joys.
 - b. Third = 50 joys, fourth = -500 joys.
 - c. Third = 6000 esbits, fourth = 5999 esbits.
 - d. Third = 10 joys, fourth = 11 joys.
2. Write a typical question in the questionnaire that would map Morris’s indifference curve.

3.2 Paradise Regained: Measurable Utility

What to Read For	What is average utility? How is average utility related to gambling? In terms of average utility, when will a smart gambler take a gamble? How can you use average utility to calculate “the” utility of various bundles? Why is “the” utility not unique? Is temperature unique?
-------------------------	--

The Utility of Different Prospects	In 1944 the goddess of knowledge played a great joke on economists. At the very moment they were growing confident that their theories of the consumer’s choice did not require measurable marginal utility, it pleased this goddess to give them a method of measuring it. ² The method uses gambling. Just as subjecting Alex Field to various budget lines induces him to reveal the order in which he ranks bundles of goods, subjecting him to various gambles can induce him to reveal the utility levels he attaches to bundles of goods.
---	---

Consider Figure 3.2. Suppose that Field buys a 7-room house at Cairo, Illinois, on a low-lying bank of the Mississippi. He consumes each year in addition to the 7 rooms of housing some other goods, say, 200 pounds of hamburger. The gamble he faces each year is that the Mississippi will flood and his house will be destroyed. If he is unlucky he will consume, say, 2 rooms of housing (a

² Her agents were John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton, N.J.: Princeton University Press, 1944), Chapter 3. The book is one of the classics of economics. There is no new thing under the sun, and von Neumann and Morgenstern were anticipated in this by two mathematicians, Frank Ramsey in 1931 and, for a special case, Daniel Bernoulli in 1738.

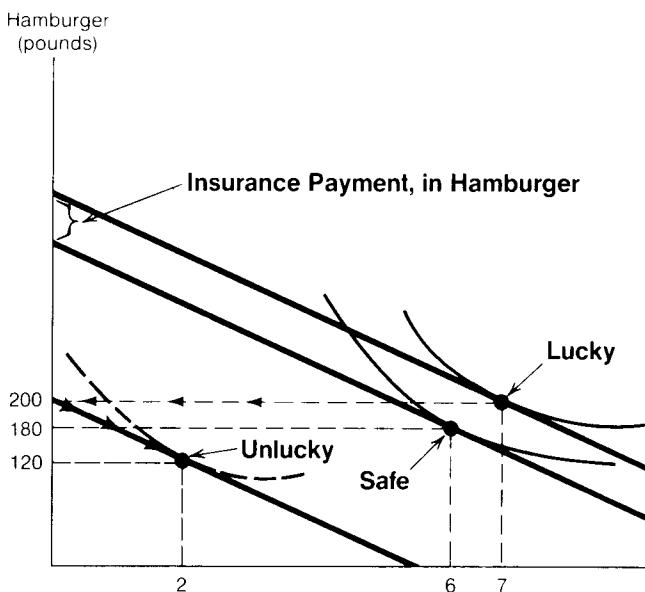
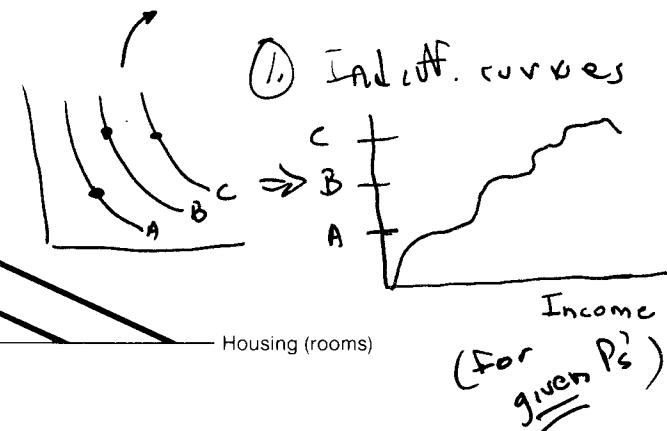


Figure 3.2
The Choice Between Insurance and a Gamble

The consumer can buy insurance that guarantees the budget line Safe or can take the gamble with the river, which will leave him on either budget line Lucky or budget line Unlucky.



mobile home) and 120 pounds of hamburger: The flood reduces his income, and he therefore consumes less. He has so far two bundles, Lucky and Unlucky. Suppose that he also has available a third bundle, namely, buying insurance on the house each year that will make him a little poorer but Safe. The insurance, unless it is provided in the form of flood relief by the government, is not free to Field, so paying it will shrink his annual budget line below the Lucky budget line of no flood. Suppose that in this Safe situation he consumes 180 pounds of hamburger and 6 rooms (cutting his hamburger consumption and renting out room 7 to pay the insurance company). His three possible budget lines and the three bundles they lead to are shown in Figure 3.2.

Notice that Field follows the arrows if he does not buy the insurance and the flood comes. The house is destroyed, but he does not sit mournfully in the rubble with 0 rooms: He uses some of his hamburger (all other goods) to buy 2 rooms. Notice, too, that the Safe budget line lies between the Lucky and Unlucky budget lines.

Introducing Probabilities of Prospects

So far there is nothing novel in the analysis. The indifference curves in the diagram are merely ordinal, merely rankings that might be revealed by subjecting Field to an experiment in shifting budget lines. But suppose that he is told (and believes) that there is a 1 in 20 chance each year that the Mississippi will flood and a 19 in 20 chance that it will not. If he does not buy the insurance, then it is reasonable to say that he has revealed a preference for this gamble over the Safe bundle. Use a vocabulary of cardinal, measurable utility, with three utility numbers (in joys) attached to the three possible bundles: U_{LUCKY} joys, $U_{UNLUCKY}$ joys, U_{SAFE} joys. He could be said to have revealed that the utility he gets *on average* from the gamble, that is, $\frac{1}{20}(U_{LUCKY} \text{ joys}) + \frac{19}{20}(U_{UNLUCKY} \text{ joys})$, is for him a greater number than is U_{SAFE} joys. This is the crucial step in the

argument. Field is assumed to behave like a smart professional gambler, who takes a gamble only when he thinks that what he will end up with on average is greater than what he can have for sure by simply not taking the gamble. The difference is that Field gambles for joys instead of dollars.³

The Utility of the Prospects Can Be Derived from the Experiment

Now suppose that the risk of the flood is raised above 1 year in 20. Were the risk very high, say, 1 year in 2, Field would buy the insurance, revealing that he preferred the Safe bundle to the gamble. Suppose that 1 flood every 10 years is the level of risk that makes him *just indifferent* between buying the insurance and taking the gamble. Therefore, the average utility in joys from the gamble must be *equal* to the utility in joys from the Safe bundle: $\frac{9}{10}(U_{LUCKY} \text{ joys}) + \frac{1}{10}(U_{UNLUCKY} \text{ joys}) = U_{SAFE} \text{ joys}$.

If one knew the number of joys from the Lucky and Unlucky bundles, one could put a number on the Safe bundle. Go ahead. Be brave. Try some numbers. Set U_{LUCKY} at, say, 1000 joys and $U_{UNLUCKY}$ at, say, 100 joys. The U_{SAFE} is then $\frac{9}{10}(1000) + \frac{1}{10}(100) = 910$ joys. This number is "the" utility of the safe bundle in much the same sense as "the" temperature today in Iowa City is 50° Fahrenheit. A scale of temperature requires two arbitrary numbers to set it (32°F freezes water, 212°F boils it), because two numbers are required to fix how much a single degree is. Likewise, the scale of utility requires two arbitrary numbers to set it ($U_{LUCKY} = 1000$ joys, $U_{UNLUCKY} = 100$ joys). Once the scale of joys is set, any bundle Field consumes can be placed on the scale by subjecting him to choices between gambles linked to the initial points.

The Army Corps of Engineers builds a levee that eliminates gambling with floods. A local sport offers him a 1 in 5 chance to win 50 additional pounds of hamburger (which Field would use to move out to point Win in Figure 3.3). If he loses, he gives up the hamburger formerly used to pay the insurance (in which case he would end at point Safe). Were he just indifferent between risking the insurance money this way and staying at a third bundle, Lucky, Win would be worth the solution of

$$\frac{1}{5} U_{WIN} + \frac{4}{5} U_{SAFE} = U_{LUCKY}$$

or

$$\frac{1}{5} U_{WIN} + \frac{4}{5} (910) = 1000$$

or

$$U_{WIN} = 1360 \text{ joys}$$

(8)

³ The assumption that Field—or anyone—gambles for joys as a wise gambler gambles for money is still controversial, more than 30 years after it was first proposed. The essence of one objection is that people may love or hate gambling itself. If Field hates gambling he may refuse the gamble (that is, buy the insurance) even though the average utility of taking the gamble is higher than the utility of the Safe bundle. The essence of another objection is that people do not calculate as closely as the theory demands, that they are ignorant of the facts, or that they are governed by habit. Such objections could be made (and have been, frequently) of all economic theory, and do not appear to have special force here.

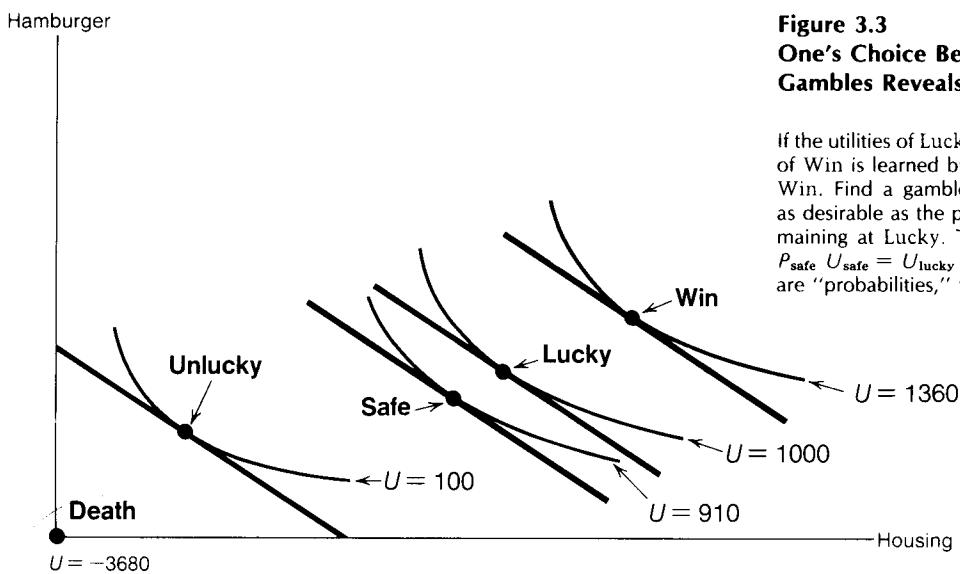


Figure 3.3
One's Choice Between Safe Bundles and
Gambles Reveals One's Utility

If the utilities of Lucky and Safe are known, the utility of Win is learned by finding a gamble that includes Win. Find a gamble between Win or Safe that is as desirable as the prospect of not gambling and remaining at Lucky. Then the equation $P_{\text{win}} U_{\text{win}} + P_{\text{safe}} U_{\text{safe}} = U_{\text{lucky}}$ can be solved for U_{win} (the P 's are "probabilities," the U 's utility numbers).

as in the diagram. You have just succeeded in attaching a utility number to the bundle Win. And by the meaning of an indifference curve, all the bundles on the indifference curve running through Win would have this same utility, 1360 joys.

Again, suppose the Corps' levee broke, leaving Field at Unlucky, a desperate man. He might contemplate some other dangerous activity, such as attempting to jump a motorcycle across the Mississippi for money. Suppose that success in this attempt would give him the bundle Win, that failure would give him the bundle Death, and that he was just indifferent between making the attempt on the one hand and staying at Unlucky on the other when the chance of success was 1 in 4. Then death would be worth the solution of

$$\frac{1}{4} U_{\text{DEATH}} + \frac{3}{4} U_{\text{WIN}} = U_{\text{UNLUCKY}}$$

or

$$\frac{1}{4} U_{\text{DEATH}} + \frac{3}{4} (1360) = 100$$

or

$$U_{\text{DEATH}} = -3680 \text{ joys}$$

as in the diagram. In like fashion one could attach a utility number to every indifference curve. If you know the probabilities Field attaches to various events, then, his utility is in one sense measurable.

Summary An experiment in alternative budget lines gives indifference curves a rank; an experiment in alternative gambles gives them numbers. The numbers measure utility in nearly the same sense as temperature is measurable.

EXERCISES FOR SECTION 3.2

1. Why is the Safe budget line between the Lucky and Unlucky budget lines in Figure 3.2?
2. If a flood risk of 1 in 5 were the risk that made Field just indifferent between a certainty of Safe or a gamble between Lucky and Unlucky, what would be the utility of Safe in joys? (Take $U_{UNLUCKY} = 100$ and $U_{LUCKY} = 1000$ as in the text.)
3. Suppose that Field is sitting at Lucky and is offered a ticket to the Illinois State Lottery costing the hamburger that would bring him down to Safe (that is, it costs the same as flood insurance). Suppose that the chance of winning is 1 in 1000 and Field is at these odds just indifferent between buying and not buying the ticket. What is the utility of Winning the Lottery?

PROBLEMS FOR SECTION 3.2

1. Paul Hohenberg contemplates buying 100 shares of Nachtschwärmer Home Improvement Company at \$1 a share. He believes that there is a 1 in 4 chance that he will gain \$400 and a 3 in 4 chance that he will lose the \$100 entirely. He finds himself unable to decide between buying the shares and doing nothing. *True or false:* His marginal utility of income around his present income is declining. (*Hint:* "Unable to decide" is "indifferent.")
2. Suppose that you knew somehow that Hohenberg gets 35 joys from the \$100 lower income, 36.25 joys from his present income, and 40 joys from the \$400 higher income. Suppose further that Hohenberg loves playing the stock market, to the extent of getting 2 joys from the thrill of buying any Nachtschwärmer stock and participating in its ups and downs.
 - a. If Hohenberg believes that there is a little over a 1 in 20 chance of the Nachtschwärmer investment earning him the \$400 gain (and a little below a 19 in 20 chance of it earning him the \$100 loss), what choice will he make between his present income and the gamble?
 - b. If you knew as before that the \$100 lower income was worth 35 joys to Hohenberg and the \$400 higher income worth 40 joys but did *not* know that his present income was worth 36.25 joys or that the joys he got from gambling was 2 joys, what number would you put on Hohenberg's utility from his present income?
 - c. What do you conclude about the necessity for the method of measuring utility of assuming no love or hate of gambling?
3. Marilyn Coopersmith, an executive in a publishing firm, will gain \$100,000 for the firm if an economics book she publishes is successful; she will lose \$25,000 for the firm if the book is unsuccessful. *True or false:* If Coopersmith will not publish any book that she believes has less than a 1 in 5 chance of success but will publish one with more than a 1 in 5 chance (that is, in such cases she preserves the firm's present income, Y , instead of risking some of it in the book), then the marginal utility of income that she is assuming on behalf of the firm is constant.

True or False

4. If rooms of housing sell for 50 pounds of hamburger per room, Field's marginal utility of income declines from Unlucky and Safe compared with Safe and Lucky.

3.3 Imperfections in Paradise?

What to Read For **Would units other than joys be suitable for measuring utility? Is a scale of utility of joys for Field unique?**

Utility Is Not Measurable Absolutely

There are, however, several different senses of measurability, and there are three senses in which the method of gambling does *not* measure utility. The first is trivial. Because the unit of marginal or total utility depends on the scale chosen and no scale is uniquely suitable, the unit in which marginal or total utility is measured has no absolute significance. The utility of Lucky could be set at 1 joy or at 10,000 joys, multiplying all the other utilities by 0.001 or 10, without affecting any conclusion drawn.

The second is somewhat less trivial, but not much less. Although one *marginal* utility for Field can be said to be three times another, one *total* utility cannot be said to be three times another. Each degree of temperature in the Celsius scale, in which the freezing and boiling points of water are 0° and 100° instead of 32° and 212°, corresponds to 1.8° in the Fahrenheit scale. Neither scale is "correct." The choice of what numbers to attach to the freezing and boiling points of water is arbitrary, just as choosing to call the utility of Field's Lucky bundle 1000 joys and that of his Unlucky bundle 100 joys is arbitrary. Another choice results in another scale. A temperature of 212°F is $212/32 = 6.63$ "times as hot" as 32°F only in the Fahrenheit scale; in the Celsius scale it is $100/0$ (that is, undefined nonsense) times as hot.⁴ It is therefore meaningless to say that the winter of 1985 was "twice as cold as normal."

On the other hand, it *is* meaningful to say that the increment of temperature (the marginal temperature) from 60°F to 212°F is 5.43 times the increment from 32°F to 60°F, because no matter what scale is used, the statement is true. Thus 60°F is 15.56°C, and $(212 - 60)/(60 - 32) = 152/28 = 5.43$ gives the same result as $(100 - 15.56)/(15.56 - 0) = 5.43$. In similar fashion, as you can show by replicating the calculations for a scale of $U_{LUCKY} = 2000$ and $U_{UNLUCKY} = 1000$, it is meaningless to say that Field gets $910/100 = 9.1$ times more utility from Safe as from Unlucky. But it is meaningful to say that his additional (that is, marginal) utility derived by moving from Unlucky to Safe is $(910 - 100)/(1000 - 910) = 9.0$ times the additional utility derived by moving from Safe to Lucky. As long as marginal utility is measurable on some scale, it is no great loss to be unable to measure total utility, since it plays no role in the economist's model of the consumer.

⁴ Once absolute zero is found (-460°F or -273°C), as it was by physicists in the nineteenth century, such statements take on meaning. Starting each scale from its absolute zero, the boiling point of water on the Fahrenheit scale is $(212 + 460)/(32 + 460) = 1.366$ times its freezing point and the same on the Celsius scale: $(100 + 273)/(0 + 273) = 1.366$. But there is no obvious way to measure absolute zero on a utility scale. Death will not usually suffice, because most people can imagine fates worse than mere death: their own deaths compounded by the deaths of their families, for example, or terrible and worthless lives. For the undignified egomaniac that economists sometimes assume typifies mankind, however, death may be zero.

Utility Is Not Comparable Between People

The third and final sense in which the gambling theory leaves utility unmeasurable is more serious. It is that utilities, whether total or marginal, cannot be compared between Field and, say, Butlin. To assert that Field (a poor man) gets more joys from an additional dollar of income than does Butlin (a rich man) is, regrettably, meaningless. What is regrettable about it is that, were such an assertion possible, we could redistribute income from Butlin to Field with confidence that the world's total happiness was being increased. What is meaningless about it is that no argument could convince someone who disbelieved it that Butlin's dollar contained more happiness than did Field's.⁵

Q: The incomes of Field (who is poor) and Butlin (who is rich) are to be taxed. Suppose that their marginal utilities of income fall as their incomes rise. *True or false:* If it is thought to be just that both make equal marginal sacrifices of utility in paying income tax, then it can be demonstrated scientifically that the tax should be progressive; that is, the rich Butlin should be charged a higher percentage tax on a dollar of extra income than the poor Field.

A: False, because science cannot compare the utilities felt by the two from an extra dollar. By contrast, this argument for progressive taxation assumes that Butlin and Field have the same curves of marginal utility (see

Figure 3.4). Field's utility from an extra dollar is the tall column in the diagram, Butlin's is the shorter one. Suppose the tax on the dollar were proportional instead of progressive, for example, 50% for both men (see the two points marked 50%). The \$0.50 taken from Field would be more valuable than the \$0.50 taken from Butlin (a poor man's sacrifice of utility would be larger than a rich man's). To bring the marginal sacrifices into equality, Field's additional dollar must be taxed at a lower rate (20% in Figure 3.4) than Butlin's (50%). That is, income taxes should be progressive. But the elaboration of the argument is silly, because we do not know whether or not Butlin has the same curve of marginal utility as Field.

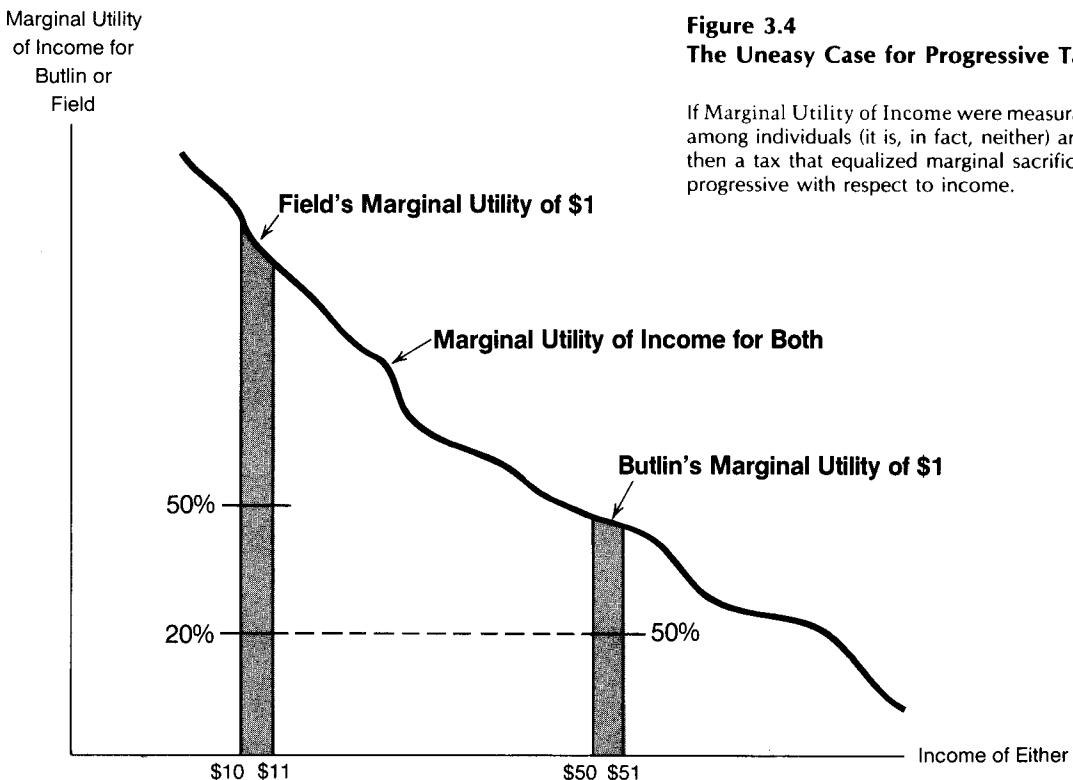
As the economist Henry C. Simons remarked in 1938, at the time the weaknesses in the argument were becoming clear,

One derives practical implications from the criterion of equality, or proportionality, of sacrifice precisely in proportion to one's knowledge of something which no one has ever known, or ever will know, anything about. Perhaps this goes far toward explaining the popularity of these doctrines among academic writers.⁶

The case for progressive taxation must rest directly on a moral premise that more equality of incomes is desirable, not indirectly on a pseudo-scientific comparison of happinesses. Such are the limits of measurable utility.

⁵ "No argument" is perhaps a little strong, for a philosopher, John Rawls, in *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), and a number of economists have in fact conceived one. They have extended in a most ingenious way the notion of revealing preferences by gambling to social as well as private choices. The gamble they propose is to imagine yourself before you knew who you were going to be (the daughter of a millionaire in Texas or the son of a porter in Hong Kong) choosing how you would wish society to be organized. The lucky event in this lottery would be to become the millionaire's daughter, the unlucky event to become the porter's son. Rawls believes that you would choose to make the society you were about to enter more equal in income, to reduce the risk in the lottery. If this were true, it would solve implicitly the problem of comparing utilities. But it is uncertain whether it is true. See, for example, Robert Nozick, *Anarchy, State and Utopia* (New York: Basic Books, 1974), especially Chapter 7, Section II.

⁶ *Personal Income Taxation* (Chicago: University of Chicago Press, 1938), portion reprinted in R. W. Houghton, ed., *Public Finance: Selected Readings* (London: Penguin, 1970), p. 21. For an excellent, nontechnical discussion of the issue, see Walter J. Blum and Harry Kalven, Jr., *The Uneasy Case for Progressive Taxation* (Chicago: University of Chicago Press, 1953), and subsequent editions.



Summary The analogy between measuring utility and measuring temperature is quite exact. Neither the units nor the proportions between levels of utility or temperature are unchangeable. What is unchangeable is the proportion between differences. Marginal utility is not measurable uniquely any more than is the increase in temperature (the "marginal temperature") between a cold and a hot day. An increase of 60°F can be expressed as an increase of 33.3°C. But the rate of decline of marginal utility is measurable uniquely, just as a rise in temperature of 60°F is half of the rise of 120°F (and in Celsius 33.3°C and 66.6°C). The analogy breaks down in only one important feature: You can make interpersonal comparisons of rates of rise of marginal temperature, but you cannot make interpersonal comparisons of rates of decline of marginal utility.

EXERCISES FOR SECTION 3.3

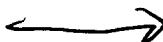
1. What is the utility of Safe for Field according to the first (1/10) gamble if the utility of Lucky is taken to be 2000 joys and of Unlucky 1000 joys? (Look back at the preceding section.)
2. Is it true, as the text says, that regardless of which scale is chosen for Field's utility the marginal utility of moving from Unlucky to Safe is nine times the marginal utility of moving from Safe to Lucky? (Make sure that you've answered Exercise 1 first.)

3. Show that if the curve of marginal utility of income were the same for Butlin and Field it would be a good idea to steal money from Butlin and give it to Field.

PROBLEMS FOR SECTION 3.3

True or False

1. If Field's utility of Lucky is set at 100 joys and of Unlucky at 10 joys (instead of 1000 and 100), then his utility of Death is -368 joys.
2. If Field's utility of Lucky is set at 1000 joys (as before) but his utility of Unlucky at 500 joys (instead of 100), then his utility of Win is 1200 joys (instead of 1360).
3. Accepting interpersonal comparisons of identical curves of the marginal utility of income, if the sum over Mrs. Rich and Mrs. Poor of utility lost from income taxation is to be as small as possible, then only Rich should be taxed (unless the government wants so much tax revenue that Rich's income is driven down to Poor's). (*Hint:* Notice that "utility lost from taxation is to be as small as possible" is not the same as the example worked in the text, "to bring the marginal sacrifices into equality.")



3.4 Living in Paradise: The Uses of Measurable Utility

What to Read For	What is the <i>utility of income</i> ? What role does it play in thinking about gambling? What is <i>risk aversion</i> ? How is it related to the shape of the utility of income curve? What is <i>fair insurance</i> ? Why does a criminal take a chance on being caught? Why shouldn't you put all your eggs in one basket?
-------------------------	---

The Utility of Income

What has been fruitful about the idea of measurable utility is not the measurements themselves (few have actually been done), but the tools for understanding consumer behavior that have been invented along the way. The premier tool is the *utility of income* curve. This curve shows the total utility that the consumer gets from different amounts of income.

Recall that the marginal utility of food is the slope of a clean slice through the hill of utility at a given quantity of all other goods. In a similar way, the *marginal utility of income* is the slope of an irregular slice through the hill of utility along the consumer's best (utility-maximizing) path up the hill (for given prices). In the top panel of Figure 3.5, the slice is the dashed expansion path. Each point on this expansion path is the consumer's best (utility-maximizing) bundle, given a certain level of income. The slope up the hill of the expansion path is the marginal utility of income.

The bottom panel of Figure 3.5 plots the total utility of income against income (measured in terms of all other goods). This is the utility of income curve. It is the way in which the expansion path slice would appear if you looked at it while standing on the All Other Goods axis.⁷

⁷ Notice that the curve will shift if relative prices shift. For the analysis to be simple, the gambles contemplated must be gambles that do not themselves shift relative prices.

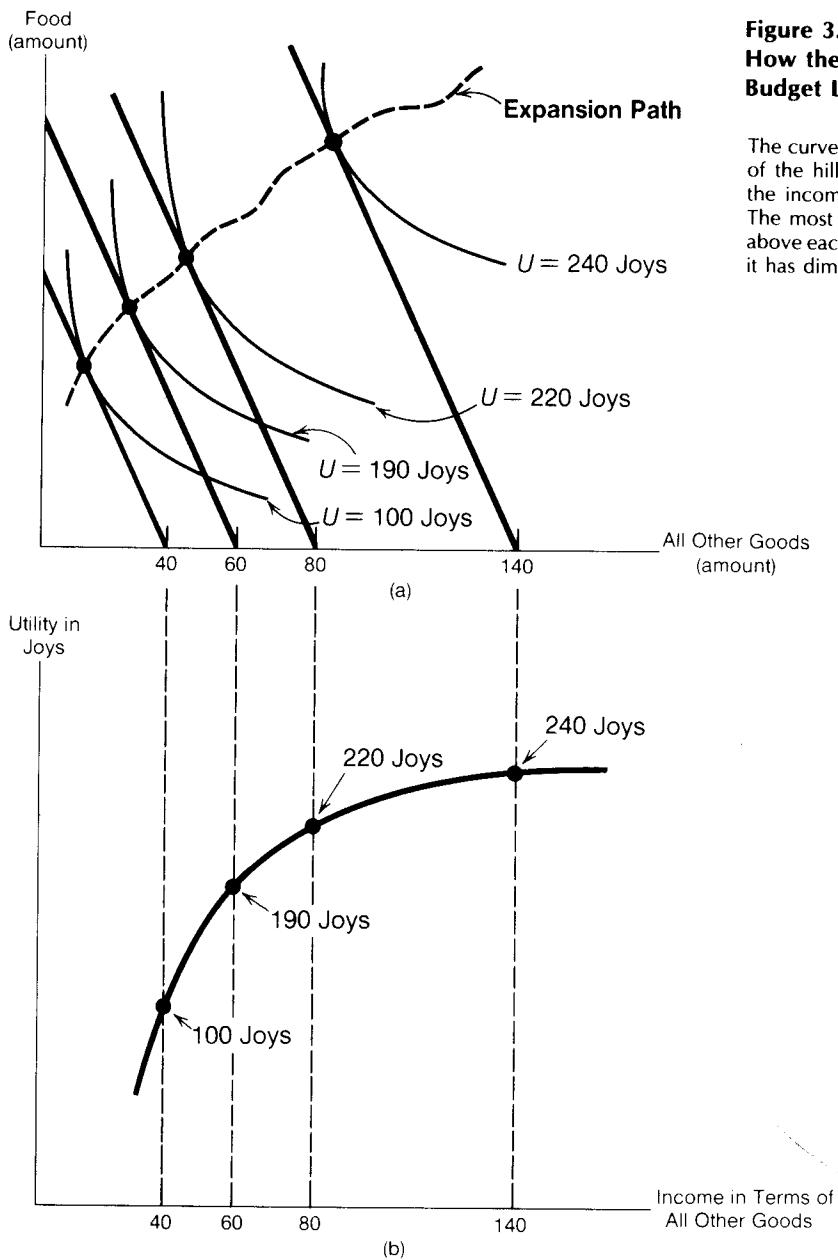


Figure 3.5
How the Utility of Income Is Connected with
Budget Lines (a) and Utility Contours (b)

The curve of the utility of income is a plot of the heights of the hill of utility for various levels of income. Here the income is measured in terms of All Other Goods. The most Joys attainable with each income are plotted above each. Notice the curvature of the utility of income: it has diminishing *marginal* utility.

Whatever one might think of its status in a world of certainty, the shape of the curve has meaning in a world of uncertainty. It does measure something, for utility is measurable, at least for a single person. And it neatly summarizes the person's condition in various circumstances, allowing the analysis to collapse all commodities into one (income) with an associated utility. The idea is useful.

Increasing Marginal Utility Implies Accepting Unfair Gambles

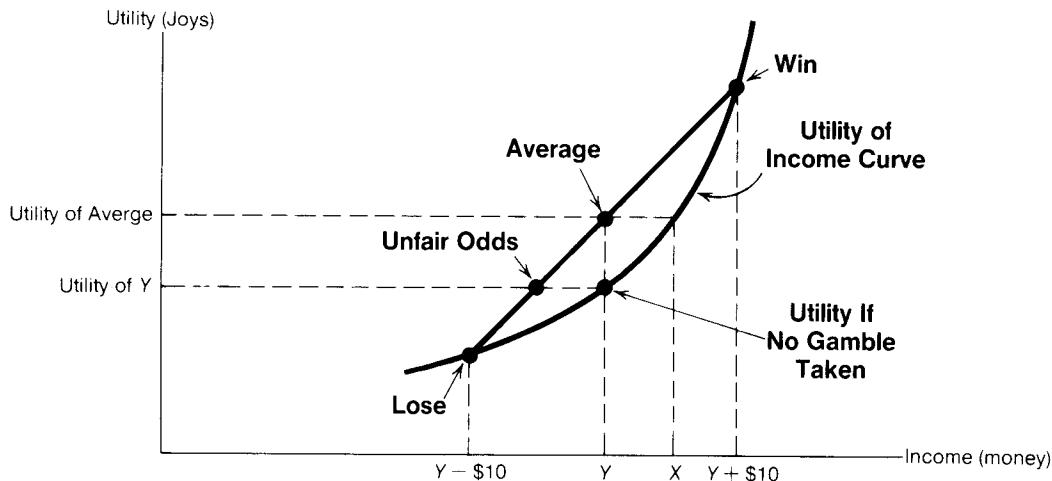
Consider, for example, the situation of Linda Zecher, a gambler playing in a casino in Las Vegas.⁸ Zecher knows that the games are unfair, in the technical sense that on average she will lose. The state of Nevada and the casinos take a certain percentage of the amounts wagered and give back to the gamblers only what is left (from 95% in blackjack, more if the gambler has a good memory, down to 70 or 60% in slot machines). Yet Zecher plays. Why? An obvious and true answer is that she likes the thrill of playing. Another is that she miscalculates the odds or believes that Lady Luck is smiling on her. But there is another possibility that would be the only possibility for a cold-blooded person with a sure grasp of the odds who nonetheless gambles.

Q: Zecher faces a 50:50 chance of gaining \$10 or losing \$10. *True or false:* If her marginal utility of income is *increasing*, she will prefer the gamble to keeping her present income.

A: The gamble is a fair one. That is, on average she neither gains nor loses income by taking it. In Figure 3.6 Zecher's utility of income is upward curving, the gamble is the line between Lose and Win, and with 50:50

Figure 3.6
Increasing Marginal Utility of Income Implies a "Preference for Risk" and Accepting Unfair Gambles

A risk-loving person may find the prospect of increasing her income by \$10 sufficiently attractive, relative to the prospect of losing \$10, to accept a gamble that will leave her, on average, with a lower income than if the gamble had not been taken. The Utility of Average is higher than the Utility If No Gamble Taken.



⁸ The first, decisive step in translating the von Neumann-Morgenstern theory into a tool for understanding behavior was taken by the economist Milton Friedman and the statistician L. J. Savage in their paper, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy* 56 (August 1948): 279-304. The following draws heavily on this paper.

odds Zecher expects to arrive on average at the point Average halfway along the line (that is, at the income she started with, measured along the horizontal axis).

Zecher will accept the gamble because the average utility she derives from it (namely, the utility of Average, measured along the vertical axis) is greater than the utility derived from holding on to her present, riskless income, Y . Another way of putting it is to say that, because Zecher's marginal utility of income is rising (the slope of total utility is rising), the \$10 increase

in income is valued more than is the \$10 decrease, and if the two are equally probable she expects to gain utility on balance by taking the gamble. In fact, she would accept worse odds than 50:50 and still take the gamble. Odds that placed her at Unfair odds in the diagram (unfair because they lead to an income on average below her present income) give Zecher's utility a little above the utility of her present, riskless income. She would therefore accept them. She plays the tables at Las Vegas although the odds are unfair. In short, true.

So a person with an upward-curving utility of income (that is, a rising marginal utility of income) prefers fair (and even some unfair) gambles to a secure income. Equivalently, were the present, riskless income at X in Figure 3.6 instead of Y , Zecher would be willing to pay $X - Y$ dollars to get into a casino that offered an uneven chance of ending up at a new, lower income, Y . This explains why casinos are able to charge big admission fees to poker games held on their premises. A person who will pay to gamble is said to exhibit a *preference for risk*.⁹

Risk Aversion and Unfair Insurance

Similarly, a person who will pay *not* to gamble is said to exhibit an *aversion to risk*, and the marginal utility of income is declining. Paying not to gamble is insuring.

Q: Joe Reid pays \$400 a year for automobile insurance that covers him against suits for damages if he kills somebody with his car. He reckons (correctly) that during the year there is a 1 in a 1000 chance that he will kill somebody and be sued for \$100,000. *True or false:* Reid is irrational to buy the insurance.

A: The insurance he buys is, to be sure, unfair, again in the technical sense that he pays more for the insurance (\$400) than the income he would expect to lose *on average* from taking the gamble: $100,000 (1/1000) + 0 (99/1000) = \100 . But if his marginal utility of

income is declining, as in Figure 3.7, he is in fact willing to pay as much as \$800 to avoid having to take the gamble (to make the argument clear, Figure 3.7 is purposefully not drawn to the correct scale).

His position without an accident is the point Win, but he faces the risk of Lose, and on average will end up at Average. The insurance company offers to give him a riskless income of $Y - \$400$. Because his marginal utility of income is declining, this amount of riskless income yields him more utility than does Average. He therefore buys the \$400 insurance, quite rationally. So, false.

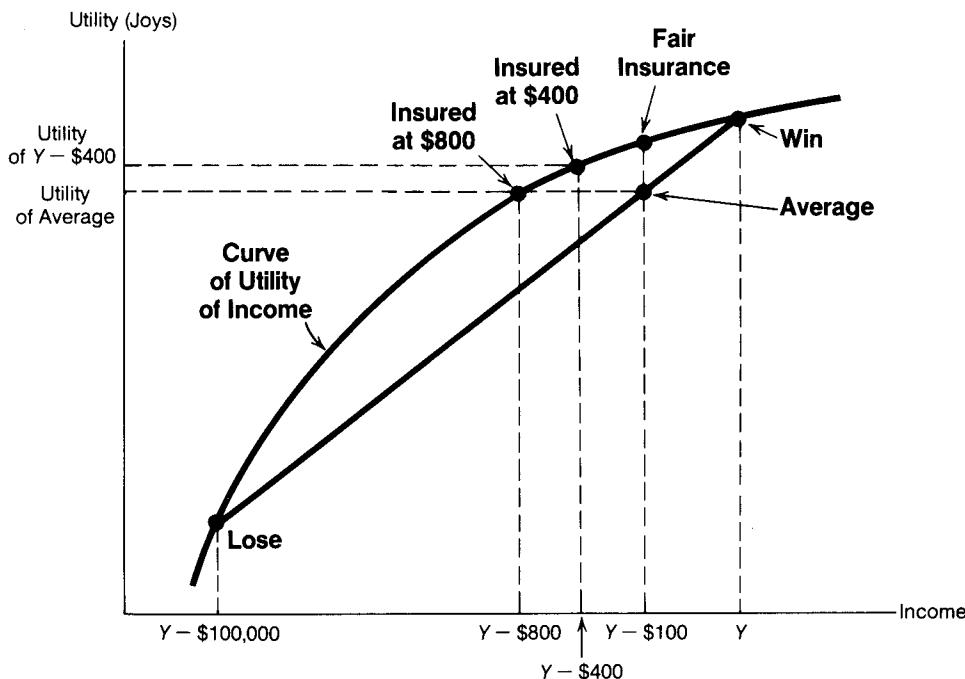
COMMENT

Notice that the insurance company, insuring thousands of Joe Reids, will pay \$100 a year in claims to the average Joe while collecting \$400 a year from each. Its profits (which are practically riskless if the number of Joes is large) will be \$300 times the number of Joes. But if the insurance companies compete with each other by tempting each others' clients away with lower

⁹ A preference for risk in this sense is not the same as a love of gambling. The distinction is between liking the results of gambling and liking the act of gambling. If you find the distinction elusive, you are not alone—many economists agree with you. The question is whether or not acts of gambling themselves enter the utility function. As was shown in Problem 2 at the end of Section 3.2, the gambling technique for measuring utility breaks down if gambling itself is a good or a bad. One can imagine an experiment to test whether or not acts of gambling enter the utility function. If our gambler in Las Vegas would be willing to pay to get into the fair game more than she could possibly earn by winning, that is, \$10, then it is clear that she likes the act of gambling, not merely its results.

Figure 3.7
Decreasing Marginal Utility of Income Implies an Aversion to Risk and Accepting Unfair Insurance

Someone with diminishing marginal utility of income finds the prospective loss of \$100,000 so unattractive that he buys insurance against the loss, even at premiums so high that he pays the insurance company more than it will on average pay him.



insurance premiums, the companies will drive the premium down to \$100 plus the administrative costs of insuring the clients. That is, the insurance will become fair, the Joes will be even better off than they were accepting the \$400 offer (they will be at the point Fair Insurance rather than at Insured at \$400), and insurance companies will earn the normal profit of enterprise. If insurance were not fair on average, it would suggest that insurance companies were to some degree protected from the full rigors of competition—by collusion to stop price cutting, for example, or, more innocently but with the same result, by the ignorance of their customers that cheaper prices exist.

Crime and Punishment

The applications of the utility-of-income curve are not merely to gambling in casinos or insuring against rare disasters. All life is a gamble, and insurance is everywhere. Consider, for example, a life of crime.¹⁰

¹⁰What follows draws on an article by Gary S. Becker, "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76 (March–April 1968): 169–217, reprinted in William Breit and Harold M. Hochman, *Readings in Microeconomics*, 2nd ed. (New York: Holt, Rinehart and Winston, 1971), pp. 339–369.

Burglary is a good case, for it is a passionless, planned, and nonviolent crime that few would claim is beyond the reach of economic analysis. Suppose that burglars were risk averse. This means that their marginal utility of income is declining and that they will only take gambles that raise their income on average (and not even all of these). It does not mean they will take no gambles. The gamble they take per burglary in the United States is about a 1 in 10 chance of being caught, convicted, and sent to prison for about two years on the one hand and a 9 in 10 chance of getting away with whatever they can steal and sell on the other. Since it is sad but true that there are people who will accept such odds, apparently the situation of the typical burglar is as shown in Figure 3.8.

The burglar's income in prison is Y_{LOSE} , the money equivalent of such misery; his income if he is not caught is Y_{WIN} . His average utility from the gamble exceeds the utility of the income, Y_{HONEST} , he could earn in legal employment. That is why he is a burglar. His average income (distinct from his average utility) is also greater in burgling. To put it another way, in this case crime pays for the same reason that being a human cannonball or a coal miner or a test pilot pays. Burglary is risky relative to, for example, working at a filling station, and a risk-averse and dishonest person will move from working at the filling station during the day to breaking into it during the night only if the odds of success are better than fair.

Each of the standard proposals for reducing burglary fits the diagram. The object is to place the point Average below the point Utility of Honest Income, making crime not pay in a utility sense. One approach is to improve the economic lot of potential criminals, by reducing unemployment, for example, or by improving the welfare system. This amounts to raising Y_{HONEST} enough that the utility from it exceeds the utility of Average. Another approach is to reduce the level of Win, by not leaving \$50,000 worth of diamonds on the dresser table or by making it more difficult to fence stolen television sets. Another approach is to

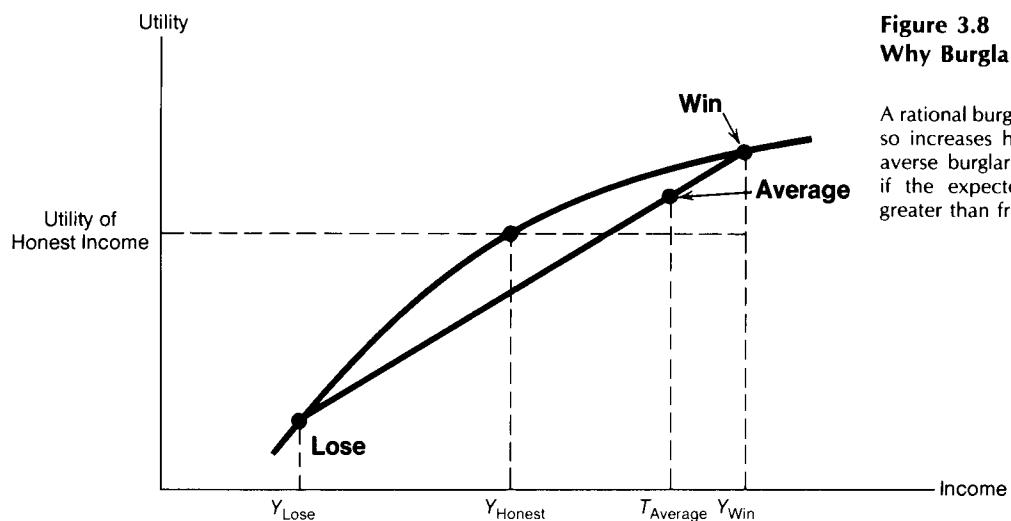


Figure 3.8
Why Burglars Burgle

A rational burglar commits a crime if doing so increases his expected Utility. A risk-averse burglar will commit a crime only if the expected income from crime is greater than from Honest Income.

move Average down the (straight) gambling line, increasing the risk of getting caught by spending more on discrete burglar alarms and indiscrete police officers. Another is to make his "income" in jail, Y_{Lost} , lower, by giving longer sentences (spending more on prisons) or making prison more unpleasant (spending less on each prison). And still another is to change his utility function, making him more risk averse (in which case he will demand more favorable odds before embarking on a life of crime) or making him feel that crime is evil (in which case he will dislike the act of taking a criminal gamble and will not in this sphere conform to the model of gambling with utilities), by giving him religion or psychiatry.

Other Risk-Averse Behavior: Diversification

Risk aversion rather than risk preference is normally assumed to characterize human behavior, for the good reason that it explains much human behavior, the behavior, for example, of economists:

T or F: A person with a doctoral degree in economics employed by General Motors earns 40% more than one employed by the University of Michigan. Part of the difference can be explained by noting the greater security of academic employment and by supposing that the academicians are relatively risk averse.

A: True. Were economists risk preferrers, it would be false; that is, they would pay to gamble on employment

by General Motors. But they are in fact risk averters and pay (in a lower salary) to avoid a gamble. There are, of course, compensations besides security that contribute to the charm and therefore to the low salaries of academic life: abuse in student riots, grading examinations, serving on committees with verbose colleagues, and so forth.

The assumption that aversion toward risk is the normal attitude leads to another, related diagram equally useful for thinking about risky choices. If Peter Hill, a cattle rancher in Montana, is risk averse, he must be compensated in higher average income for taking a risk. He feels that risk is a bad, and only by giving him more of a good (higher average income) will he accept more of it. Figure 3.9 portrays as solid lines the contours of Hill's hill of utility between average income and variability around that average.¹¹ If Hill were risk *preferring*, the contours would slope *downward* instead of upward; if he were risk neutral (with a straight line for a utility-of-income curve), the contours would be straight, horizontal lines running out of the average income axis.

The usefulness of the diagram derives from the following reasoning. Suppose that, if Hill puts all his wealth into savings accounts at the First National Bank of Bozeman, he achieves point Safe on the average income axis (with zero variability); and if he puts it all into cattle, he achieves point Risk, with a higher average income but also a variable income. At Safe, he earns, say, \$10,000 a year every year for sure; at Risk, he earns on average, say, \$30,000, but in some years nothing at all and in other years, say, \$60,000.

T or F: Hill chooses point Best; that is, he holds part of his wealth in savings accounts and part in cattle.

A: True. By varying the amounts he puts in the two assets, he can move anywhere along the Safe-Risk line.¹²

¹¹ You should be able to convince at least your intuition that there is a connection between these contours and the utility-of-income curve. If the left side of your brain, too, wants satisfaction, it will first have to learn some statistics, not a bad thing to do.

¹² If Safe had any variability, the line would in general no longer be straight but curved. And if the income from Safe moved inversely with the income from Risk, the curve would have the opposite

If he puts half his wealth in each asset, for example, he achieves an average income of $\frac{1}{2}(\$10,000) + \frac{1}{2}(\$30,000) = \$20,000$ and a variability of income on the upside of $\frac{1}{2}(\$60,000) + \frac{1}{2}(\$10,000) = \$35,000$ and on the downside of $\frac{1}{2}(\$10,000) + \frac{1}{2}(0) = \5000 . Both

the average and the variability are halfway between Safe and Risk. That is, he is on a straight line between them, his budget line. The best point on or below the budget line is Best.

This is the moral of the tale. Hill does not in general put all his wealth into one investment but, rather, holds a mixture of safe bank accounts and risky cattle, the better to achieve the best combination of risk and return. With a different attitude toward risk, he would hold a different bundle. Similarly, Diane Lindstrom might own a Ph.D., a bank account, and a piece of swamp in Florida, with various amounts of her wealth tied up in each asset. Similarly, General Electric owns on behalf of its stockholders a diverse portfolio of land, patents, and factories. The diagram, which can be elaborated in various ways, exhibits the behavior of risk-averse people, namely, people who do not put all their eggs into one basket.

Summary

Spain searched for Eastern spices and found Mexican gold. In like fashion, the search for measurable utility—an apparently pointless academic expedition—found something better. The utility-of-income curve and related curves can describe behavior toward risk: gambling, insurance, and other choices among risky activities. Life is a lottery, and the logic of lotteries is applicable to the economics of gambling, insurance, crime, capital punishment, investment, knowledge, technical change, war, unemployment, occupational choice, agriculture, and enterprise—that is, to much of the theory of price.

EXERCISES FOR SECTION 3.4

1. *True or false:* If the indifference curves in the top panel of Figure 3.5 were labeled 100, 120, 220, and 1000 joys the resulting utility-of-income curve would show *increasing* marginal utility.
2. Describe how the following are insurance, that is, how they let one have a stable income at the cost of some income (or convenience) on average.
 - a. Tenure of professors.
 - b. Membership in a community that helps you out when the floods come.
 - c. Holding a lot of different assets rather than one kind.
 - d. Carrying an umbrella when it hasn't started raining yet.
 - e. Frisking of airline passengers to stop hijacking.

PROBLEMS FOR SECTION 3.4

- 1. Consider Richard Sylla deciding how much of a \$100 fund he is going to invest in two assets, money balances (amounts in the cookie jar or in his checking account, which are perfectly secure but yield no interest rate) and bonds (which yield an interest rate but are not perfectly secure). Sylla is risk averse.
 - a. Describe for a given average interest rate on bonds how he chooses the proportion

curvature as the indifference curves. The reason is simple. When he is holding two assets whose incomes tend to move inversely, their variations will on average offset each other. The 50:50 mix, for example, will give the same average income as will the straight line, but because of the offsetting, the same average income will be associated with lower variability.

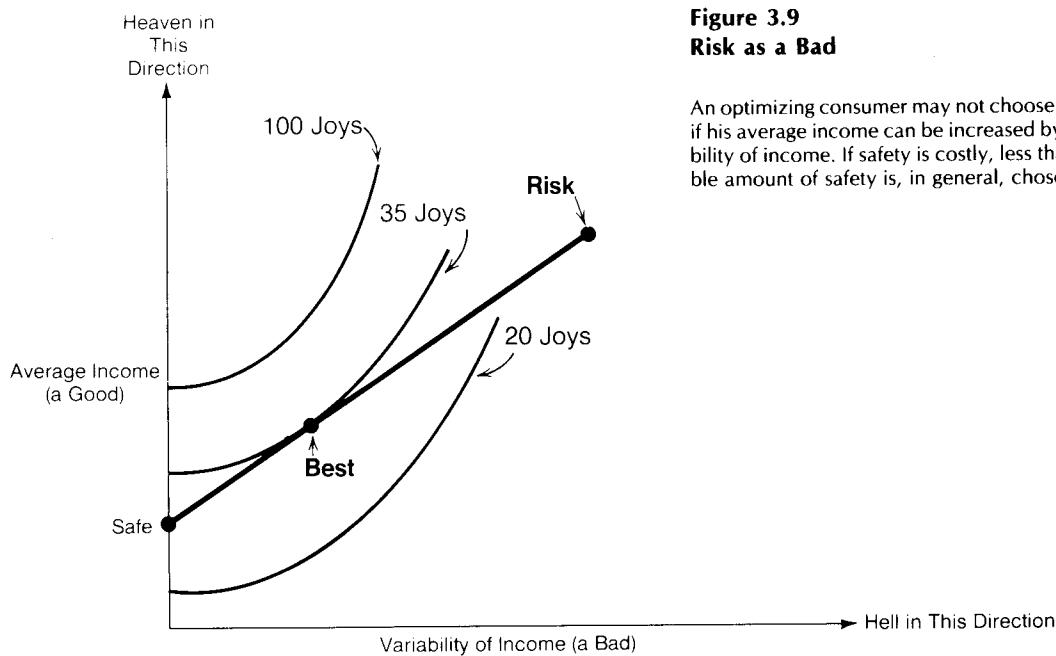


Figure 3.9
Risk as a Bad

An optimizing consumer may not choose a perfectly safe income if his average income can be increased by accepting some variability of income. If safety is costly, less than the maximum possible amount of safety is, in general, chosen. Compare pollution.

to invest in money balances. (*Hint:* Use Figure 3.9, thinking of money balances as being similar to Hill's savings accounts; interest is average income, and the variability of interest causes the income to vary.)

b. What happens to the proportion if the average interest rate on bonds falls but the variability of the interest rate stays the same? (*Hint:* The point like Safe in Figure 3.9 moves down; Risk stays the same).

True or False

- 2. It is impossible to imagine a utility-of-income curve that would imply simultaneously taking unfair gambles for large but improbable gains and taking unfair insurance against large but improbable losses. (*Hint:* Draw one that is concave for low incomes and convex for high incomes, that is, Z-shaped).
- 3. It costs John 1 cent worth of time and trouble to fasten his seat belt on each trip in his car; putting on the seat belt reduces the probability of dying in a crash from 1 in 2 million on each trip to zero; and John's utility curve of income is a straight line. Therefore, if John were just indifferent between fastening the seat belt and not fastening it, he would be willing to be shot one year from now for \$20,000.
- 4. If criminals were risk preferring, crime would not pay. (*Hint:* Set up Figure 3.8 with a curve concave *upward*. When would the criminal do crime? Is his average income higher or lower than honest income?)
- 5. A risk-averse burglar will prefer a harsh but improbable punishment to a mild but more probable punishment if his average income from the two is the same. (*Hint:* Go back to Figure 3.8 again; "more probable" means "Average is closer to Lose".)
- 6. A risk averter will never specialize in holding a risky asset. (*Hint:* Corner solution?)
- 7. A risk averter minimizes risk.

CHAPTER

4

Indifference Curves and Demand

What to Read For

What exactly happens when a consumer faces a drop in the price of something, say transportation? Is he made virtually richer? How much income would the consumer be willing to give up to get the privilege of buying the transportation at the lower price? How do membership fees relate to this amount? What is the substitution effect? The income effect? In an indifference curve diagram, how can a fall in the price of one commodity be broken into the price (or substitution) effect and the income effect? In a demand curve? What is a compensated demand curve? An uncompensated demand curve? How is an ordinary demand curve related to income-compensated curves? What are Giffen goods? Do they exist? What is a "need" for a good?

Indifference curves, then, tell you about the consumer's choice among bundles, in both an uncertain and a certain world. Another major use of indifference curves is less direct, namely, to give an alternative way of looking at the demand curve. You know from the discussion about Figure 1.10 (p. 24) that indifference curves and the demand curve are related. The job here is to look closely at their relation.

Free Purchase or Entry Fee

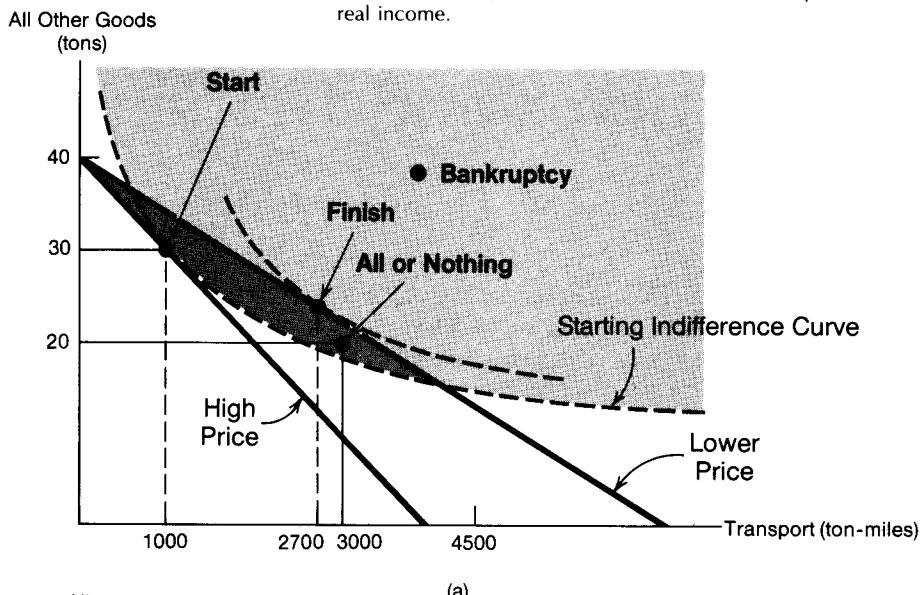
Consider Robert Fogel, a pioneer farmer living in 1869 at Ogden, Utah, near the about-to-be-completed Union Pacific Railway. Fogel buys transportation for himself and his crops by giving up some of his income of all other goods. At the price of transport in terms of all other goods that he faced before the railway (that is, High Price in the panels of Figure 4.1), he purchased 1000 ton-miles of transport at a cost of 10 tons of all other goods, arriving at the point Start. Now the railway opens, and could sell transport at a Lower Price than the teamsters and mule drivers of Ogden whom he bought from before. What happens to the amount of transport Fogel buys?

All or Nothing

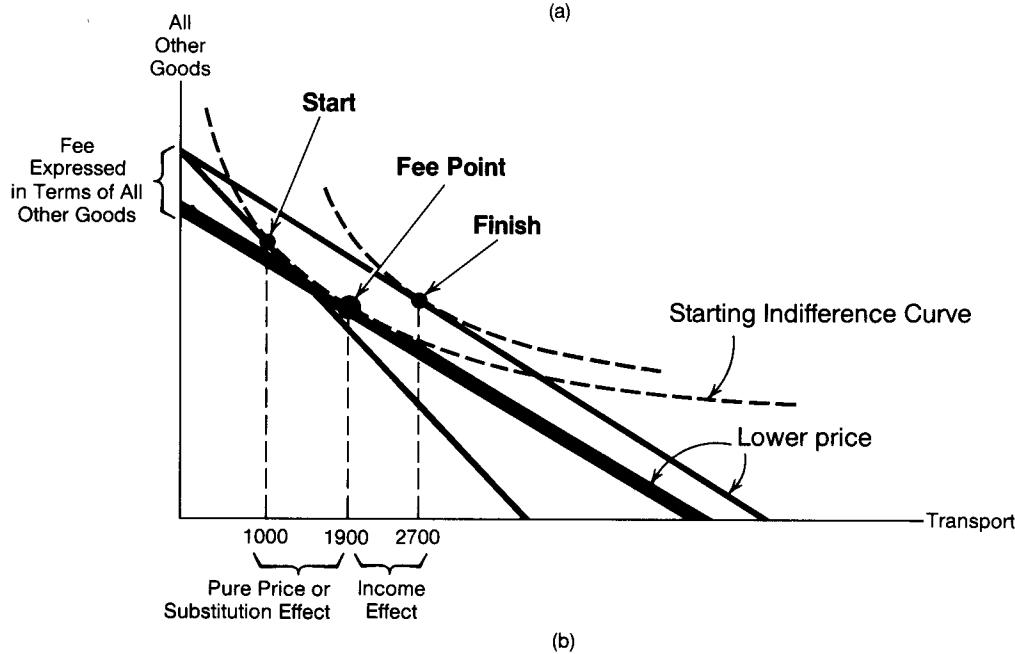
In a general way the answer is pretty obvious: He buys more if the price is lowered. In detail, though, the answer depends on the details of the offer the railway makes and he accepts. Look at the top panel. Fogel will not accept the

Figure 4.1
The Indifference Curves for All or Nothing and for Free Purchase (a) and for an Entry Fee (b)

The consumer is offered a lower price and free choice of quantity, choosing to increase consumption from 1000 ton-miles to 2700 ton-miles. The increase consists of two logically distinct parts: from 1000 to 1900 in response to the change in price alone, holding real income constant, and from 1900 to 2700 in response to the price-induced change in real income.



(a)



(b)

offer if it does not put him somewhere better than or equal to the point Start. He will insist on being in or on the entire shaded region above the Starting Indifference Curve. Think about that for a second: To be worth taking, an offer has to make you better off (or at least no worse off). The railway, on the other hand, will not make an offer that is outside its budget line, Lower Price. The point Bankruptcy, for instance, is an offer of so much transport in exchange for so little all other goods that the railway would be nuts to make it. Think about that for a second, too: *Both* parties must be made better off or it's no go. Somewhere in or on the edge of the heavier shaded area is where Fogel is going to end up. In that area both conditions are met. Fogel is better off *and* the railway is better off.

The railway could say to Fogel, "Look, Bob old pal, we'll sell you 3000 ton-miles of transport for 20 tons of all other goods. If you don't like the offer, you needn't take it, but it's our final offer, Bob ole buddy." He pays 20 out of his endowment of 40 tons of all other goods, and for that he gets exactly 3000 ton-miles of transport, no more, no less. Though it sounds ordinary enough, such an offer is in fact uncommon in the economy, because the seller must be without competitors to make it. It is called an "all-or-nothing" offer, marked in the top panel as All or Nothing. You take all the transport specified in the offer at the price the railway asks or you get nothing from the railway. A contract to buy from Your Friendly Local Natural Gas Monopoly 3 billion cubic yards of natural gas at 1 cent a cubic yard is an all-or-nothing offer. Take it or leave it, "it" being a particular amount at a particular total cost. Big, lumpy purchases of complicated things like houses and vacations have elements of all-or-nothing to them. You buy all the house at 6926 Constance Street, crabgrass and all, or not at all; all the vacation in St. Petersburg, shuffleboard court and all, or not at all. But of course in such cases there are lots of other sellers of houses and vacations trying to get your business elsewhere. Ernestine Austen may tell you "You buy all my house at \$150,000 or nothing," and it's literally true. But you can get another house pretty much the same by going to her competitors. Not so for the Union Pacific, because the mule drivers competing are so much more expensive than the railway.

On the other hand the railway might offer Fogel the opportunity to buy as little or as much as he wanted at a certain price. This is the usual offer in a market. At \$1.10 each you can buy all the hamburgers you want at the local hamburger joint; at \$6 each you can buy all the copies of Wayne C. Booth's *Modern Dogma and the Rhetoric of Assent* you want at the local book joint. You need not buy ten or a thousand to get any at all. Likewise Fogel: The railway might offer him, for instance, the straight budget line Low Price, along which of course he pays a constant price. Choosing to buy 2700 ton-miles of transportation, he comes to rest at Finish. Notice that Finish is the best he can do in or on the heavily shaded area. Of all the feasible offers it is the best for Fogel. Letting him choose how much he wants is always going to be better for him than letting him accept only some crummy all-or-nothing offer more favorable to the railway.

A third possibility is not as common as the what-you-wish offer but more common than the all-or-nothing offer. The railway could charge Fogel a flat fee, unrelated to how much transport he buys, *for the privilege of buying transport at the Lower Price*. The fee is like a cover charge in a nightclub or a membership fee in a club or for that matter a tuition charge at a university. The fee

lets you into a club, so to speak, in which prices of certain things are cheaper than they are outside—dancing (for free), golf (\$5 green fees), or courses in economics (\$50 in books). In the bottom panel of the diagram the Fee is expressed as an amount of all other goods he pays in order to get access to the heavy budget line. The heavy budget line has the same slope as Lower Price, but has lower income by the amount of the fee. That's why it's parallel: same price, lower income. If the railway is powerful (in other words, has no competitors) and if it is smart it can stick Fogel with a Fee so high that he is no better off after buying the railway transport than he was sitting on a mule at Start. He will move to the best position he can, a point on the Starting Indifference Curve marked Fee Point.

Notice that Fee Point is the best the railway can do in or on the shaded area. The railway gets the Fee as net gain, then lets Fogel trade along the heavy price line *at cost*—that is, at no net gain. The *only* net gain to the railway is the Fee. You can see that no other line parallel to the initial indifference curve extracts as high a fee as the heavy line. In view of what supplying transport actually costs the railway in all other goods given up (iron for rails, for instance, or coal for engines), Fee Point makes the most for the railway.

From Fogel's point of view the Fee is the income he would be willing to give up to get the privilege of buying at the lower price. Read that sentence again, putting yourself in Fogel's shoes. The Fee, in other words, is the amount Fogel is made virtually richer if he is allowed to buy at the lower price.

Now the second, what-you-wish offer lets Fogel buy at the lower price *without* paying a Fee. It makes him virtually richer. If the price of a thing you buy falls, of course, you are made virtually richer. By how much? By the amount you would be willing to give up to get the privilege of buying at the lower price. How much is this in Fogel's case? As was just said, the amount of the Fee.¹

If Fogel is permitted to buy out to Finish, he buys 2700 ton-miles of transport, as you can see in the diagram. If he must pay the Fee he only buys 1900. The difference, $2700 - 1900 = 800$, is the *income effect* of the virtual rise in income if he does not have to pay the Fee. Since transport is here a normal good, more is purchased when the lower price is allowed to have its full, good effect. The increase in the amount Fogel would buy even if he *was* charged the fee is $1900 - 1000 = 900$. It is the pure price effect. That is, it is the result of substitution of transport for all other goods that naturally occurs when transport becomes relatively cheaper. Therefore, it is also called the *substitution effect*.

In other words, the way in which a consumer facing the usual offer reacts to a fall in price splits naturally into two parts. On the transport axis, the substitution effect is the move from 1000 to 1900 ton-miles, the income effect is the move from 1900 to 2700 ton-miles. Inside the diagram, the substitution effect is the move from Start to Fee Point, the income effect being the move from

¹ Another answer is possible to the question, how much richer is Fogel made by a fall in price? It is the amount of income (in all other goods) you could take away from him after the fall in price and still leave him enough to buy the initial bundle, Start. This answer is called the Slutsky measure, the answer in the text is called the Hicks measure. They are pretty much the same if the fall in price is not very large.

Fee Point to Finish. The sequence of moves does not actually take place. True, if the railway first made Fogel the offer with the Fee and afterward, in a fit of generosity, gave the Fee back to him, he would move first to Fee Point and afterward to Finish. But the real point is that the increase in transport bought after a fall in price depends on two features of the consumer's indifference map. It depends, first, on how sensitive he is at a given real income to changes in price, the substitution effect (that is, how great is the curvature of an indifference curve). And it depends, second, on how sensitive he is at a given price to changes in real income, the income effect (that is, how much more transport he buys as he moves up to a higher indifference curve).²

**The Offer Curve Is
an Alternative
Representation of
the Demand Curve**

Bearing these notions in mind, consider how Fogel would react to a succession of offers of low prices, such as Low, Lower, Lowest in the top panel of Figure 4.2. With a free choice of how much to buy, he would take up the tangency points indicated, and if such points were connected up, they would form a curve. It is called the *Offer Curve*. The bottom panel plots the *slopes* of the successive budget lines (that is, the relative price ratio) against the quantities of transport to which they lead (as in Chapter 1), giving the solid Demand Curve.

The ordinary demand curve, then, is merely an alternative way of showing the offer curve. They both represent Fogel's best choices when he faces different relative prices, given a fixed amount of money income. By contrast, with the maximum fee charged at each low price, Fogel would move along the Starting Indifference Curve. The dashed demand curve plots the succession of slopes of this indifference curve. Such a demand curve is called *income compensated* because it compensates for (that is, takes away) the income effect. In this case the income effect is the effect of prices falling from High (notice that the two demand curves coincide at High Price). It is the demand curve holding real income constant. The indifference curve through point Start is here taken as the standard for judging changes in real income. But any indifference curve could be the standard. In other words, corresponding to each indifference curve there is an income-compensated demand curve. It is derived by imagining Fogel to be charged fees that keep him on the indifference curve. His free choice demand curve (derivable from his offer curve) connects up such entry fee demand curves (derivable from his indifference curves), as shown in Figure 4.3.

As in the indifference curve diagram, the effect of a fall in the price of food from, say, Low to Lower can be broken into two parts, as the arrows in the diagram and on the transport axis show: the substitution effect, which is the movement down along the dashed income-compensated demand curve that runs through Low; and the income effect, which is the movement out along a fixed price (namely, Lower) to the new, higher compensated demand curve running through Lower.

² If you find the income and substitution effects hard to grasp you have my sympathy. In my junior year in college I failed quizzes on it twice before getting the point, sort of. I didn't really get the point until I started teaching it to other people, five years later. It is not the most important idea in price theory, though not unimportant either. Among the moderately important points it is for some reason the hardest to grasp. For now just get the diagrams straight and watch for the light. It comes in Chapters 7 and 25, with scattered glimmers in other chapters.

Figure 4.2
Offer Curves (a) Correspond to Demand Curves (b)

The offer curve is the set of tangency points between indifference contours and all possible straight budget lines drawn for a given Endowment of all goods. To construct the demand curve from the offer curve, read quantity off the horizontal axis and plot this as a function of price, as given by the slope of the budget line. If the consumer is held on a given indifference contour rather than at a given Endowment of all other goods, then the Income-Compensated Demand Curve results.

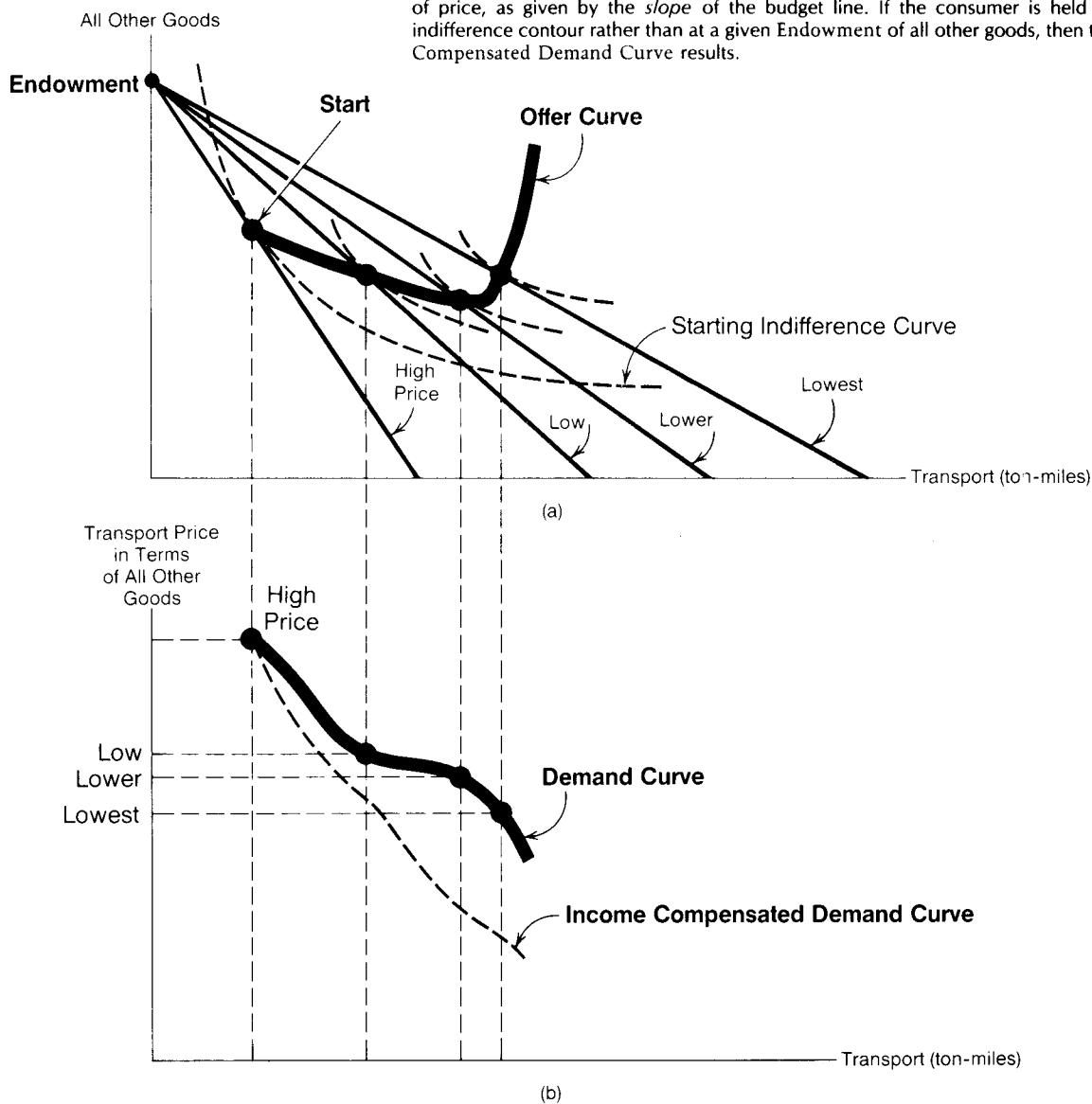
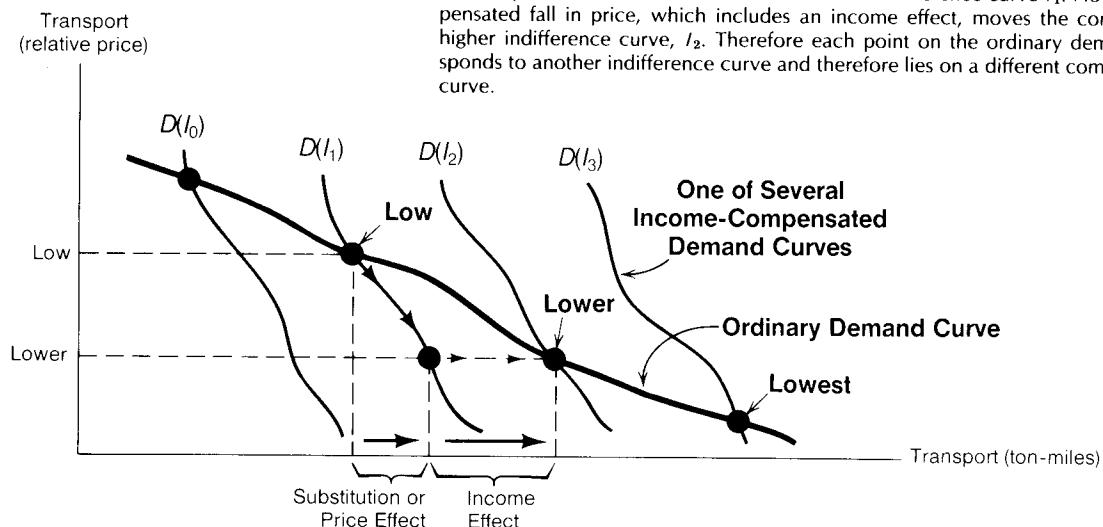


Figure 4.3
An Ordinary Demand Curve Is a Collection of Points on Income-Compensated Curves



An income-compensated demand curve $D(I_1)$ shows only the pure substitution effect of a fall in price because it holds the consumer on indifference curve I_1 . However, an uncompensated fall in price, which includes an income effect, moves the consumer to a new, higher indifference curve, I_2 . Therefore each point on the ordinary demand curve corresponds to another indifference curve and therefore lies on a different compensated demand curve.

**The Uses of
Compensated and
Uncompensated
Demand**

The payoff to the correspondences between diagrams of indifference curves and diagrams of demand curves will become apparent later. Put briefly, it allows one to move easily from speaking of particular markets for transport or food or housing to speaking of their general effects on the whole society's income or happiness. In the present case, for example, the maximum fee that Fogel would pay to be permitted to buy cheap transport from the railway (that is, the income of the income effect) is a measure of the benefit he derives from the opening of the railway. If you knew the shape of his demand curve and of the demand curve of other typical buyers of transport in the nineteenth century (farmer Fishlow in Illinois, for example), you could estimate the social benefit derived from the introduction of railways. This has in fact been done, arriving at the surprising conclusion that the social benefit was small.³ The same techniques, described in Chapter 10 in more detail, can be applied to projects of any sort: a subway in San Francisco, a pipeline in Alaska, a shopping center in Cleveland.

In the present chapter the main payoff to the logic of income and substitution effects is in the production of examination questions of diabolical difficulty:

³ R. W. Fogel, *Railroads and American Economic Growth* (Baltimore: Johns Hopkins, 1964); and Albert Fishlow, *American Railroads and the Transformation of the Antebellum Economy* (Cambridge, Mass.: Harvard University Press, 1965).

T or F: If transport is a normal good for Fogel, his ordinary demand curve for transport has a lower slope than do any of his income-compensated curves.

A: True. Look back at Figure 4.3. The rightward march of the income-compensated demand curves— $D(I_0)$, $D(I_1)$, $D(I_2)$, $D(I_3)$ —expresses the normality of transport. At a given price relative to all other goods Fogel consumes more transport if his income allows him to

The logic involved is clearer if you consider the opposite case.

T or F: If transport is an *inferior* good for Fogel, his ordinary demand curve has a *higher* slope than do any of his income-compensated curves.

A: True. Higher indexes now march leftward, and the virtual enrichment from a fall in price now causes him

reach a higher indifference curve. He will always move to a compensated curve with a higher index—from $D(I_0)$ to $D(I_1)$, say, or from $D(I_1)$ to $D(I_3)$ —when the price falls, for a falling price always makes him for all practical purposes richer. Since normality implies that compensated curves with higher indexes are more rightward, a fall in price is associated with a rightward move in the amount consumed. Not so difficult after all.

to buy less, not more. The income effect of a fall in price from, say, Low to Lower is now negative. It offsets to some degree the substitution effect, resulting in a *smaller* increase in transport bought than would occur if Fogel were fixed by fee to the indifference curve he had attained at the Low price.

Giffen Goods

As was shown in an elementary way in Chapter 1, the substitution effect of a fall in price is always positive. It is conceivable that food or starchy foods or potatoes or Irish potatoes in 1846 could be so inferior that the negative income effect would overwhelm the positive substitution effect. The English economist Alfred Marshall awarded (incorrectly) the credit for realizing that this was conceivable to Sir Robert Giffen. In the *Giffen paradox*, a demand curve would have a perverse, upward slope.

The Law of Demand, then, can be put as follows: There are no Giffen goods. George Stigler has offered a characteristically economic bit of reasoning on the empirical success of this proposition:

If an economist were to *demonstrate* its failure . . . he would be assured of immortality, professionally speaking, and rapid promotion. Since most economists would not dislike either reward, we may assume that the total absence of exceptions is not from lack of trying to find them.⁴

The Peculiar Economics of Need, Priority, and Other Pathologies in Demand

If you know something definite about the shape of indifference curves (such as that the income effect is strongly negative, as in the Giffen paradox), you can, of course, say something definite about the shape of the corresponding demand curve. A particular shape lies behind the economic thinking of the man in the street. It turns on the word “need,” a word that economists find much more difficult to use than do statesmen, journalists, and other people innocent of economics. Perfectly ordinary sentences which might occur in any political speech or newspaper editorial leave many economists unhappy: “We *need* natural gas, a *basic commodity* that we *cannot do without*. Providing this *necessity* at a price people can *afford* should be high on our national *priorities*.” The nub of the issue is the shape of indifference curves and the corresponding demand curve.

⁴ George Stigler, *The Theory of Price*, 3rd ed. (New York: Macmillan, 1966), p. 24, his emphasis. The fable that Irish potatoes in the terrible famine of the 1840s is a case in point is demolished by Gerald P. Dwyer, Jr., and Cotton M. Lindsay, “Robert Giffen and the Irish Potato,” unpublished paper, Emory University.

Q: Draw Gary Hawke's indifference curves between food and all other goods and his demand curve for food (given an income in all other goods) if Hawke needs 1 pound or more of food a day to survive.

A: Consider Figure 4.4. The indifference curves are all outside the Shadow of the Valley of Death. Outside the

Valley, Hawke has an ordinary demand curve; inside it (that is, at a price of food such as Death Price so high that he literally cannot "afford" to buy 1 pound), he is dead and, needless to say, has no demand curve.

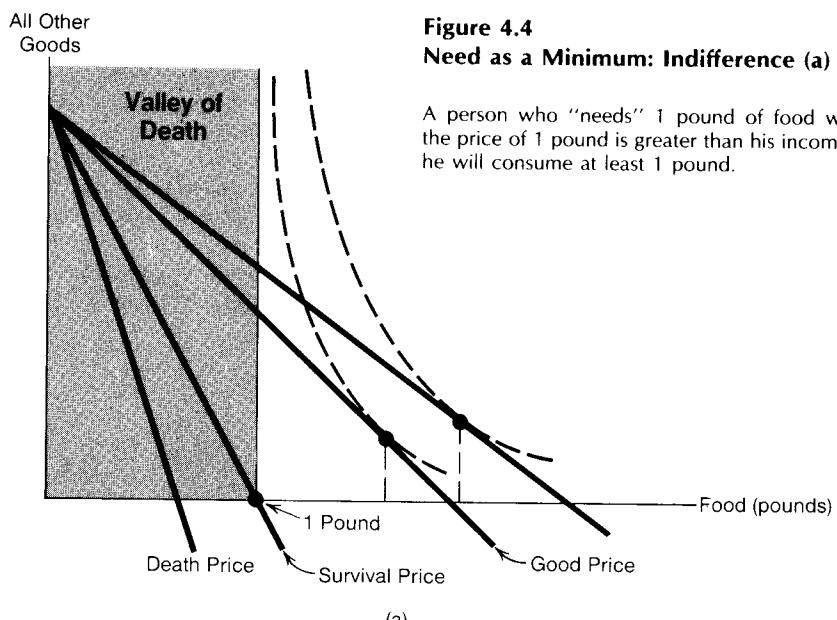
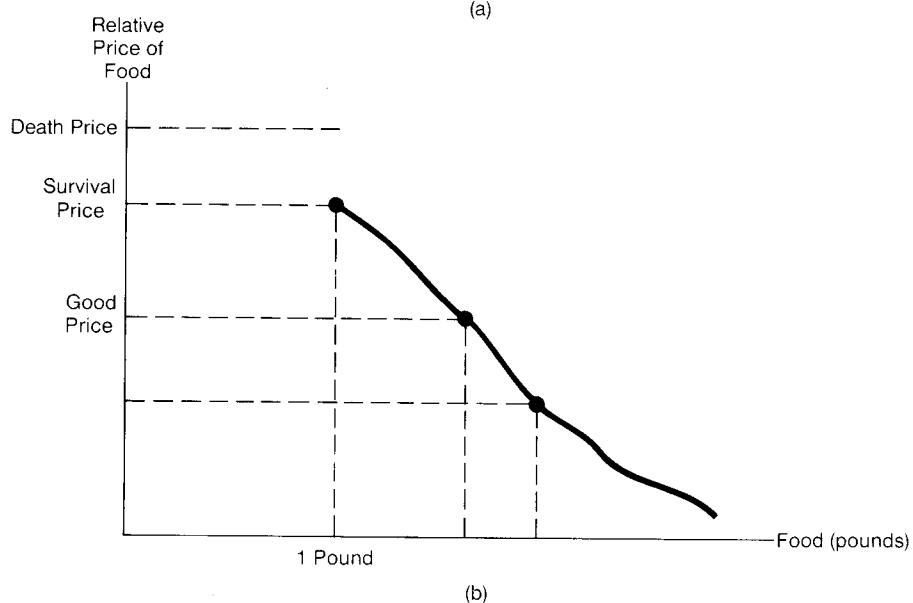


Figure 4.4
Need as a Minimum: Indifference (a) and Demand (b) Curves

A person who "needs" 1 pound of food will never voluntarily consume less. If the price of 1 pound is greater than his income, he will (involuntarily) die; otherwise he will consume at least 1 pound.



This is one interpretation of "need": some minimum amount below which utility is zero or at some unpleasantly small amount. The mental experiment that people perform when they assert that food or steel or health care is "basic" is to imagine life without them, below the need point. If you ask a noneconomist whether health care or rock music is the more appropriate object for veneration, regulation, and subsidy, he will answer, "health care." If asked why, he will answer, "because it is a basic need." But the needfulness of health care is irrelevant to social choices beyond the need point. Most social choices are choices between more or less health or rock music, not between some or none. The mental experiment behind the notion of "basic needs" is usually the wrong one.

A slightly different meaning of need lies behind the word "priority" (which despite its technical sound is never used in economics).

Q: Larry Neal values only food until he has consumed 2 pounds; beyond this point he values only shelter. Draw his indifference curves between food and shelter.

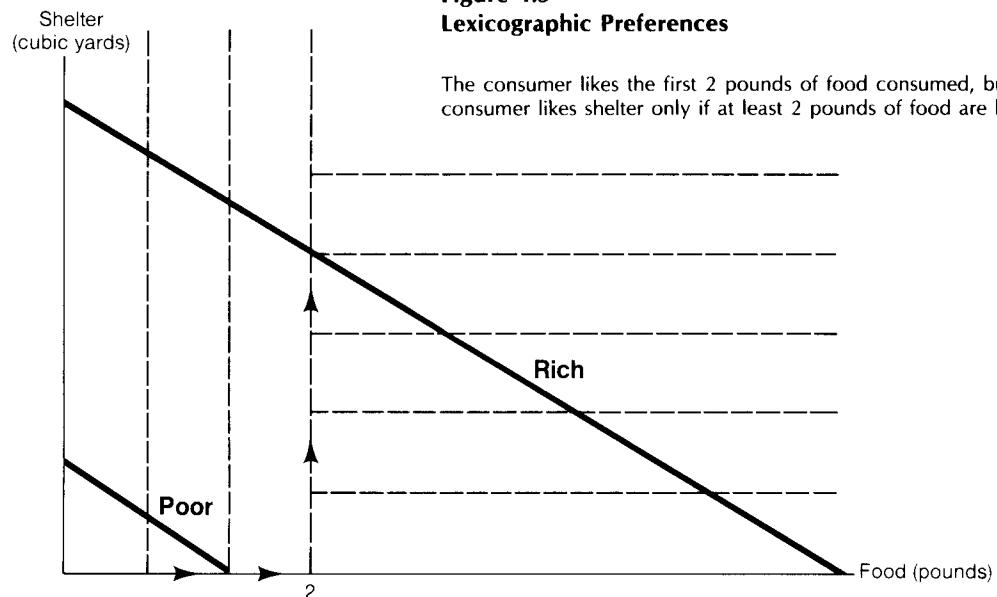
A: When 2 pounds of food is unattainable (outside a budget line such as Poor in Figure 4.5), he consumes

only food; when it is attainable, he consumes 2 pounds of food and spends all his remaining income on shelter. In other words, food has a higher priority than does shelter. First he buys food. Only when he can buy all the food he needs (namely, 2 pounds) does he buy shelter.

COMMENT Although the generalization cannot be drawn in two dimensions, the idea can be generalized to a list of successive needs, such as food (1), shelter (2), clothing (3), . . . , Sony Walkman (350). It is influential in psychology, where it is called the "hierarchy of needs": in ascending order, physiological, safety, social, ego, and self-fulfillment needs. It is not influential in economics, where it is called *lexicographic preferences* (if you look up "lexicographic"

Figure 4.5
Lexicographic Preferences

The consumer likes the first 2 pounds of food consumed, but not any other pounds. The consumer likes shelter only if at least 2 pounds of food are being eaten.



in the dictionary, you will first look at the "L's"; then second, having met that need, the "E's"; then third the "X's," and so forth). Such preferences violate common economic observation. It is not true that the poor live by bread alone, nor, equivalently, is it true that the demand curves of all commodities become perfectly vertical at low prices or high incomes (that is, after the need for something can be satisfied). And even were it true that the preferences of each consumer were lexicographic, if consumers did not all have the same preferences, consumers as a group would behave as though they had ordinary preferences. In short, economists do not believe it is useful to speak of "priorities" and suspect that this word, as is true for the phrase "basic need," is merely a verbal club with which to batter people who do not value a new nuclear weapon or a new program for the poor as highly as does the speaker. Interpreted literally, these words of noneconomic economics entail implausible or irrelevant shapes of indifference and demand curves.

Summary A demand curve in a diagram of the quantity bought of transport and the relative price faced corresponds to an offer curve in a diagram of quantity bought of transport and quantity bought of all other goods. The distinction between an offer of a free choice and an offer of a fee to be able to have the choice leads to the distinction between income and substitution effects. A compensated demand curve contains only the substitution effect; an ordinary demand curve contains both the substitution and the income effects.

The distinction between the income and substitution effects of a fall in price is applicable to the pathologies of demand curves: demand curves that slope upward rather than downward (the Giffen case) or demand curves that are vertical (at the point of fulfilled needs). The pathologies are factually unimportant and are useful chiefly as mind-expanding exercises in the connection between indifference and demand curves.

EXERCISES FOR CHAPTER 4

1. *True or false:* Fogel would like the railway to choose the offer marked Bankruptcy in Figure 4.1.
2. *True or false:* The railway would like Fogel to choose an offer below his Starting Indifference Curve.
3. *True or false:* The railway could and would offer Fee Point as an all-or-nothing offer. (*Hint:* Remember that Fee Point is the *best* the railway can do, considering that Fogel will stay on or above his Starting Indifference Curve.)
4. If you spent half of your budget of \$6000 on college tuition, what would be the virtual fall in your income if tuition increased 20%? If you spent 1% on paper and pencils, what would be the virtual fall in your income if the prices of paper and pencils increased 20%? Is the income effect of an item that is a small part of your budget likely to be large?
5. *True or false:* Offer curves are the same shape as demand curves.
6. *True or false:* Demand curves, like indifference curves, are convex (bulging) inward to the origin.
7. When you really and truly "need" a new stereo your demand curve for it does *not* slope down; it is a vertical line. No price, no matter how high, will cut your demand for

the good; and no price, no matter how low, will increase your demand for it. In this sense do you *need* a new stereo? A piece of steak? A course in economics? A college degree? An apartment? Food? An appendectomy when suffering from acute appendicitis?

PROBLEMS FOR CHAPTER 4

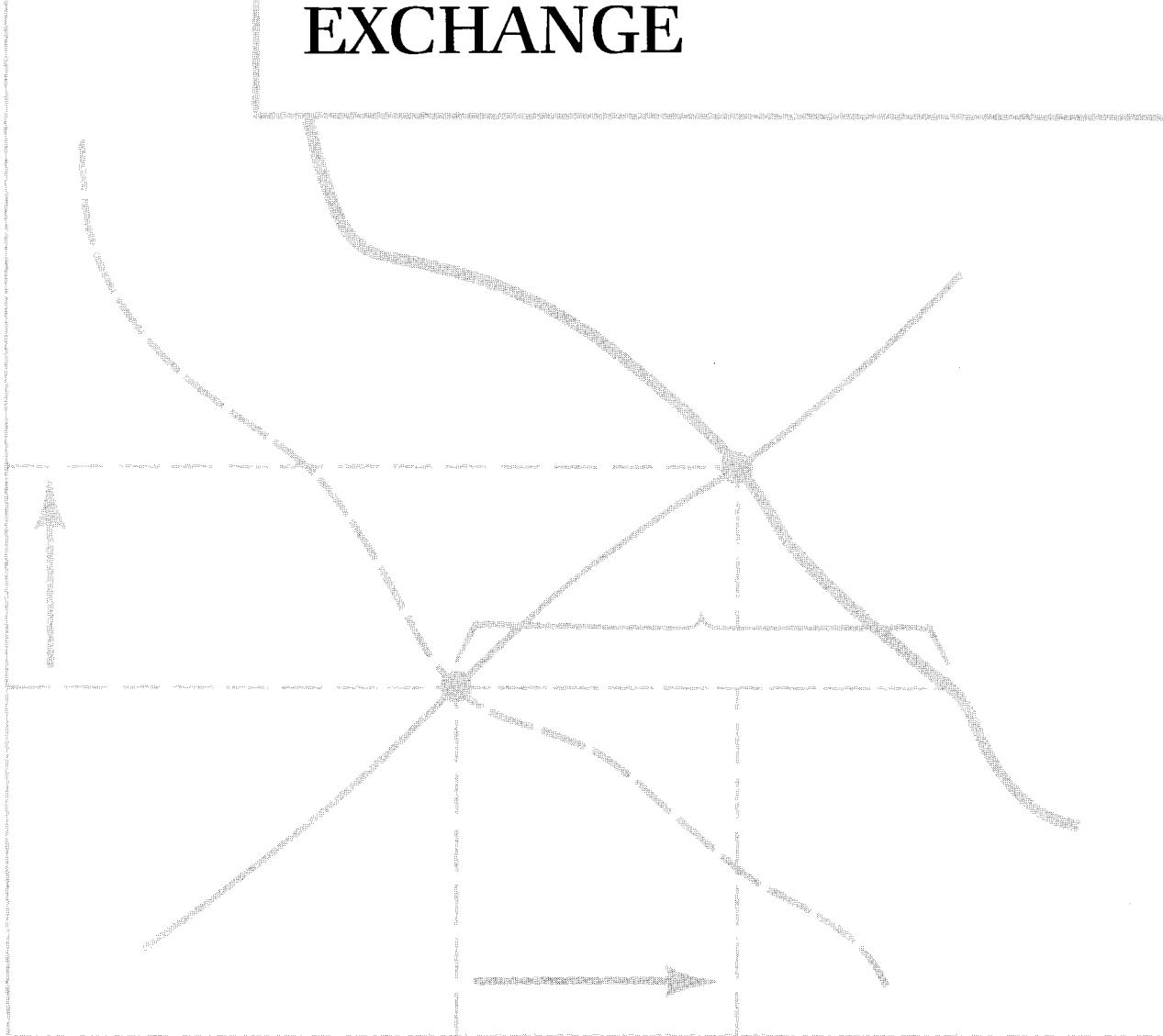
1. Movie theaters typically offer candy bars in large sizes only. Alice Galenson would like 2 ounces of candy at 10 cents an ounce but must buy 8 ounces at 40 cents. *True or false:* This is an example of an all-or-nothing offer. (*Hint:* Once inside the theater does Galenson have any competitor to go to?).
2. Universities typically discourage part-time students, either by charging them higher prices per course or by insisting that all students be enrolled in a degree program. *True or false:* This is an example of an all-or-nothing offer. (*Hint:* All the house at 6926 Constance Street or none?)
3. The Woodlawn Bar and Tap, popularly known as Jimmy's, can serve each drink at a cost to Jimmy of 25 cents.
 - a. Draw the budget line and indifference curves of a patron if he faced the 25 cent (zero profit) price. How much would he buy?
 - b. Pick any point inside the budget line of (a). If the patron chooses this point, what would be Jimmy's profit (expressed in all other goods) from the patron's patronage? (*Hint:* It is a certain vertical distance.)
 - c. Now draw the indifference curve running through the patron's endowment of all other goods. Can Jimmy force the patron to a point below this curve? What point along it maximizes Jimmy's profit? (*Hint:* A vertical distance between any two curves reaches a maximum when the slopes of the two curves are equal along a vertical line, that is, when the curves are vertically parallel.)
 - d. Describe on the one hand a cover charge and price per drink and on the other a required minimum number of drinks and price per drink that would leave the patron at the point described in (c).
 - e. What would happen to these schemes if Jimmy's faced competition from other bars?
4. A heavy smoker climbs the walls if he has too few cigarettes a day and coughs convulsively if he has too many. Draw his demand curve for cigarettes.
5. Betsy White can trade off hours of leisure for tons of all other goods (namely, food, housing, and so forth) by selling the hour of leisure at the going wage per hour. She is endowed, of course, with 24 hours of leisure. Draw White's offer curve for leisure by pivoting her Leisure–All Other Goods budget line around the point of 24 hours of leisure. Draw the corresponding demand curve for leisure. Note that the supply curve of hours of labor is simply 24 hours minus the demand curve for leisure. Does the supply curve ever bend backward; that is, does the number of hours supplied ever fall when the wage rises?

True or False

6. The standard 8-hour day is an example of an all-or-nothing offer.
7. If Eric Jones's indifference curves have the usual shape (that is, a convex shape), then his demand curve for books necessarily has a convex shape.
8. If housing and all other goods are perfect complements, then the income-compensated demand curves for housing are vertical straight lines.

II

EXCHANGE



Trade

5.1 Supply and Demand

What to Read For How can a demander of goods at the same time be a supplier? What is the *veil of money*? Do consumers value the goods that money can buy, or the little green pieces of paper the money is painted on? How can one relate the usual indifference curve diagram to a consumer's demand curve for one commodity and her supply curve of all other goods? Is a *balance-of-payments deficit* a Bad Thing? Is American self-sufficiency good? At what price does one become a demander? At what price a supplier?

To Demand Is to Supply

In the ordinary way of speaking an American buying fifteenth-century Italian paintings abroad is a "d demander." At \$5000 per painting John James might demand 10 paintings. But he is in fact simultaneously a supplier, for he supplies the foreigner with \$50,000. That is, he supplies \$50,000 worth of American goods. The money throws a veil over the essence of the transaction, which is a swap of 10 paintings for some computers or wheat or whatever bundle of all other American goods the foreigner buys with the American money. That the American buyer of the paintings is not literally a maker of computers or a grower of wheat is irrelevant. In his job of oil drilling or museum curating he has acquired the power (the money) to purchase these other American goods the foreigner demands. The giving of money makes it unnecessary for James to lug around wheat and computers in his trips to foreign art dealers. But in all essential respects that is what money allows him to do. Buying and selling, to repeat, are swaps of one good for another, thinly veiled behind money.

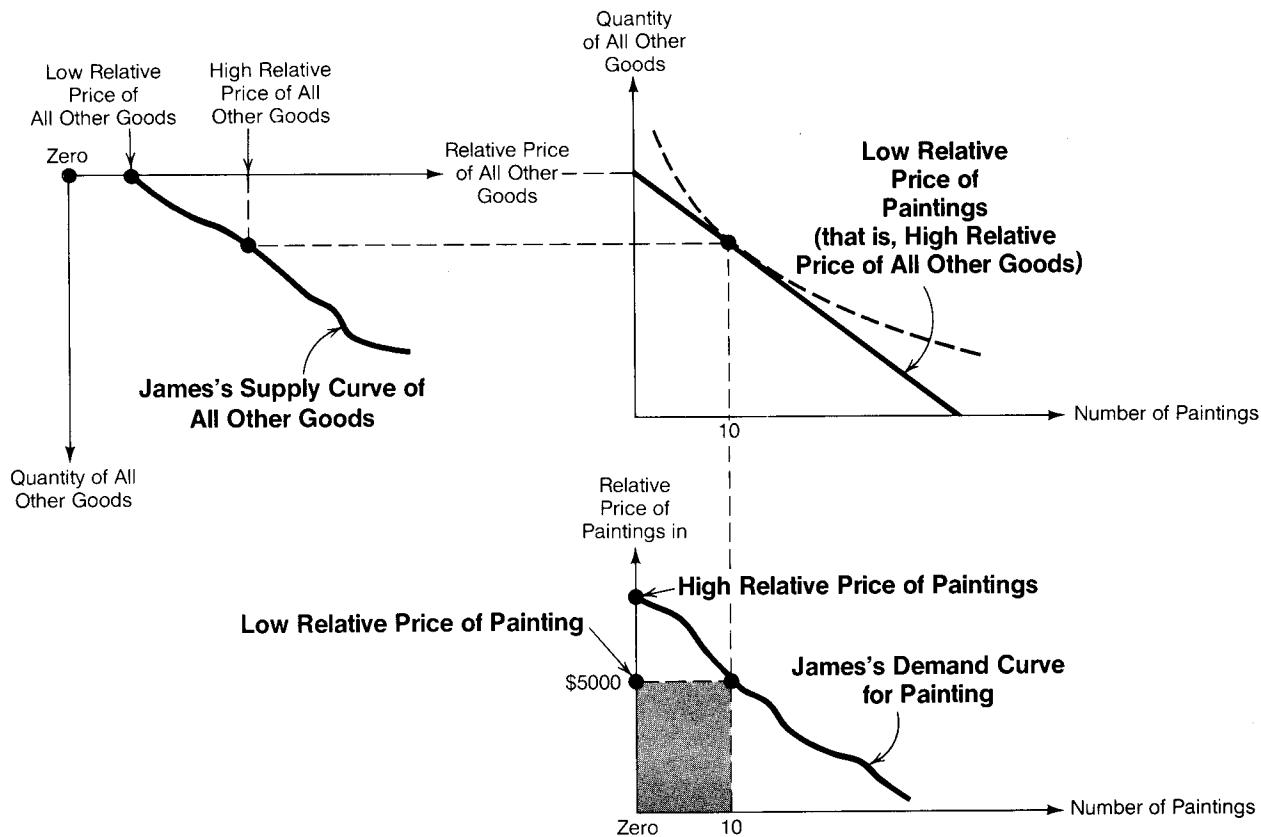
This argument has already been used in the discussions of demand. Consumers have been said to trade (supply) all other goods (not money) for food; consumers have been endowed with incomes in all other goods (not money); and the price of food relevant to its demand curve has been expressed relative to all other goods (not money). Exchanges are exchanges of goods for goods, not "really"

goods for money. People value the goods that money can buy, not the little green pieces of paper.

Figure 5.1 depicts the argument that a demander is also a supplier. It shows to the left of the usual indifference curve diagram James's *supply curve* of all other goods. Looked at from the side it is an ordinary supply curve. At some Low Relative Price of all other goods (a High Relative Price of paintings), James supplies Zero all other goods; that is, he demands zero paintings. At a High Relative Price of all other goods, the position portrayed inside the diagram, he does supply some all other goods, that is, he demands some paintings (10, in fact). Corresponding to the demand curve for paintings there is a supply curve of all other goods. The supply curve slopes upward, not downward, as you can see by turning the book counterclockwise on its side. The supply curve

Figure 5.1
A Demand for X Is a Supply of All Other Goods

Only a price above the Relative Low Price of All Other Goods, the consumer will trade some of his endowment of all other goods for paintings. He will be a supplier of all other goods and a demander of paintings.



and the demand curve are merely alternative ways of looking at the same act of exchange by James.

It is a sad commentary on the failure of economists to educate lawmakers that this simple idea is applicable to great issues of policy.

T or F: If international trade were limited to trade in goods and services (not also in gifts, IOUs, and money balances), it would be impossible for the United States to succeed in a policy of increasing its exports and reducing its imports.

A: Under such circumstances, the United States could not supply wheat to the world without simultaneously demanding shoes, small cars, and the other things with which the world pays for the wheat. Therefore, true.

The point is that exchange is a move along a budget constraint. The value of the goods you provide to someone else in an exchange must be equal to the value of the goods you receive from that person. When applied to whole economies, the point is known as one version of *Say's Law*.¹

In fact, you can go further, allowing for trade in money as well as in goods.

In exchange for goods

Q: The United States has had a balance-of-payments deficit with Japan. The value of American wheat and so forth exported has been less than the value of Japanese autos and so forth imported into the United States. The Japanese have cheerfully accepted American dollars to make up the difference. The deficit has alarmed American policymakers. *True or false:* Their alarm is misplaced, for Americans as a whole are made better off by the deficit: The Japanese accept green pieces of paper in exchange for autos, television sets, and other objects of American demand.

A: The correct answer is contrary to what you may have gathered from the evening news. Dollar bills cost American society the trivial cost of printing them. If the Japanese lust after them to the extent of offering real goods to get them at the value printed on their faces, fine. The willingness of the Japanese to give \$1 worth of goods for a piece of paper costing a few cents to produce allows Americans to consume outside their real budget line between autos and wheat.² The veil of money makes an exceptionally good bargain look like an alarmingly bad one. In short, true.

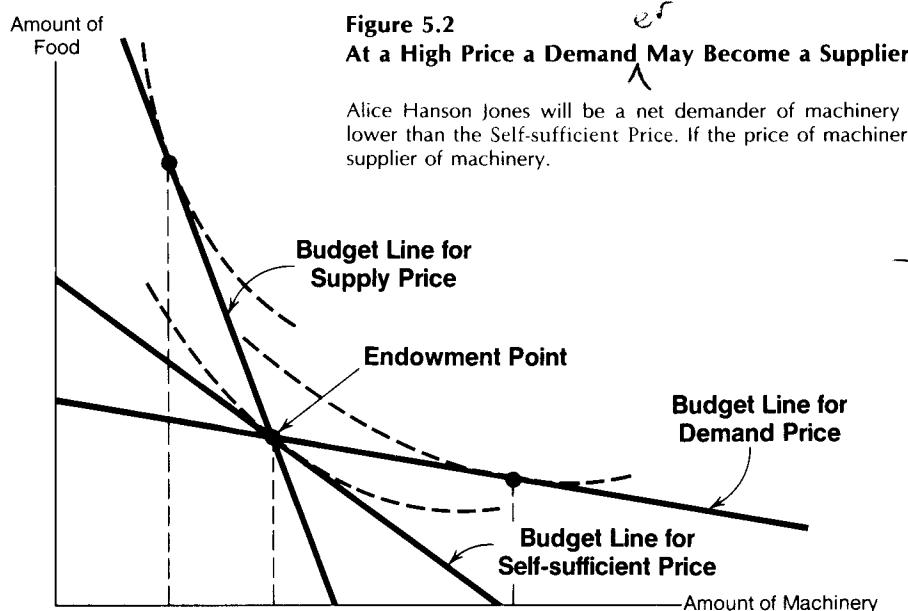
A Demander at One Price Might Be a Supplier at Another

For a person or a country, then, the budget constraint connects the demand for one good with the supply of all others. Furthermore, a person with an initial endowment of, say, food and machinery will at some prices be a supplier of machinery and at some prices a demander. Alice Hanson Jones is happy with her endowment only if she faces a price, labeled *Self-sufficient Price* in Figure 5.2, that offers her no better indifference curve. That is, only at one price is Alice satisfied with the machinery she begins with.

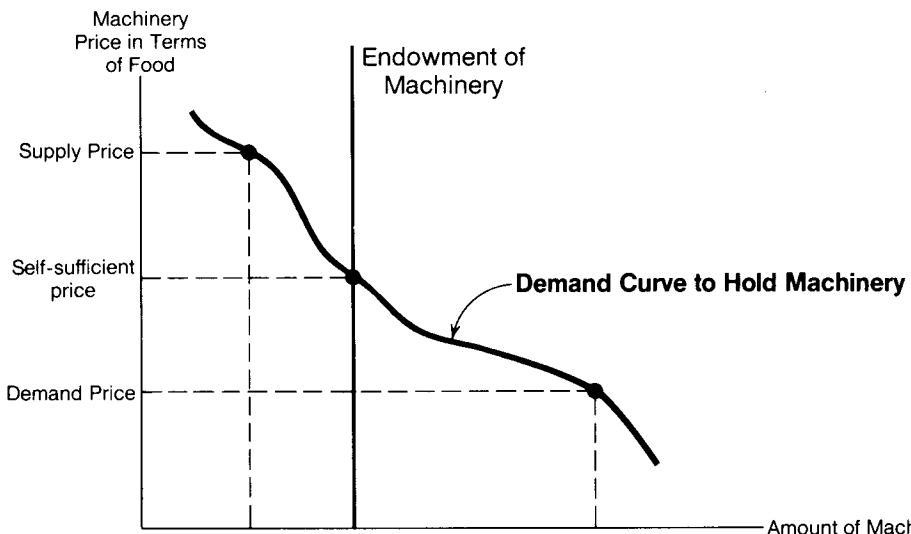
Look at the bottom panel. At a low price, such as *Demand Price*, she demands more than she has initially; at a high price, *Supply Price*, she demands less. At the high price of machinery she willingly gives up some machinery (supplies it) to acquire more food. The portion of her demand to hold machinery above *Self-sufficient* is a supply curve, where the quantity of machinery supplied in-

¹ After Jean-Baptiste Say (1767–1832), the first French academic economist, immensely popular in his time for the sterile lucidity and system he brought to economics (in contrast to the fruitful chaos in contemporary British economics).

² That is, it allows them to violate *Say's Law*, while obeying *Walras's Law*, namely, that the summed monetary value of the goods and *money* you provide to someone else must be equal to the monetary value of the goods and *money* you receive from that person. M. E. Leon Walras (1834–1910), pronounced “vall-ras,” a little like the well-known aquatic mammal was a French academic economist, immensely popular after his time for the sterile lucidity and system he brought to economics (in contrast to the fruitful chaos in contemporary British economics).



- idea of an endowment (= bundle)
- got an indif. curve
- if P hi, what?
- so: people S-evs
or D-evs depending
on P



creases from right to left starting at Endowment of Machinery. The upshot is that the analysis of demand over which we have labored since page 1 is also an analysis of supply.³ Alice's demand curve for machinery gives the amounts

³ The earlier diagrams yielded only demand curves (and not, as here, supply curves as well) because the endowment was always placed on the vertical axis instead of inside the diagrams. The endowment was the income in all other goods. An endowment on the food axis in the present diagram would signify a zero initial amount of machinery. You cannot supply what you do not have.

at various prices she will demand from sellers. Her supply curve of machinery—which is an upward continuation of her demand curve when she has an initial endowment of machinery to sell—gives the amounts at various prices she will supply to buyers.

Q: A peasant grows 125 bushels of grain and 50 pounds of wool. *True or false:* He will buy some more grain if his marginal valuation of grain at 125 bushels is higher than the market price of grain, and he will sell some of his own grain if his marginal valuation is lower than the price.

A: The marginal valuation (relative to wool) is the slope of his indifference curve. If the (relative) price is equal to this slope at 125 bushels, he remains self-sufficient. At a higher price, he is a supplier, at a lower price a demander. Self-sufficiency is a matter of prices. True.

Summary The inquiry into consumer budget lines and tastes has arrived at exchange, that is, at supply and demand. A consumer's demand curve is also the consumer's supply curve, in two senses. First, to demand something Alice Jones must supply something else. That she usually supplies money obscures the nature of the exchange: goods for goods, with money as merely a convenient intermediate step. Second, a consumer demanding a car at one price will at another price supply it. Jones must, of course, have a car to supply if she is to supply it. Her supply-demand curve is simply her initial endowment of cars minus her demand for cars. If she demands many cars relative to her endowment (the price is low, she is rich, or her endowment is small), then she will have an "excess demand" for cars; if she demands few, then she will have an "excess supply."

EXERCISES FOR SECTION 5.1

1. If money didn't exist, how would John James buy paintings?
2. Explain why Alice Hanson Jones is not satisfied with her Endowment Point if the relative price of food in terms of machinery is anything other than Self-sufficient Price. Use the top panel of Figure 5.2.
3. Do the amounts of the endowments in the question in the text about the peasant matter at all to the answer?

PROBLEMS FOR SECTION 5.1

1. Rondo Cameron has and consumes oats and books in Lochielside, Scotland. Because transportation costs to and from the market town are high, the relative price he faces for oats is high if he buys oats and low if he sells oats. Describe his situation, supposing the cost of transporting books to be trivial. What will happen if the cost of transporting oats falls?
2. Transportation costs are like taxes. Suppose that Ann Harley grows wheat and owns some books, each of which she might trade to get more of the other if the price is right. Suppose that at her point of self-sufficiency (where she wishes to be neither a buyer nor a seller) she values one book at four bushels of wheat. Suppose finally that the government of Ontario takes for taxes exactly half of anything she buys. Will she deal in the market if the price before tax in the market is one book for three bushels of wheat? One for five bushels? Beyond which two prices of books in terms of wheat will she enter the market? (*Hint:* Try her out buying wheat by selling books at the two prices; then buying books).

Now we
put them
together
to see what
happens. I
middleman
is not mere
fifteen.

3. If the "money" in the problem about the balance-of-payments deficit was not paper currency (which is cheap to make) but gold (which is expensive to make), would the answer be the same? That is, would it be a Good Thing to have a deficit?
4. Develop an argument that "money is a veil" if it does not appear in the utility function (that is, the utility function has amounts of cars, television sets, and so forth in it, not amounts of cash and checking accounts).
- 5. The notion that middlemen and the like are nonproductive is deeply rooted. A man growing beans is said to be productive; a man selling beans, a parasite. Such attitudes have driven many a New York ad man to dream of working a little farm in Vermont. But consider:
 - a. If our economy consisted of two people, farmer Haines and shoemaker McQuaid, how many transactions would there be (counting an exchange of beans for shoes as one transaction)? If it consisted of three people, Haines, McQuaid, and pencilmaker Menard, how many transactions would there be (suppose throughout that each person would want to buy at least a little of every other person's wares)? If four, how many? If five? If six?
 - b. [Optional] Work out the general formula for a society of N people, supposing that each of the N must go around to each of the others to get his supply of beans, shoes, pencils, movies, education, windows, cars, insurance, and so forth. [*Hint:* The N th person opens communication with $N - 1$ people, giving $N - 1$ links. The $(N - 1)$ th person opens it with $N - 2$ people, because one of the people is the N th person, who has already established a link. Likewise, the $(N - 2)$ nd person opens it with $N - 3$ people, since two links with him have already been established (by Mr. $N - 1$ and Ms. $N - 2$). What is the sum? Check it with the simple cases of $N = 5$, $N = 6$.]
 - c. Now suppose that Donald Gordon, a grasping middleman of the very worst sort, opens a General Store to which everyone goes to buy and to sell, no longer having to go to each other individually. How many transactions (links) are there now? Compare it with the number without the middleman. How do you feel about middlemen now?

5.2 Exchange Between Two People or Nations

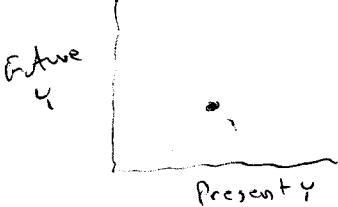
What to Read For	Why is exchange <i>mutually beneficial</i> ? How can this be represented with indifference curves? Is a starving farmer who takes a loan from a moneylender in order to survive being "exploited"? Is perfect knowledge necessary for exchange to be mutually beneficial? What is an <i>Edgeworth box</i> ? What points within the Edgeworth box are unacceptable to each bargainer? What is the <i>contract curve</i> ? What is a <i>Pareto optimal point</i> ? An <i>efficient point</i> ? Is there any best point? Does the <i>Edgeworth box</i> tell where bargaining between two people will end up? Why?
------------------	--

Exchange Is Mutually Beneficial

(27e of David
Suckman &
farm debt)

A perfectly ordinary but widely applicable thought about an exchange is that both parties to it must be made better off. Force, fraud, or mistake aside, an exchange that happens must be one that is as economists put it *mutually beneficial* or *mutually advantageous*. Otherwise it would not happen. It takes two to tango, and if either partner is made worse off, he or she will sit out the dance.

Ordinary and obvious though this first principle of exchange is, it is commonly



misunderstood. The starving farmer taking a loan from the moneylender is commonly said to have "no choice" but to take the loan at an "exorbitant" interest rate from the moneylender, who is "exploiting" the farmer and making him a "virtual slave to indebtedness." Most economists have difficulty with these words. The farmer does have a choice, albeit a dismal one: starve. He is made better off by the loan. In certain technical senses to be described in Chapter 24 he may well be "exploited" by the moneylender, but not in the sense of being made worse off by the "exorbitant" interest rate or his "virtual slavery" than he would be if the moneylender did not make the loan and left the farmer to starve. The phrase "mutually advantageous exchange" has a cheerful sound. But it promises only that exchange makes people better off, not that it makes them best off in the best of all possible worlds.

Q: John Pynchon was the richest man in seventeenth-century Springfield, Massachusetts. He made loans to the townsfolk, many of whom died in debt to him or lost to him their pitiful scraps of land. *True or false:* Pynchon was evidently exploiting the townsfolk.

A: He and the townsfolk were engaging in mutually advantageous exchange of present for future goods. That the townsfolk died in debt to him is evidence that they could get present goods in exchange for unfulfilled promises of future goods, escaping debtless to a better world. Therefore, false.

As it happens, Pynchon charged interest rates well below the prevailing rates elsewhere on the frontier. Were interest rates the appropriate standard, one could say that the townsfolk were exploiting him. But in return for his liberality on this score he got deference and political power (supplementing the political power granted to him by the Massachusetts General Court). The loans were mutually advantageous. Miserable as the townsfolk were, their voluntary choice to take the loans indicates that they believed the loans made them better off.

The exchangers, however, must know what they are getting into. When either side to a bargain is foolish or badly informed, the exchange can hurt. Some cars, for example, are "lemons," that is, bad ones that look in all respects aside from their repair bills the same as good ones. Owners of lemons naturally try to sell them to other people, which loads the secondhand market with lemons. Owners of good cars cannot sell them at a high enough price because there are so many lemons selling at the same low and lemony price (a lemon, remember, looks the same as a good car). The result is the sale of a lot of lemons worth less than their price. Ignorance of the quality of a car that is for sale will hurt the buyer.

But the argument must go further. If people are aware of their own ignorance—the first stage of true wisdom—then they will not buy any secondhand car, knowing that it must be a lemon. The exchange will simply not happen, one way of avoiding the hurt of a bad exchange. Or dealers who can determine whether a car is a lemon will spring up, guaranteeing the car (at a price). Or knowledge that one is buying a lemon if one buys from Sam's Fly-by-Night Auto Sales will become widespread, and the price of the cars that do end up on Sam's lot will fall closer to their true value. Perfect knowledge is not necessary for exchange to be mutually beneficial, only knowledge that one is not perfectly knowledgeable.⁴

⁴ The analysis is that of George Akerlof's "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84 (1970): 488–500, a brilliant and lucid essay accessible to students of price theory.

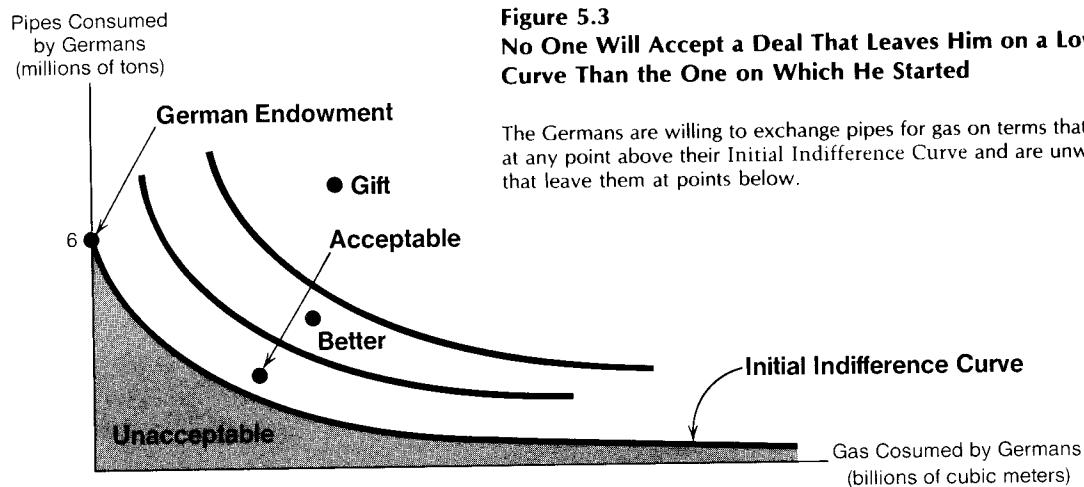
The Beneficability of Exchange Exhibited by Indifference Curves

"Open-eyed" exchange, then, is beneficial. Getting beyond this first principle requires indifference curves. What appears to be the simplest market group is two people trading with each other. Some time ago West Germany and the Soviet Union concluded a deal whereby a certain amount of Soviet natural gas was exchanged for a certain amount of West German steel pipes. The question is, how were the "certain amounts" arrived at? Consider the German indifference curves between gas and pipes (shown in Figure 5.3).

The Germans start at the point German Endowment, with a certain number of pipes and no natural gas. Needless to say, they would like the Soviets to offer their natural gas at a very low price in terms of pipe, allowing the Germans to move, for example, to the point Better. Indeed, they would like the Soviets to offer it at a zero price, or still better a negative price, making a gift of not only some gas but also some steel pipes of their own (acquired from the Japanese, say), leaving the Germans at a point such as the point Gift or beyond. If necessary, however, the Germans would accept any exchange that improved their welfare, that is, any exchange that put them on a higher indifference curve than the initial indifference curve. Any point outside the shaded area is such an exchange. The Soviets, undoubtedly with a different set of indifference curves, would behave in a similar way. The Soviets start with some gas but no pipes. Again, they would like the other party to sell its goods at a low, zero, or negative price. So, too, some shaded area is unacceptable to them. If a mutually advantageous deal is possible, it must fall within the acceptable area for both the Germans and the Soviets.

The Edgeworth Box

What is needed is a device for exhibiting the acceptable areas simultaneously, exhibiting the two consuming nations bumping together and constraining each other's behavior. The device is called an *Edgeworth box*, after the same F. Y. Edgeworth who a century ago invented indifference curves. It is evident that any actual exchange reduces the natural gas the Soviets have by the same amount that it increases the amount the Germans have; likewise for steel pipes. This

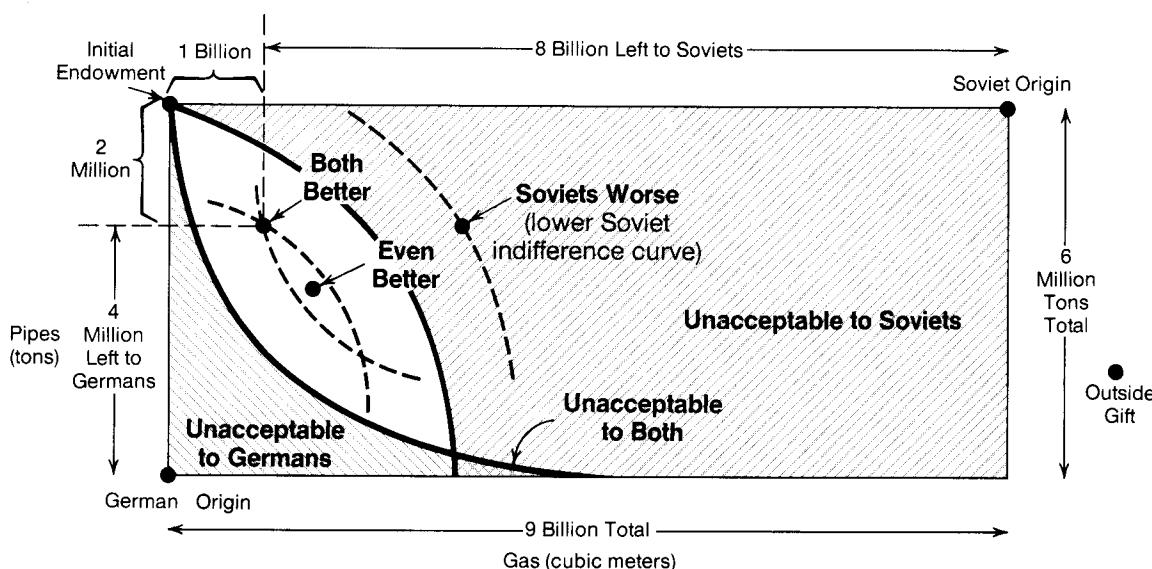


is what is meant by an exchange. Furthermore, there is a fixed total amount of gas to be distributed among the two and a fixed total amount of pipes. Consider Figure 5.4. If one tipped the Soviet pipes-gas axis and its indifference curves upside down and placed it on the German axis, forming a rectangle with the German endowment of pipes as its height and the Soviet endowment of gas as its length, any point within the rectangle would represent a distribution of gas and pipes after a deal between the Germans and the Soviets.

Any point outside the rectangle represents a deal unattainable with the initial endowments. In the Outside Gift off to the right of the box an amount of gas ends up heating German homes in excess of what the Soviets have to offer. The acceptable points—the deals that increase the utility of both parties—are within the lens-shaped, unshaded area. For example, an exchange of 2 million tons of German pipes for 1 billion cubic meters of gas leaves the two at the Both Better point. Notice the little dashed indifference curves through the point: both are better than the utilities of the initial, pre-exchange distribution of pipes and gas. That is, after the exchange both are on higher indifference curves, one curve looked at from the German, the other from the Soviet origin. By contrast, the distribution represented by the point Soviets Worse makes the

Figure 5.4
The Edgeworth Box Represents Exchange

A deal between the Germans and the Soviets must take both sides to higher indifference curves than their initial ones if the deal is to be mutually acceptable. The only points that are acceptable to both sides are those inside the lens formed by the indifference curves passing through Initial Endowment. At the point Both Better the Soviets value the marginal quantity of gas less than do the Germans; the same holds for the Germans and pipes. A lens of mutually acceptable points remains, so further trade is possible, to Even Better, for example. Trade is finished when there are no more lenses beyond the point arrived at.



Soviets worse off, for it is on a lower Soviet indifference curve than they started on before exchange. The Soviets would balk at such a deal, and it would not come off.

This peculiar diagram represents the notion that exchange must be mutually beneficial if it is to take place voluntarily, and voluntary exchanges are mutually beneficial. Deal making stops when the deal made leaves no lens-shaped area of further benefit to both parties. The Germans and Soviets might conclude a deal to move from the point Initial Endowment to Both Better. At Both Better, however, there is *still* available a lens-shaped area of further benefit, and both parties will want to move to a point such as Even Better inside the lens; that is, they will want to exchange more. They will stop only when no lens exists or when their indifference curves are tangent or when their marginal valuations of the goods are equal or when no further dealing is mutually beneficial: four ways of saying the same thing.

There Is No One Best Point Consider the following:

T or F: For a given lens of advantageous exchange between the Germans and the Soviets, there are not one but many exchanges that will leave them unwilling to exchange further.

A: Look at the lens in Figure 5.5. Any point of tangency between a Soviet and a German indifference curve will leave no lens of further exchange beneficial to both. But there are many such points (in fact, infinitely many, along the heavy line), four of which are shown in the figure. So, true.

If they move to any of these points, they will not both wish to make further deals. The collection of all such points is known as the *contract curve*, the notion being that the Germans and Soviets would both be willing to enter a contract to move from a point off the curve to a point on it.

Moving to the contract curve, then, makes both parties better off. But note the following.

T or F: Once the Germans and Soviets have taken up a position on the contract curve, the Soviets cannot be made better off without making the Germans worse off.

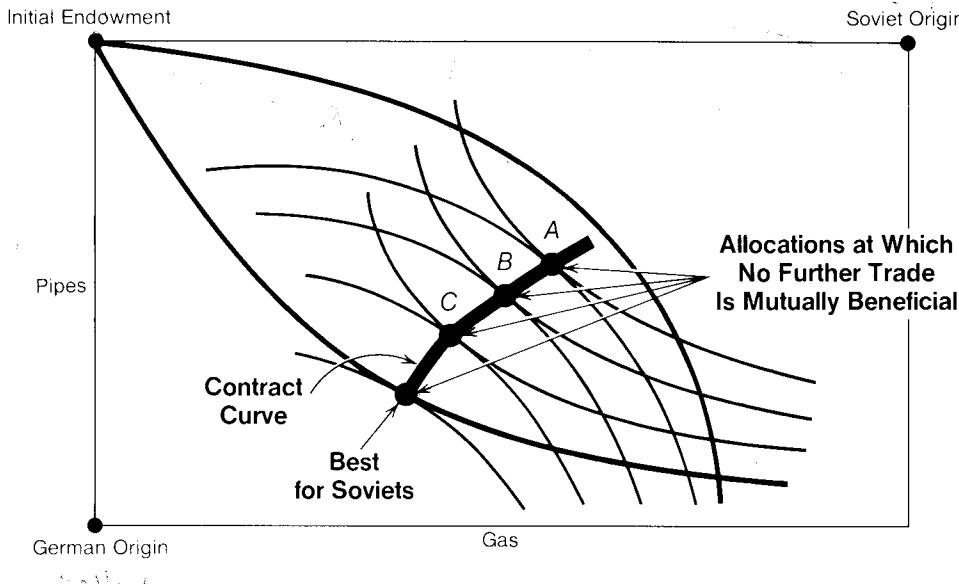
A: Look again at the diagram. If they arrive at point *A*, the Soviets would like to move to point *B* or, better still, to point *C*. Moves along the contract curve in this direction—away from the Soviet origin—obviously make the Soviets better off. But they make the Germans

worse off, that is, drive them to lower indifference curves than at *A*. Such a move reduces the German holdings of both gas and pipes, which is clearly disadvantageous to the Germans and advantageous to the Soviets. Consequently, the Germans will not agree voluntarily to move from *A* to *B* or *C*. The Soviets might try to force the Germans to do so by trying to bully them with words or bombs. But force aside, the move will not occur.

Any point on the contract curve, in other words, is a point of rest in exchange. To put it in economic jargon, any point on the contract curve is *Pareto optimal* or *efficient*. A deal is efficient if it leaves the Soviets unable to make themselves better off without making the Germans worse off. At the same time, the Germans cannot make themselves better off without making the Soviets worse off. The

Figure 5.5
The Contract Curve Is All the Points of Tangency of Indifference Curves

Points on the Contract Curve are “efficient” in the sense that at these points neither party can be made better off without making the other party worse off. Because German and Soviet indifference curves are tangent to each other at point C, a move in any direction, such as to the southeast, makes one party (in this case the Soviets) worse off. Any point on the contract curve can be reached from a point off the contract curve by a deal that is beneficial to both sides. The point actually reached by the Germans and the Soviets will depend on their bargaining abilities.



idea of efficiency is the main idea in economic thinking about society's happiness, and it will later come up again and again.

Where Do Two People Bargaining End Up?

For the moment, however, the concern is not with how people achieve happiness but with how they behave. Consider what has been said so far about their behavior. If the Soviets and Germans conclude a bargain free from force, it is clear that they have moved to a point inside a lens-shaped area. If they are skillful bargainers, furthermore, they will conclude a bargain somewhere on the contract curve, because any bargain off it leaves a further bargain that makes both better off. This much economists suppose they know about bargaining behavior.

Economists do not know, however, *where* on the contract curve skillful bargaining between two parties will lead. That is, they do not know how many pipes the Germans will give up in exchange for gas. Exceptionally subtle bargaining on the part of the Soviets, with skillful use, for example, of threats to break off bargaining entirely, might lead to a point such as the point Best for Soviets in the diagram. Here the Soviet welfare is as high as it can be subject to the constraint that the Germans are no worse off than when they started

(if they were worse off, one can assume that they would not accept the deal).

But economists have not discovered very much about what constitutes exceptionally subtle bargaining.⁵ There are profound reasons why this is so. Suppose that some bargaining technique were known to be useful when employed by the Soviets—such as threatening to abandon bargaining altogether in a week if the Germans do not accede by then to the terms demanded. The Germans would come to understand this. The Germans would use it *themselves*. Its usefulness to the Soviets, therefore, would vanish. In general, any knowledge that the analyst of the situation acquires can be expected to be acquired by the participants. They will alter their strategies in view of the knowledge, making the knowledge obsolete. The Soviet bargainers make a “last” offer. The German bargainers know that the offer is insincere (that there are quotation marks around “last”) and ignore it, making their own “last” offer. But the Soviets know that the Germans know that the Soviets’ “last” offer is insincere and prepare a “real” last offer. But the Germans know that the Soviets know that the Germans know that the Soviets’ “last” offer is to be replaced by their “real” last offer, itself insincere. And so forth.

The conclusion, then, is that bargaining between two people can end up anywhere on the contract curve—or even, if the bargainers are clumsy, off the contract curve. Economists cannot predict exactly how bargainers will behave, which is not surprising. Look at common experience. Within limits, the amount you will pay Dario Comi for a seconhand Chevy depends on your narrative skill in telling him about your poverty or your distaste for Chevrolets or the superior offer that George Dunsmore has made to you. Within limits, the frequency with which you can bluff the other players in a poker game with a busted straight depends on how much the others know about your personality, your history as a poker player, and your mood at the moment. Within limits, the success with which a congresswoman can change a military procurement bill to make her district better off depends in part on how well she can conceal her intention to vote for the bill anyway. Poor-mouthing, bluffing, and empty threats are not very well understood.

Summary

What would seem to be the simplest case of exchange—bargaining between two people in isolation—is in fact the most complex. The Edgeworth box brings indifference curves to the task of illustrating its complexity. For an exchange to take place, clearly, the exchange must make both parties better off. That is, there must exist in the Edgeworth box a lens-shaped area of mutual advantage. Or, to put it another way, at the pre-exchange allocation of money and cars or pipes and gas, the supplier’s marginal valuation of the thing that she is supplying must be less than the demander’s. An exchange that puts the bargainers on the contract curve leaves them with tangent indifference curves by definition, that is, it eliminates the lens shape. Such a position is called “efficient.” Along the contract curve one bargainer cannot be made better off without making the other worse off: No further exchanges are possible that are advantageous to both. But where the bargainers will end up on the contract curve, or even

⁵ They do have something to say about it. See, for example, Thomas C. Schelling, *The Strategy of Conflict* (Cambridge, Mass.: Harvard University Press, 1960), and other works in the “theory of games,” for which see Chapter 21, “Competition Among the Few.”

whether they will end up on it at all, is indeterminate. Within limits, the trading of cars or votes exceeds the predictive power of economics. The next section, however, will show that more bargainers narrow the limits.

EXERCISES FOR SECTION 5.2

1. "Landlords in Iowa City exploit the students who rent rooms from them at high rents." Make a hostile and relevant comment in light of the Pynchon case.
2. A student unlikely to repay a loan is like a lemon car. Do such people spoil the deal for honest folk?
3. In the definition of efficiency, does it matter whether "Soviets" and "Germans" were reversed in: "A deal is efficient if it leaves the Soviets unable to make themselves better off without making the Germans worse off?"

PROBLEMS FOR SECTION 5.2

True or False

1. If the Soviets put a very high marginal valuation on the last cubic meter of the 9 billion cubic meters of gas they had initially, exchange might not take place.
2. If the Germans and Soviets each had equal initial endowments of gas and pipes, they would not trade.
3. The greater the difference in two traders' initial endowments, the larger the gains from trade.
4. A contract curve represents the greatest happiness for the greatest number.
5. A law prohibiting people from buying medical treatment from quacks (as defined by the American Medical Association) keeps patients and quacks off the contract curve.
6. A point outside an Edgeworth box is a gift from the rest of the world.

5.3 Trade Among Many People: Behavior

What to Read For

Why don't you bargain and haggle over most deals? How much influence over the going market price of bread do you have when you purchase bread? What is a *price taker*? What is the *one man-one toilet principle*? What was the definition of the *offer curve* given in an earlier chapter? How is it used here to show the *equilibrium* achieved in a market of many buyers and sellers? Is a price that is less than the equilibrium price sustainable? Who is to blame for the equilibrium?

Why You Do Not Bargain Every Day

Fortunately for someone trying to understand and predict it, most economic behavior is not bargaining between two people. The reason is simple: There are many people. What prevents you from haggling with the storekeeper about the price at which he will sell you a quart of milk? The answer is that both you and the storekeeper can go to someone else to make the exchange. If you do not like the price he offers, you can go to another store; if he does not like the price you offer, he can wait for another customer to come in. Only

when the alternative stores or customers are known to offer worse deals or are expensive to contact can bargaining occur. A bunch of flowers that is sold during the week at the North End market at a fixed price becomes at 11:55 Saturday night (the market is closed on Sunday) an object of bargaining between the last customer and the last flower vendor. During the week, competition for the flowers and for the money determines their price. This competition takes place among many buyers and sellers of the same item. In such circumstances each seller or buyer takes the price obtainable as given, for with many sellers and buyers no one of them can alter the market price by threatening to refuse the bargain. If you individually refrain from buying a loaf of bread, the price does not change perceptibly, because another of the millions of consumers of bread steps into your place and buys it. We are, in short, back to one consumer facing given income and prices.

One Man, One Toilet

This is one of the most important bits of reasoning in economics. It bears restatement. One buyer of bread among a hundred or a million cannot perceptibly affect the terms on which he or she buys bread. The buyer is, as the jargon has it, a *price taker*. The price may vary from time to time, but not because of variations in the amounts the buyer individually buys. It would be irrational for the buyer to think otherwise. If all people flushed their toilets at once, the city water mains would burst. But a man bearing a grudge against the city would be insane to chortle in gleeful anticipation of a water main break as he flushed his toilet because he is only one of many owners of toilets. *One man, one toilet*: That is the limit of his influence over the water supply, as is his one vote over an election or his one decision not to buy a loaf of bread over the price of bread.

Therefore, all the apparatus constructed earlier of a consumer facing prices and income "given" to him (that is, over which the consumer individually has no influence) applies to exchange in populous markets. And it is intuitively clear why the analysis of such a situation of price taking will lead to definite conclusions while the analysis of two-person bargaining over the price leads nowhere. Two-person bargaining is complex because my offer affects thine which affects mine which affects thine, and so on to mental exhaustion. Price taking cuts the story short at the first words: I buy a pound of bread at the going market price without reflecting on how my act will change the price at which the storekeeper will offer the bread, because it will not in fact change his price.

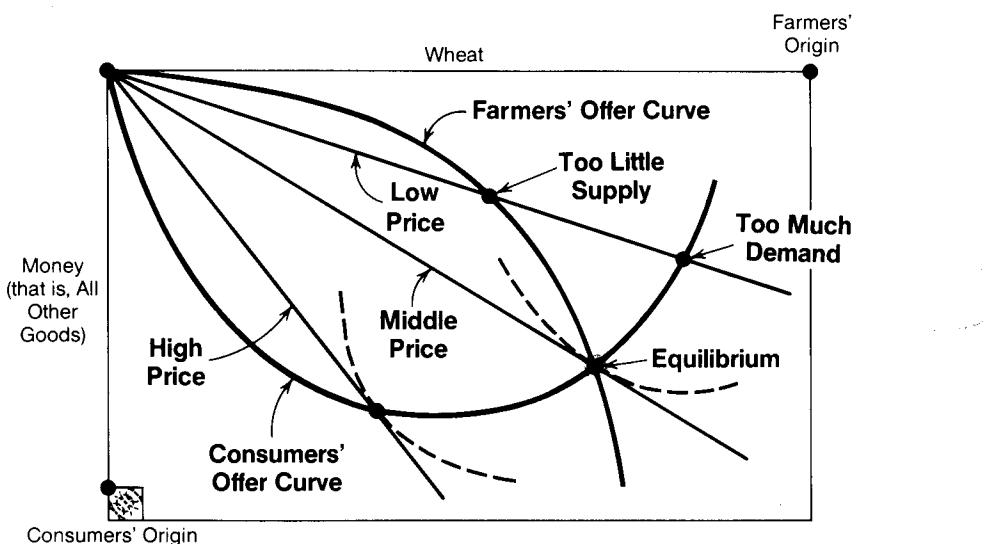
How the Market Achieves Equilibrium

A radical simplification in the theory of exchange is permitted by price taking. It eliminates bargaining. One can still, however, look at the behavior of the market in an Edgeworth box, and a great deal can now be gotten out of it. Consider Figure 5.6. Each farmer has an endowment of wheat to sell for other goods and each consumer an endowment of money to sell for wheat. Thousands of farmers and consumers meet in Chicago to exchange their little bundles of wheat and money.

The tiny box in the lower left-hand corner is the Edgeworth box for an exchange between the farmer Susan Howson and the wheat eater Steven Webb. The whole box is the Edgeworth box for exchanges between thousands of Howsons and Webbs, that is, for the whole market. Each Webb is unable to influence

Figure 5.6
Price Taking Implies Market Offer Curves

The market offer curves reveal that the Middle Price line is tangent to the indifference curves of both farmers and consumers at point Equilibrium, so farmers wish to sell the same quantity that consumers wish to buy. At Low Price, consumers demand a larger quantity of wheat than farmers offer to sell. It therefore cannot be an equilibrium.



the money price of wheat he faces in the market and therefore takes the price as given. When the price varies, he varies his consumption of wheat, as do all the other Webbs. The wheat-consuming Webbs, in other words, move along the heavy Consumers' Offer Curve just as does a single consumer in the analysis of demand in the last chapter. Two of the Webbs's indifference curves tangent to the High Price and to the Middle Price are shown as dashed lines to jar your memory of how such offer curves are constructed. The Farmers' Offer Curve is similarly constructed from the farmers' indifference curves. Of course, since the Edgeworth box flips the farmers' indifference curves, their offer curve has the opposite curvature. At the crossing point of the two curves, Equilibrium, the Middle Price induces farmers to sell exactly as much wheat as consumers are induced to demand at that price. That is, Equilibrium is the only point on the two curves for which the offer made by farmers as a group is consistent with the offer made by consumers as a group. At any price other than Middle Price, the offers will be inconsistent.

Low Price, for example, induces farmers to supply wheat only out to point Too Little Supply, whereas it induces consumers to demand at point Too Much Demand. Quite naturally, in other words, a low price of wheat causes suppliers to supply less than demanders demand. Equally naturally, the price will tend to rise. With demanders banging on his door and telephoning him day and night in a search for more wheat at the Low Price, it will be in the self-interest

of each supplier of wheat to withhold his wheat until the price rises; likewise, it will be in the self-interest of each unsatisfied demander to raise the amount of money he offers, enticing wheat supplies away from other demanders. The self-interest of both, then, pushes the price up, and continues to do so until Middle Price—the price of Equilibrium—is achieved.⁶ And a similar story applies to High Price: It is not sustainable, and it tends to fall to Middle Price. At Middle Price demanders do not wish to buy more wheat than is supplied by suppliers; suppliers do not wish to sell more wheat than is demanded by demanders. The market price is determined; the self-interest of thousands of price-taking suppliers and demanders has led to the glorious end of the story, Equilibrium.

The Equilibrium Is No One's Plot

The determination of market price by the self-interest of large numbers of price-taking suppliers and demanders is routinely misunderstood by the man in the street. A failure of the Brazilian coffee crop and the consequent rise in the price of coffee leaves people muttering about a conspiracy by the Brazilian government. A rise in the price of natural gas or meat leads their representatives in Congress to open hearings into "price gouging" by big oil companies or big meat wholesalers. The implicit model of price determination used in such analysis of "price gouging" is a bargaining model, with one evil mind determining the quantity supplied and the price, altering the price at its pleasure. Accurate as the model is for some markets (see Chapter 17), it is inaccurate for most to which it is applied. Markets of price takers have many participants none of whom has a noticeable influence on the price faced. No participant intends the price to be what it is. However high the price of meat, suppliers would like it to be higher; however low the price, demanders would like it to be lower. The equilibrium price is the unintended consequence of the self-interested exchange. Suppliers charge as high a price as they can, what the market will bear; if you as a demander are irritated by this lack of charity, feel free to call it "price gouging" or charging an "exorbitant" price. But keep in mind that you as a consumer offer as low a price as you can, what the market will bear: in the supplier's eyes you are "exploiting" the supplier or paying an "unfair

⁶ Another assumption has slipped into the argument, the assumption known as *recontracting*. The notion is that farmers and consumers make tentative contracts to exchange at Low Price, but recontract when it is discovered that Low Price does not produce equilibrium. The contracts become binding and the exchange is actually made only when the price has no tendency to change (that is, at Middle Price). The purpose of this artificial assumption (which does apply, though, to certain well-organized commodity markets) is to eliminate exchanges at nonequilibrium prices. The point is that were these to take place the endowments of the parties would be in effect altered. They would move first to some intermediate position short of equilibrium, such as Too Little Supply, and then trade from that position. The diagram would be pointlessly complicated, because then offer curves would have to be redrawn out of Too Little Supply, and redrawn again out of the next intermediate position, and so forth. In applications the defense of the simplifying assumption of recontracting can run along various lines. Recontracting may literally exist (as in the well-organized markets just mentioned). Equivalently, the goods may be auctioned off (an auction ends when the price has no tendency to change). For most markets the defense has to be that the gains and losses of trading at nonequilibrium prices average out. Some wheat will be exchanged for money at a higher than equilibrium price, some at a lower than equilibrium price. The defense is not always persuasive. If for some reason much wheat, for example, is bought at a very high price (the market falsely anticipating a harvest failure, say), wheat consumers may be so impoverished by the transaction that the equilibrium price ultimately achieved is very different from what it would have been with recontracting.

price." These words imply intent and therefore moral responsibility. To repeat, however, no single price taker is responsible for the going price.

Summary One person's decision to buy or not to buy a bushel of wheat, a house, or a suit of clothes has a trivial influence on the going price of wheat, housing, or clothing. Having no influence, the person behaves as a "price taker," taking the market price as given and making the best choice possible in view of it. In other words, the person is the isolated consumer of earlier chapters facing a given budget line. Adding up the offer curves of each isolated demander and supplier results in market offer curves (demand and supply curves). The intersection of the market offer curves determines the equilibrium price and quantity toward which self-interest will drive the market. This process is blind: No participant conspires to make it happen, but it does.

EXERCISES FOR SECTION 5.3

1. You are fortunate to meet a man in the middle of the Sahara, three days camel ride from anyone else, and propose to trade some of your water for some of his food. Is either of you a price taker?
2. "Collective bargaining" is bargaining by a labor union about wages. For example, Iowa Beef Company bargains about meat-packing wages with a union. But if the union is a small part of the labor force in Iowa (it is) and the company takes a small part of the labor force (it does), is there really anything to bargain about?
3. The United States regularly produces more than 60% of the world's soybeans. There are hundreds of thousands of American farmers who grow soybeans. *True or false:* The United States as a whole has the power to bargain about soybean prices (perhaps the government of the United States in trade negotiations with Japan), but single farmers are price takers.
4. *True or false:* The offer curves are the same as indifference curves.
5. "Nader Charges Gouging by Coffee Retailers." If coffee retailing (by the cup or the pound) has a lot of alternative suppliers, is Nader right?

PROBLEMS FOR SECTION 5.3

- 1. The proposition that the little man is rational to be a price taker applies to more than markets. It applies to politics. Suppose that Claudia Goldin and Peter Temin are candidates for the Senate in what is expected to be one of the closest races in American history. A million people are expected to vote, yet a swing of only a thousand votes one way or another is expected to make the majority. If each arrangement of the thousand votes is equally probable, it follows that the probability of an exact tie—the only circumstance in which an individual vote counts—is expected to be 1 chance in 2000. You value a Goldin victory at \$100. That is, if some political magician could guarantee Ms. Goldin's victory for a fee (he cannot, of course), you would be willing to pay him \$100. Set aside any pleasure you get from fulfilling your patriotic duty. *True or false:* If it costs you more than 5 cents in time and trouble to go to the polls and vote, it would be irrational for you to do so.
- 2. Under common law, the "enclosure" of an English village before the eighteenth century required the consent of every one of the 40 or 50 owners of rights in the village ("enclosure" involved the gathering of scattered plots of land into consolidated holdings). During the eighteenth century the English Parliament developed another method of enclosure, which

could take the place of the common law method and which required only a vote by a majority of the 40 or 50 owners to enclose, not unanimous consent.

a. Under the common law method, what would happen if one owner refused to enter the bargain? What could the recalcitrant owner demand of those who wanted the enclosure to go forward? Who else would have an incentive to make this demand?

b. Under the new parliamentary method, what happens to the power of one owner to block an enclosure? Why? Would you expect enclosure to become more common after the new method was introduced?

3. When the Bally Corporation decided to build a casino in Atlantic City it announced that it would pay \$100,000 to each owner of a house lot on the proposed site. Even though the assessed value of the houses was less than \$15,000, some homeowners still held out for more. Why?

5.4 Trade Among Many People: Happiness

What to Read For Is the equilibrium also the best the society can do? In what sense of "best"? What is Adam Smith's Theorem? What is the relationship between the equilibrium point of intersection of the offer curves, and the point of intersection of the demand and supply curves corresponding to the offer curves? Do people have to be unselfish to achieve efficiency? Is efficiency good? Should we always let the market work freely? What is the economist's way of fixing bad results of markets? How does this compare to the noneconomist's way?

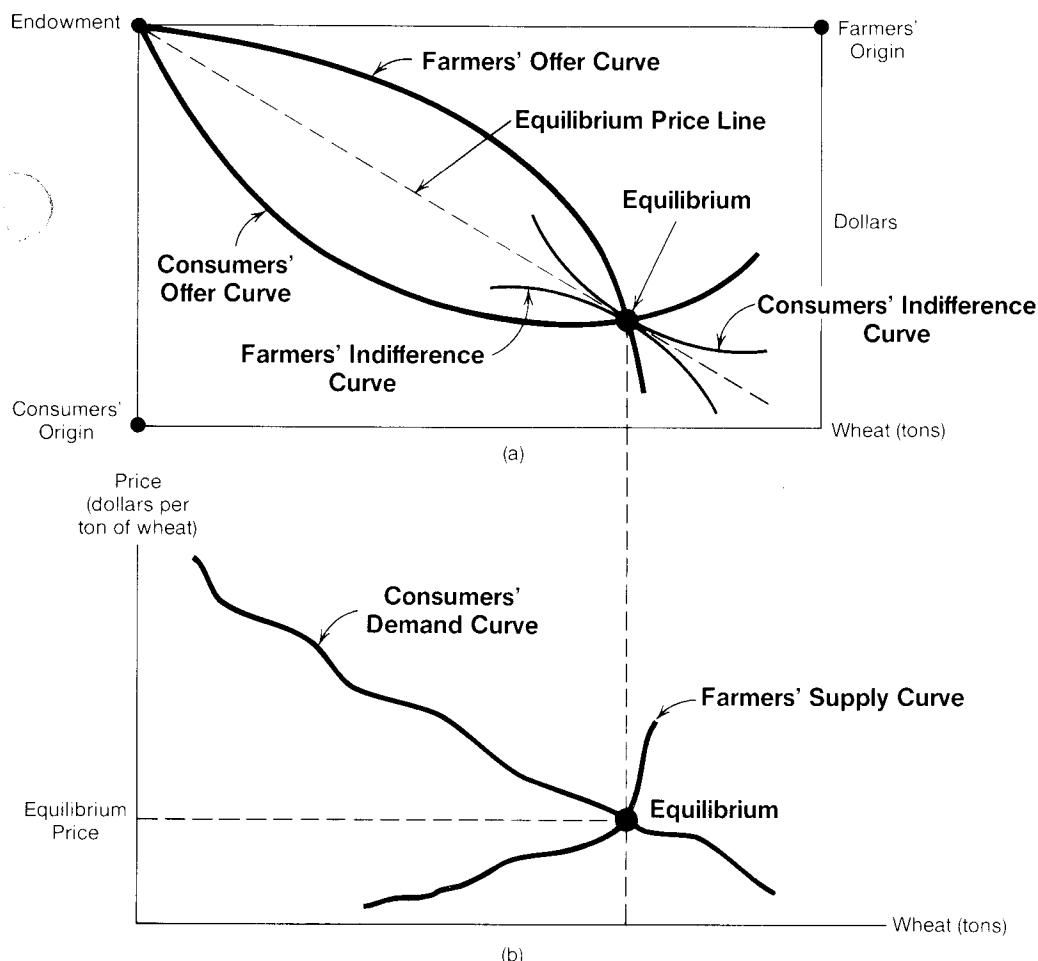
Adam Smith's Theorem It is now possible to prove a central and celebrated theorem in economics: that the exchange brought about by the pursuit of self-interest in a market of price takers is *efficient*. In other words, the equilibrium of supply and demand leads suppliers and demanders into a position in which one cannot be made better off without making another worse off. The proof wraps the arguments of this and the preceding chapter into one lovely diagram (Figure 5.7). Well, at least it looks lovely to an economist. The top panel shows the offer curve of farmers and consumers of wheat. The bottom panel shows the demand and supply curves corresponding to the offer curves.

It has just been shown why the point Equilibrium (in either panel) is the equilibrium exchange and why therefore the slope of the dashed Equilibrium Price Line (in the top panel) is the equilibrium price. Look closely at Equilibrium in the top, Edgeworth box panel. Because of the way in which offer curves are constructed, a farmer's indifference curve (drawn in lightly) touches the Equilibrium Price Line at Equilibrium; likewise, a consumer's indifference curve (also drawn in lightly) touches the Equilibrium Price Line there. That is, at Equilibrium the two indifference curves, touching the same price line at the same point, touch each other. But exchanges at points of tangency between two indifference curves are on the contract curve (not drawn). That is, Equilibrium is on the contract curve. At the point Equilibrium the farmers cannot be made better off without making the consumers worse off, and vice versa. In other words, there is no reallocation that can make everyone better off. The equilibrium of supply

Figure 5.7

Exchange Among Price Takers Is Efficient: The Offer Curves Intersect at a Point of Mutual Tangency (a) So That Supply Equals Demand (b)

Along the Equilibrium Price Line, farmers offer a certain quantity of wheat in return for a certain amount of money. Consumers offer an identical amount of money in return for an identical amount of wheat. Furthermore, at Equilibrium no person is made better off without making someone else worse off: The equilibrium is efficient.



and demand, arrived at by the struggle for advantage in the marketplace, is efficient. *Q.E.D.*

The theorem, proven in this form at the end of the last century and reproven repeatedly in more subtle forms down to the present, was made a leading theme in economics in 1776 by Adam Smith. *Smith's Theorem* is a paradox of selfishness. The least admirable of human motives leads in a market of small sellers

and buyers to an admirable result, the full exploitation of mutually advantageous exchanges. That is, selfishness leads onto the contract curve, and we have seen that a move from a point off the curve to certain of the points on it is desirable. Supply and demand selects one of these desirable points.

That Supply Equals Demand Is Efficient Does Not Imply That It Is Good

Recall, however, that the contract curve consists of infinitely many such points. There is nothing morally sacred about the particular point selected by supply and demand. The moral choice between points on the contract curve is a choice between one person's happiness and another's. This moral choice is much harder than the choice between making both happier by moving to the curve or leaving both with the little happiness they have by staying off the curve.

Q: By altering within the initial lens of mutually advantageous exchange the endowments of wheat and money in the hands of farmers and consumers, a market of price takers can arrive at any point on the portion of the contract curve inside the lens. Show it.

A: In Figure 5.8 the offer curves running out from the initial endowment meet at the point Unimpeded Equilibrium. By altering the endowment to Ceiling Supply, one

generates another pair of offer curves (dashed), meeting at another point, called here Morally Better. Both points are the result of price-taking markets, and both are on the contract curve. But at Morally Better farmers are worse off and consumers better off than at Unimpeded Equilibrium. It is pretty clear from the diagram that an endowment can be found that leads to any given point on the contract curve.

In other words, changes in the distribution of goods cause changes in the offer curves (or supply and demand curves) and changes in the final distribution of goods. Unless the present distribution of goods among people is thought to be morally best, there is no case that economics can make without a moral judgment that leaving the market to work unimpeded is best. This is what it means to say that the point on the contract curve selected by supply and demand is not morally sacred.

The Economist's Presumption in Favor of Letting the Market Work

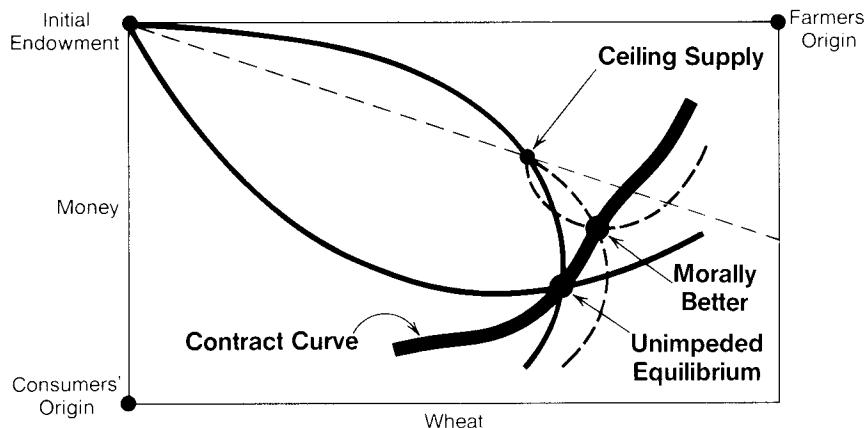
Yet Smith's Theorem began a long love affair between economists and the market. Economists have a reason for their love of the market. Suppose it were determined that the unimpeded equilibrium in the market for wheat was socially undesirable, that in a better world the wheat consumers *should* be made better off and the farmers worse off. There are two ways of achieving this, the noneconomist's and the economist's way.

The noneconomist's way is to interfere with the market directly, setting, for example, a price ceiling on wheat that will benefit consumers. The economist's way is to change the initial endowment in favor of the consumer by taxing or subsidizing income. Afterward the economist would let the market work unimpeded. In Figure 5.8 the dashed straight line representing the price ceiling leads farmers to supply wheat out to Ceiling Supply, which may indeed put consumers on a higher indifference curve than Unimpeded Equilibrium (the lower amount of wheat consumers get may be more than outweighed by the smaller amount of money they have to give up to get it).

Economists argue, however, that Ceiling Supply can be improved upon by making it, for example, the initial endowment and then letting both parties trade out to the point Morally Better on the contract curve. At Morally Better both will be better off than at Ceiling Supply. The economist's way makes both parties

Figure 5.8
Changes in Initial Endowments Change the Offer Curves. Redistribution and Free Markets Can Improve upon Unimpeded Markets

Any point of the contract curve can be reached by a series of exchanges that exhausts the possibilities for mutually beneficial exchange. A price ceiling gets consumers to the Morally Better point but leaves society at Ceiling Supply, which is inefficient. A redistribution that makes Ceiling Supply the endowment point guarantees Morally Better to consumers and permits farmers and consumers to improve their happiness further by free trade to the Contract Curve.



better off. That is, the society should reshuffle the endowments of wheat and money instead of interfering with the market directly. Any distribution of happiness achieved by interfering with the wheat market can be made better by interfering instead with the distribution of income and then letting the market work.

This is the reasoning behind, for example, what has been called the "negative income tax": Replace the present maze of subsidies to housing, food, and medical care for the poor with money payments to the poor. As James Tobin, one of the inventors of the negative income tax, put it:

While concerned laymen who observe people with shabby housing or too little to eat instinctively want to provide them with decent housing and adequate food, economists instinctively want to provide them with more cash income. Then they can buy the housing and food if they want to, and if they choose not to, the presumption is that they have a better use for the money.⁷

The best way in which to enrich the poor, in other words, is to enrich them, not to tinker with the markets in which they operate. If the employees and stockholders of Zenith Television are considered to be appropriate objects of public charity, the charity should take the form of money gifts to them, not

⁷ "On Limiting the Domain of Inequality," *Journal of Law and Economics* 13 (October 1970): 263–277, reprinted in E. S. Phelps, ed., *Economic Justice* (Harmondsworth, England: Penguin Education, 1973), p. 449.

quotas to protect them (at the expense of consumers) against Japanese competition. If farmers are considered to be an especially desirable group for the nation to subsidize, the subsidy should again take the form of money gifts, not (again at the expense of consumers) price supports, officially sanctioned marketing boards, acreage allotments, and payments for nonproduction.

The argument is not conclusive. As Tobin remarks, it "rarely satisfies the intelligent egalitarian layman . . . [because he knows] that there are pragmatic limits on the redistributive use of taxation and cash transfers."⁸ To be sure, one of the "pragmatic limits" is precisely the openness with which redistribution of income puts the social question. In a democracy it is good, not bad, to put social questions openly. There would undoubtedly be more opposition to a subsidy to Zenith Television if it appeared as a line on the income tax form ("Zenith Subsidy: Add \$3 to calculated tax") than if it appeared, as it does, as a small item in the financial section of the newspaper ("Zenith Hits Unfair Sony Practices, Calls for Quota"). But the main point is that the taxes required to get the money for the subsidy are themselves interferences in a market, the market for the taxed good (income or cigarettes or whatever). The economist's way of helping the poor does not in fact avoid all interference in markets.

This is one among many sophisticated objections to the economist's love for the market that will be discussed in more detail in later chapters. At any rate the flaws in the unsophisticated objections are by now clear. It was noted earlier that the ignorance of people is not an objection if they realize their ignorance or if businesses arise to teach them. Nor is the poverty of people a decisive objection to permitting the market to operate unimpeded: Their poverty can sometimes be alleviated directly, leaving the market to achieve the additional benefits of efficiency. Smith's Theorem, in short, is not easily dismissed.

Summary

Smith's Theorem asserts that a selfish price taker competing with others "intends only his own gain, and he is in this, as in many other cases, led by an invisible hand to promote an end which was no part of his intention,"⁹ namely, a better price for his customers and an efficient exchange. Capitalism is altruistic, not in intent but in result. Smith's Theorem, however, does not guarantee social bliss. The point on the contract curve that the intersection of supply and demand selects is not necessarily the best point. Yet the economist can argue, if he argues with care, that questions of efficiency—whether society should be on the contract curve—can be treated separately from questions of equity—where on the contract curve society should be.

EXERCISES FOR SECTION 5.4

1. *True or false:* In the Equilibrium of Figure 5.7 no one can be made better off.
2. *True or false:* The point on the contract curve chosen by market competition is the best.
3. What change of income distribution would an economist recommend to replace the following programs of the American government?
 - a. Quotas on imports of Japanese steel. (*Sample answer:* Abolish the quota and send

⁸ Tobin, in Phelps, ed., p. 449.

⁹ Adam Smith, *The Wealth of Nations*, first published in 1776 (Chicago: University of Chicago Press, 1976), Book IV, Chapter II, p. 477.

government checks directly to steelworkers and to owners of the steel companies. This would have the same effect on the distribution of income as the quota, but would allow exchange out to the contract curve.)

- b. Restrictions on what land can be used to grow tobacco, which enriches the owners of the land that can.
- c. Rules that certain American exports can be shipped only in ships of American registry.
- d. Taxes on imports of foreign cars.

PROBLEMS FOR SECTION 5.4

1. The United States imposes an effective quota on its imports of Italian shoes. *True or false:* By comparison with the free trade point, the quota makes both Americans and Italians worse off. (*Hint:* Think about it in terms of Figure 5.8.)

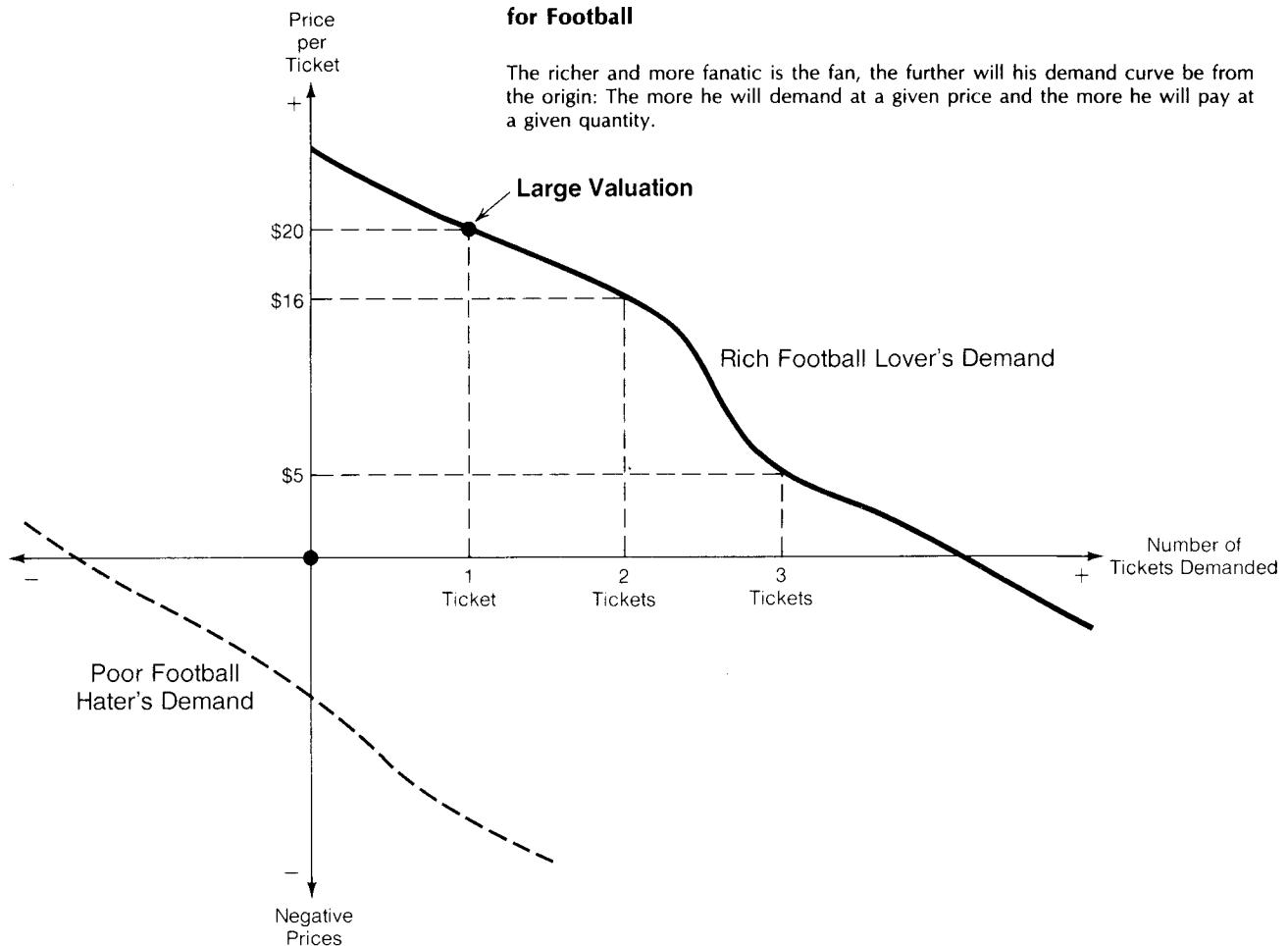
- o 2. The guns for today's war can be paid for by two methods. On the one hand, the government can force its citizens to give up resources today by taxing them. On the other, it can entice them to give up resources today by offering them an interest rate if they will loan the government the resources, to be paid off tomorrow. To pay off the loan tomorrow it will have to tax its citizens tomorrow. *True or false:* In a sense the method of borrowing to pay for the war transfers the burden from today's generation to tomorrow's. (*Hint:* "Entice" is a voluntary exchange; "force" is a push. How do you value each?)

Using Market Supply and Demand

6.1 The Uses of Equilibrium

What to Read For	What determines the position of someone's demand curve for football tickets? Is a demander always a demander, regardless of how the price or his initial number of tickets changes? Why is allocation of tickets necessary? Are tickets scarce? What are some alternative ways of allocating tickets? Is bribery always bad? Is there a victim in an exchange? Is scalping of tickets bad? Why do economists favor making tickets resellable? When the market is allowed to work, who ultimately gets the tickets?
The Problem of Allocation	<p>Thinking about exchange can be done simply with curves of supply and demand. It is startling how far this simple idea can travel, as the next couple of chapters will show. The starting point for the journey is the Law of Demand for tickets to the Iowa stands in the Iowa-Minnesota football game. Everyone has a demand curve for tickets. If the person is the Football Hater in Figure 6.1 his demand curve is low—or even as drawn below the origin, which says that he would have to be paid to accept the seat. If he is a Football Lover, however, his demand curve is far above and to the right of the origin. So too if he is rich, since a large income moves out the demand for normal goods.</p> <p>At a given price a Rich Football Lover will demand more tickets than a Poor Hater. Or, what amounts to the same thing, at a given quantity of tickets—say 1 ticket—the Rich Football Lover will be prepared to pay the Large Valuation (namely, \$20). Think about the common sense: Clearly it makes sense that a lover will want a lot of what he loves and will pay much for whatever he can get.</p> <p>If he already has a lot of tickets, however, he'll not want more. If the Football Lover in the diagram had 100 tickets given to him, even he would sell off some of them. Although he is a football lover, he would become a net <i>supplier</i> of tickets to other people. Furthermore, if other people are willing to pay a</p>

Figure 6.1
The Position of the Ticket Demand Curve Depends on One's Love for Football



great deal—that is, if other people have an even greater love—the Football Lover will become a supplier even with few tickets, or one. For instance:

T or F: If tickets are selling for \$100 each, the Football Lover as drawn in the diagram will sell his ticket even if he is given one free.

A: Look at his demand curve at 1 ticket: He is willing to pay \$20 or a little more. If he keeps the ticket he

is getting \$20 worth of pleasure but sacrificing the \$100 he could get from the football fanatics willing to pay \$100. So he will give up the ticket if he can, selling it outside the stadium for \$100 on game day. True.

The reasoning comes from the downward slope of even the Football Lover's demand curve. That is, from the Law of Demand comes a Law of Supply: A

big initial endowment of tickets or a lack of love for football will make Jonathan Pincus a supplier of tickets. And at some high price he'll be a supplier even though he has few tickets and loves each one. The higher the price, the greater the supply.

All this, and more, refers to the individual demand. But football is watched in large crowds, and it is in understanding the behavior of large crowds that supply and demand shows its worth. Unhappily, it is a fact of scarcity that not everyone who has a positive demand for seats can fit into Kinnick Stadium in Iowa City to watch the game. Somehow the 70,000 seats have to be *allocated* among the hundreds of thousands of people who would gladly take them if they were free. How to do it?

Understand that the question is among the most important a society faces. True, allocating football tickets, when all is said and done, is not very important. But *everything* scarce needs to be allocated, and some of the scarce things are very important. The choice of how to allocate—whether by the competitive market or by a central plan or by chance—must be faced for housing, water, education, food, clean air, playgrounds, books, clothing, cars, roads, and every other scarce good. The football example is merely a simple one (unclouded by strong feelings about "need") which represents the whole Problem of Allocation.

The first step to solving the problem is to notice that people will not just sit there, but will try to deflect the tickets in their direction. They will, in other words, *compete* for the tickets. The forms of competition that are permitted by the society will determine who wins. For instance, if the society permitted people to use violence to compete for the tickets, the strongest would win. If the tickets were allocated by first-come-first-serve, those best able to compete with patience for standing in line would win. If the tickets were allocated by political influence, the politically influential would win. Competition is an inevitable consequence of scarcity and of the ability of people to compete in one or another way.

The way that is commonly used, of course, is the market, or in other words bribery. The Football Lover bribes whoever now has the tickets. This sounds nasty. But on reflection you will agree that bribery is a lot less nasty than most of the alternatives. If big, mean people with guns can go around shaking people down for tickets, the victims get no reward for losing their tickets. If the same folks go around with oily smiles offering bribes, the victims are no longer "victims" at all, but willing recipients of payments for services rendered. If tickets are allocated to those who make the most persuasive case that they "need" them the most, the people who give them up are again made no better off. Smart talkers win, which is one reason why literary intellectuals prefer allocation by political dialogue to the grunts of the market. If on the other hand the smart talkers must offer bribes, it will be their money not their mouths that does the talking.

T or F: Scalping of tickets—buying tickets from the ticket office wholesale and marketing them retail at the stadium gate—is a great evil.

A: So it is commonly believed. But the belief does not

have much to recommend it. The scalpers are no more objectionable than ice cream or hot dog retailers. No one forces you to accept their deals. You get the retail convenience of being able to buy on the spot on the spur of the moment.

One way to allocate the tickets is to allocate them randomly, by dropping them out of a helicopter or delivering them to every twentieth house. No economist would object to such a scheme *as long as people could scalp (resell) the tickets*. Usually they cannot. Tickets allocated to students by university rule, for instance, often cannot be sold to nonstudents, and student IDs are checked at the door. The economist views such restrictions as lunacy. What is lunatic about them is that the random allocation will give some tickets to Football Haters. The Football Lovers would gladly pay money to the Haters to get the tickets, but if resale is forbidden they cannot. In other words, the economist recommends that the tickets be allocated not by violence or lining up or status in the university but by the market. The reason he recommends it is simple. The market is voluntary. And the market, when it is allowed to work, *puts the scarce and limited tickets in the hands of those who value them the most*.

This is the key point. No matter how the tickets are allocated at first, the market will redivide the population of Iowa into two groups. This is like *any* system of allocation. One will be inside, watching the team crush Minnesota; the other will be outside. But what makes the market different and better is that every person inside will value his or her seat more than anyone outside. Otherwise, each would have made a deal with an eager outsider. If the tickets are allocated by strength or beauty, the strong or the beautiful get them and the rest are hurt. If they are allocated by the market, those willing and able to pay the most get them and the rest (if they had the tickets at first) are compensated.

Summary

As goes the allocation of football tickets, so goes the allocation of any scarce thing, from chewing gum to kidney machines. The thing has to be allocated, separating people into those who get it and those who don't, those who see the game and those who don't. The demand curve of each person expresses how much he would be willing to pay for tickets. A poor person with lots of tickets initially who hates football will on all these counts be a net supplier. A rich person with few tickets initially who loves football will be a demander. No matter how the tickets are initially allocated, if bribery—the market—is allowed to work, these suppliers and demanders will make deals. The deals put the tickets in the hands of those who value them most—not in the hands of the most strong or the most patient or the most well connected. That's how the market works. It's also why the market springs up all the time: Participants like it. And it's why economists like it, too.

EXERCISES FOR SECTION 6.1

1. Suppose that the University of Iowa gave you all 70,000 tickets to The Game. Describe your situation in terms of your demand curve and your initial allocation. What would you do with the tickets?
2. Normally, Michael Edelstein demands furniture. From time to time he buys a couch or a rug. Someone dies and leaves him a houseful of furniture. What happens?
3. Vietnam was once a net exporter of rice. Since its war and revolution it has become a net importer. Does this mean that its demand curve has moved out?
4. Name six alternative ways that people might compete for one of the 70,000 tickets.

5. Which of the six ways hurts other people? Which wastes resources in the very process of competing itself?
6. In New York City under rent control a new tenant often pays “key money” to an old tenant to induce her to move out of a cheap apartment. Is key money bad?
7. Countries such as Romania have price controls and rationing. That is, the price of meat is set below what it would sell for in an open market and the meat is allocated instead by political influence or need. There are “black markets” in which meat can be bought at prices far above the official price, coming from people who sell the meat allocated by the government to them. Black marketeers are often jailed or shot. From the point of view of the participants in the market, is there anything bad about black markets? From the point of view of government officials who wish to keep the power to allocate meat, is there anything bad?
8. Under a market system, the rich *could* get all the tickets. Do they? Why not?
9. Under a beauty system (tickets are given to beautiful people) the beautiful get all the tickets. What is the essential difference between the two methods of allocation? (*Hint:* Does a scalper ask to see your bank account? On the other hand, do you give up some of your pile of beauty when competing in a beauty contest?)
10. If there are rents fixed by law, then tenants of apartments can’t compete with money. If the tenants can’t compensate with money, will landlords be pleased to rent to tenants with children or dogs? Students? Poor people? Racial and ethnic minorities?

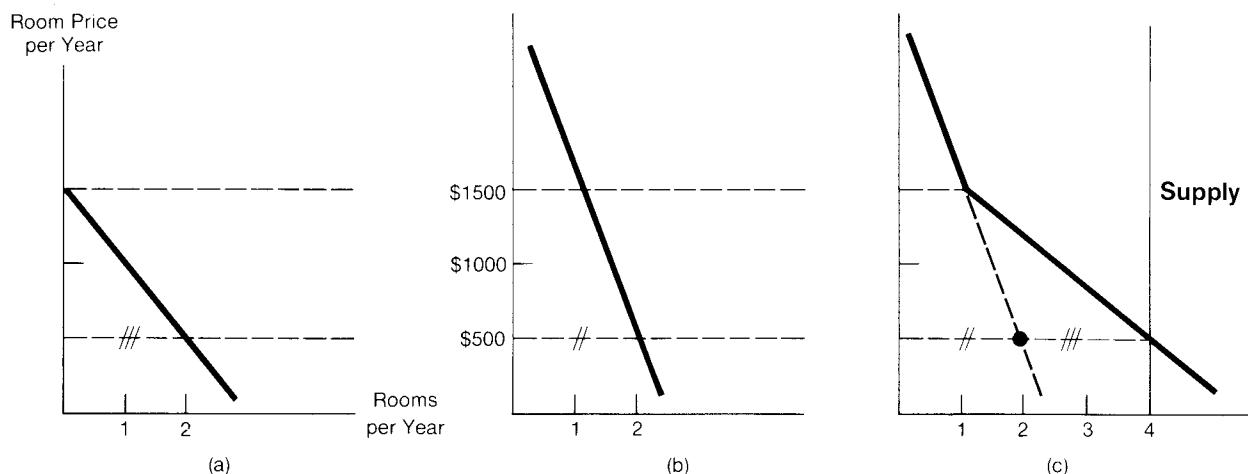
6.2 Adding Up Supply and Demand: Equilibrium

What to Read for	How do you go from a single person's demand curve to a whole market's demand curve? The market's supply curve? What is equilibrium, geometrically speaking? What is the significance of the phrase "the market must clear"? What is the relation between quantities supplied and quantities demanded at the equilibrium price? How do movements in the price and quantity depend on shifts in the supply schedule? On shifts in the demand schedule? How can one infer from price and quantity movements what has happened to the market supply and demand curves?
How to Add Up Individual Demands and Supplies	Suppose that there are many demanders of housing in Iowa City. Each demander is a price taker because there are many: Recall that a “price taker” thinks (quite sensibly) that his purchases or sales will not change the price. He takes the market price as given. At any given market price each will want to buy a certain quantity of housing according to his demand curve. At any given market price the quantity demanded by the market (that is, by all demanders) will be <i>the sum of all the individual quantities</i> . The market demand curve, in other words, is the horizontal sum—the sum at various given prices—of the demand curves of all demanders. Were there only two demanders, Ransom and Sutch, the market demand curve for housing in Iowa City would be constructed as shown in Figure 6.2. Notice the kink in the market demand at the price (\$1500) below which Ransom begins to demand housing. Notice too the slashes that identify in the style of high school geometry horizontal segments of equal length (in the case shown, at a price of \$500).

Figure 6.2

Market Demand Is the Horizontal Sum of Individual Demands: Compare Ransom's Demand (a), Sutch's Demand (b) and Demand for Both (Market) (c)

The quantity demanded by the market at any price equals the quantity demanded by Ransom plus the quantity demanded by Sutch. Thus the market demand curve equals the horizontal sum of the individual demand curves.



The same sort of diagram gives the supply curve of housing. Add up at each price the quantities supplied in Iowa City by four landlords, DeCanio, Higgs, Reid, and Wright. Mathematically the trick is to express the individual quantities demanded and supplied as functions of the price of housing and then add up the functions. The demand function would be, speaking in generalities, $D(P)_{\text{RANSOM}} + D(P)_{\text{SUTCH}}$ and the supply function $S(P)_{\text{DECANIO}} + S(P)_{\text{HIGGS}} + S(P)_{\text{REID}} + S(P)_{\text{WRIGHT}}$. To be specific, Ransom's demand curve for rooms might be $Q_R = 3.0 - 0.002P$. This equation says that Ransom will buy, for example, no rooms at a price of \$1500 per room [$3.0 - 0.002(1500) = 0$] and 3 rooms at a price of zero (do it to see if you understand). Likewise, Sutch's demand might be $Q_S = 2.5 - 0.001P$. The market demand curve would therefore be the sum of these two, $Q_R + Q_S = 3.0 - 0.002P + 2.5 - 0.001P = 5.5 - 0.003P$, which is in fact the equation for the line below the kink in the diagram.

Against this demand curve might be set a fixed supply of one room from each of the four suppliers. A fixed supply is insensitive to price; that is, the amount supplied is *not* a function of the price. The supply curve, therefore, is a vertical line at 4 rooms, as in the diagram. The *equilibrium price* is the price that makes the two demanders consume exactly 4 rooms in total:

$$\text{Quantity supplied} = 4 = \text{quantity demanded} = 5.5 - 0.003P$$

The price is the solution to this equation, \$500, as (again) in the diagram.

The equilibrium price could be put back into the two individual demand curves to find out how much Ransom and Sutch each demand "in equilibrium."

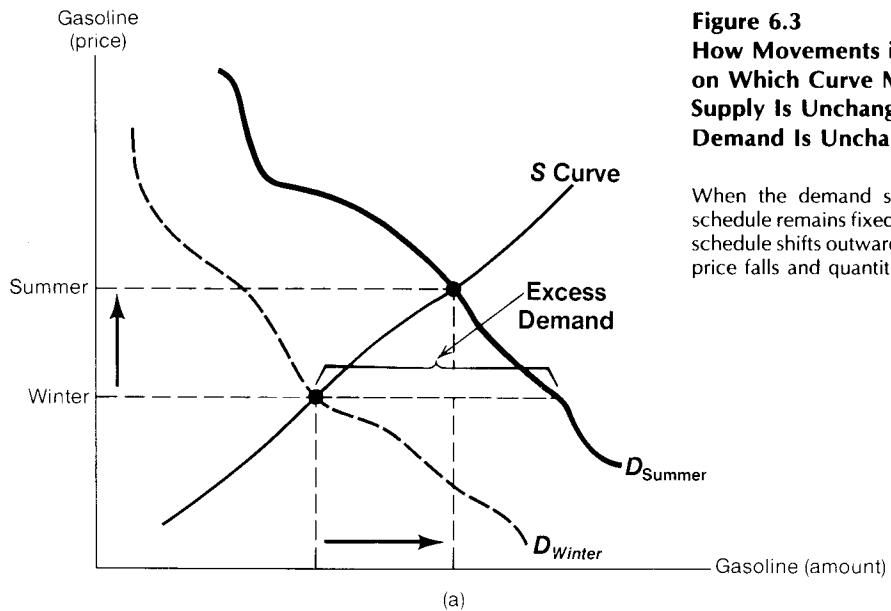
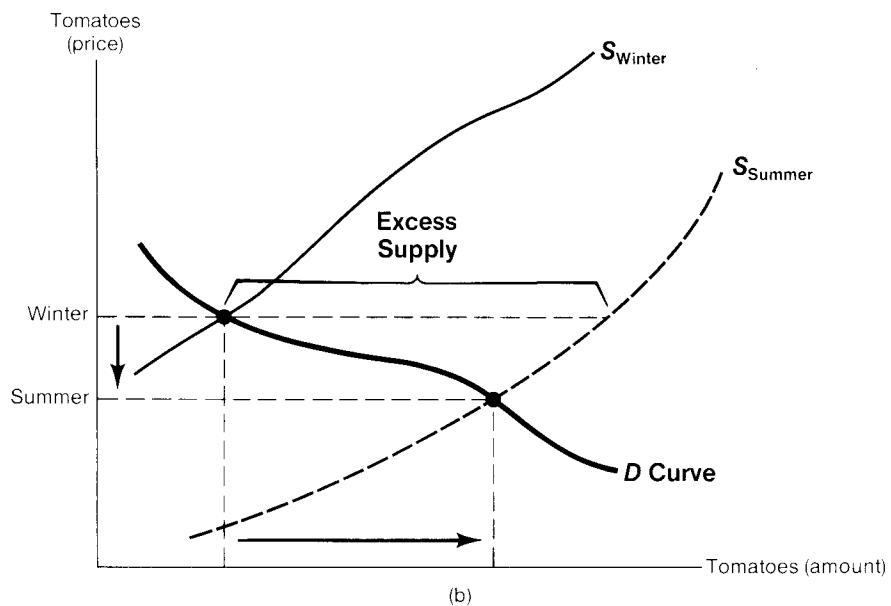


Figure 6.3
How Movements in the Price and Quantity Depend on Which Curve Moves. When Demand Rises and Supply Is Unchanged (a); When Supply Rises and Demand Is Unchanged (b)

When the demand schedule shifts outward and the supply schedule remains fixed, price and quantity rise. When the supply schedule shifts outward and the demand schedule remains fixed, price falls and quantity rises.



Ransom, for example, demands $2.5 - 0.001(500) = 2.0$ rooms per year. The diagram mimics the algebra exactly. The individual demand and supply curves are added up into market supply and demand curves. The intersection of the two determines the equilibrium price and quantity for the entire market. And the equilibrium price can then be run back to the individual supply and demand

curves to determine the allocation of quantities among demanders and suppliers.

An astonishing range of economics problems reduces to the simple assertion that *the market must clear*, that is, that the price adjusts to make quantities supplied and demanded equal. They are equal in equilibrium, where the supply curve crosses the demand curve. The problem of predicting what will happen to the market price and quantity when supply or demand curves move is one example (see Figure 6.3).

Q: More gasoline and tomatoes are sold in the summer than in the winter. Yet the price of gasoline is higher and the price of tomatoes lower in the summer than in the winter. Why?

A: The “yet” is misleading. There is nothing unexpected about the way in which the prices move. The supply curve of gasoline does not vary from season to season, because the technique of making gasoline from crude oil does not vary. Likewise, the demand curve for tomatoes does not vary from season to season, because the tastes and incomes of tomato eaters do not vary. But as people take to the road in the summer to visit Aunt

Louise or to prevent Aunt Louise from visiting them the demand curve for gasoline moves out, and as northern gardeners harvest their crops the supply curve of tomatoes moves out. This is shown in Figure 6.3. The prices must adjust to clear the market, that is, to make equal the quantities demanded and supplied. At the low winter price of gasoline in the face of the summer demand the market would not clear. There would be *excess demand*, which would force the price up. At the high winter price of tomatoes there would be *excess supply*, which would force the price down. As the diagram says, the movements in price depend on which curve moves—supply or demand.

Inferring Supply and Demand from Price Movements

The argument can be run in reverse, inferring what has happened to market supply and demand curves from how price and quantity behave (see Figure 6.4).

Q: Since the 1940s both the quantity and the relative price of medical care have increased. Therefore, the demand curve has moved out faster than the supply curve.

A: You observe the two points in the left panel of Figure 6.4. If the market is in equilibrium, these are points of market clearing, that is, of crossing of supply and demand curves. The two points are explained in the

right-hand panel. Notice that the demand curve has moved out faster than has the supply curve (compare the short and long arrows out of the old equilibrium, Once). So, true. You would not be surprised to learn that lots happened on the demand side of the market, such as Medicare for the elderly, inducing the elderly to buy more.

The Market Must Clear

The following amazing assertion depends only on recognizing that a fixed quantity supplied is a perfectly vertical supply curve and that the market must clear.¹

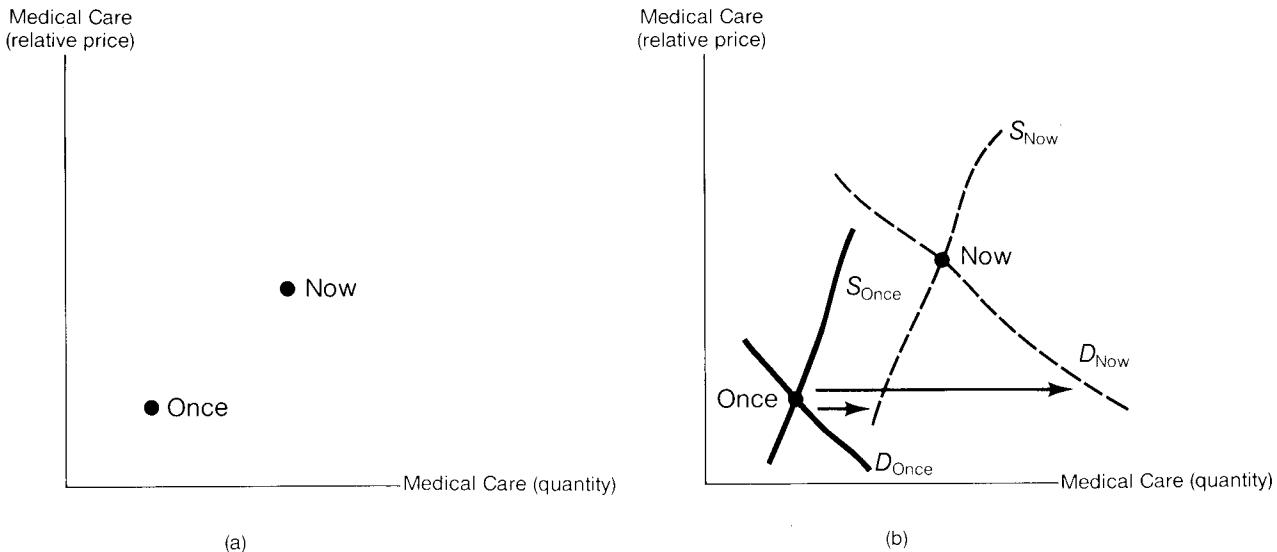
Q: In years of bad harvests in England a century ago, the rich would buy up grain and sell to the poor all they desired at half the market price. *True or false*: Contrary to its appearance and intent, the plan did not benefit the poor at all, but benefited instead the holders of grain (other rich men, grain dealers, and so forth).

A: True. Before the intervention of the charitable rich, the grain dealers sell the amount Available Supply of grain to the poor at the Equilibrium Price the market will bear (see Figure 6.5). As shown, after the intervention the grain dealers still sell the same supply to the poor, who must still be induced to consume only the

¹ Due to Mountiford Longfield (1802–1884), an Irish economist who anticipated early in the nineteenth century most of the developments in economics late in the nineteenth century. George Stigler uses the example in “The Politics of Political Economists,” in G. L. Stigler, *Essays in the History of Economics* (Chicago: University of Chicago Press, 1964), p. 60.

Figure 6.4
Changes in Price and Quantity: What You Observe (a) and How You Explain the Observations (b)

Price and quantity are greater Now than Once before. The cause must be an outward shift in the demand curve. Even if the supply curve also shifted out, the shift must have been less than the shift in the demand curve.



Available Supply. The sole difference in the situation is that now the grain passes through the hands of the rich, who mark it down to half the price they paid. Since the final price to the poor cannot change (it must be Equilibrium Price if the poor are to have only the Available Supply), competition among the rich for the

grain to sell to their poor will cause the price the rich pay to exactly double, to Double Price in the diagram. The plan enriches holders of grain (who get Double Price) and has no effect on the poor (who still pay Equilibrium Price).

COMMENT Another way in which to see the truth in the argument is to imagine in the diagram a demand curve by the charitable rich (a portion of which is drawn as a dashed line through Double Price). The rich will pay a price double whatever the poor will pay; the poor will pay Equilibrium Price for what is available. Therefore, the demand by the rich "derives from" the demand by the poor, the willingness to pay being exactly double the poor's. The quantity to be sold determines the poor's willingness to pay, which determines the rich's willingness to pay (namely, double). The rich fail in their attempt to benefit the poor because any attempt to sell the grain to the poor at less than Equilibrium Price causes the poor to demand *more than is available*. This causes the rich to demand more than is available, which causes the price paid by the rich to rise.

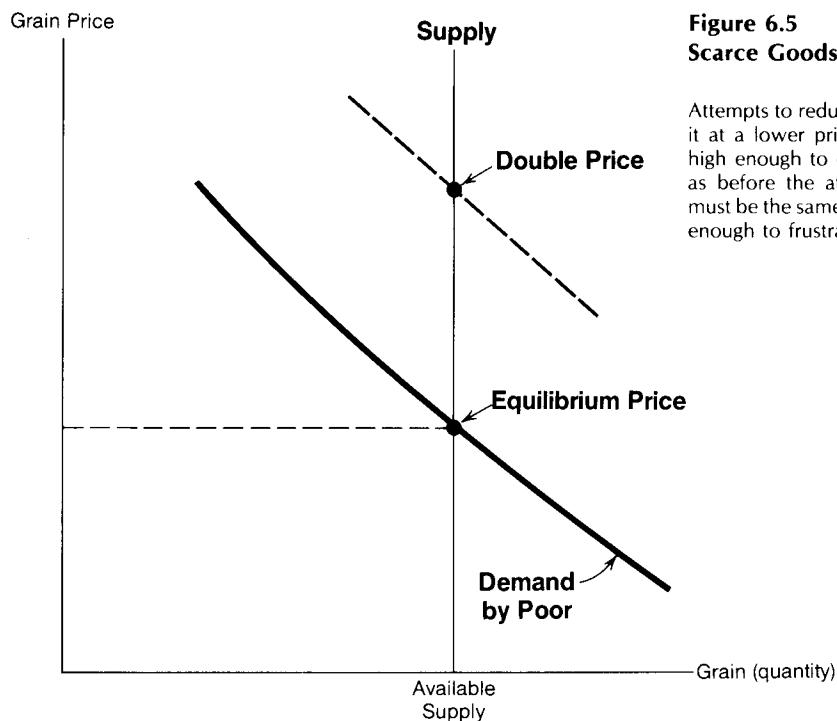


Figure 6.5
Scarce Goods Must Somehow Be Allocated

Attempts to reduce the price of a good in fixed supply by selling it at a lower price will fail. The price after lowering must be high enough to choke off demand to exactly the same extent as before the attempt to lower the price. That is, the price must be the same, implying that the market price will rise exactly enough to frustrate the attempt to help buyers.

Summary The equilibrium point of a market is the price at which the quantity supplied is equal to the quantity demanded, or equivalently, the quantity at which the price suppliers will accept is equal to the price demanders will accept. It is the intersection of the curves of supply and demand. To use the analysis of supply and demand, you must rivet your attention on this equilibrium point. The supposition that the point is achieved is remarkably powerful, explaining why prices vary, why observed prices and quantities reveal whether supply or demand varied more, and why amounts supplied must be limited if the price is to be high or the price high if the amount is limited.

EXERCISES FOR SECTION 6.2

1. State which *direction* market price and quantity for housing will move in each case relative to the original equilibrium (4 rooms bought at \$500 a room) if:
 - a. DeCanio's room burns down.
 - b. Wright builds 2 more rooms.
 - c. Higgs buys the 3 other rooms from DeCanio, Reid, and Wright (and then sells as before to Ransom and Sutch).
 - d. Sutch's demand curve changes from $Q_S = 2.5 - 0.001P$ to $Q_S = 3.5 - 0.001P$.
 - e. Ransom's demand curve changes from $Q_R = 3.0 - 0.002P$ to $Q_R = 3.0 - 0.003P$.
2. Are the supply and demand curves drawn through points Once and Now in Figure 6.4 the only ones that you could draw? That is, does the evidence of points Once and Now force you to draw these particular curves, with their particular slope and position?

PROBLEMS FOR SECTION 6.2

- **1.** The price of meat (sugar, coffee, toilet paper, or whatever) is now exorbitant. *True or false:* A good scheme for solving the problem is for consumers to boycott meat (sugar, coffee, toilet paper, or whatever), driving down its price; when its price is low, consumers can resume their higher purchases.

True or False

- 2.** The two observed points in Figure 6.4 are consistent with any number of supply and demand curves running through them.
- 3.** If the Organization of Petroleum Exporting Countries wishes to hold the price of oil above its equilibrium level, it must hold the amounts its members supply below their equilibrium levels.
- **4.** The banning of American exports of wheat to South Africa will cause more people in Zaire (another wheat-importing country) to starve.

6.3 Adding Up Supply and Demand: The Diagrams for Analyses

What to Read For

What is the *Law of One Price*? Why is it necessary for using supply and demand curves? What is the general technique for handling questions involving supply and demand? How does the equilibrium price determine how much is demanded or supplied by each segment of the demanders or the suppliers? How do you prove that a small demander or supplier should think of herself as a price taker?

The Law of One Price

Another way of putting the idea that a market must clear is the *Law of One Price*. In equilibrium, all participants in a market for housing of a certain sort must be paying or getting the same price. If the Law of One Price is violated, there are further deals to be made, further reshuffling of suppliers and demanders. The market has not yet cleared.

The practical significance of the Law of One Price is that when it is true the *source* of a shift in supply or demand does not matter. It's easy to get confused about this.

Q: Cotton cloth is cheap to transport and many countries produce it. Between 1913 and 1927 British exports of cotton cloth fell from 7.1 billion to 4.2 billion linear yards. In 1913 the chief market for British cloth had been India, taking 3.2 billion of the 7.1 billion. By 1927, however, India had developed its own cotton cloth industry and took only 1.8 billion yards of British cloth. *True or false:* If India had not developed its own industry, British exports of cotton cloth would have fallen half as fast as they in fact did between 1913 and 1927.

A: Put away your pocket calculator and ignore the arithmetic, for it is only relevant (and tenuously so even then) if India and Britain were the only consumers and producers in their markets for cotton cloth. But they were not. Both were in a single world market. The outward movement of the Indian supply curve (this is what it means to say that it developed its own industry) would push back the demand curve facing *all* exporters of cloth and would drive down the world price. This, in turn, would reduce the amount exporters, such as

Britain, would want to supply. But the fall in the Indian demand would not reduce yard for yard the demand facing Britain, as is implied by the arithmetic. That Britain in an arithmetical sense "depended" on Indian imports reflects minor advantages of political connec-

tion and of special treatment in India for British exports. Aside from these minor advantages, British exporters faced a world market (a world price). The location of the demand did not matter. False.

Why Price Taking Is Necessary

The Law of One Price is obviously necessary for the adding up of the curves: "Adding up" must be in some definite direction, for example, horizontally, at one price. It wouldn't make sense to add John's demand at \$5 to Laura's at \$300. Price taking is also necessary: If quantities are to be added at a given price, the price must be given.

T or F: A group of demanders who are not price takers do not have an aggregate demand curve.

A: The market does not give them a fixed price. That is, the price they get depends on the amount they demand (each one faces a rising supply curve, not a flat,

horizontal price). A demand curve, however, is a schedule of amounts purchased at various *given* prices. Not taking price as given, these "price-searching" demanders do not have such schedules. That is, true. The analysis of supply and demand applies only to price takers.

A great deal more than mere theoretical tidiness comes out of the idea of adding up individual supply and demand curves. It is the heart of applied economics.

Uses of Adding Up: The Keynes Problem

The general technique for handling questions involving supply and demand is pretty simple. (1) Identify all the suppliers and demanders in the market, grouping them into relevant categories. (2) Write down in one form or another the condition for equilibrium: Summed supply equals summed demand. (3) Draw the corresponding diagram. (4) Answer the question.

Being the greatest economist of the twentieth century is no protection against missing one or all of these steps.²

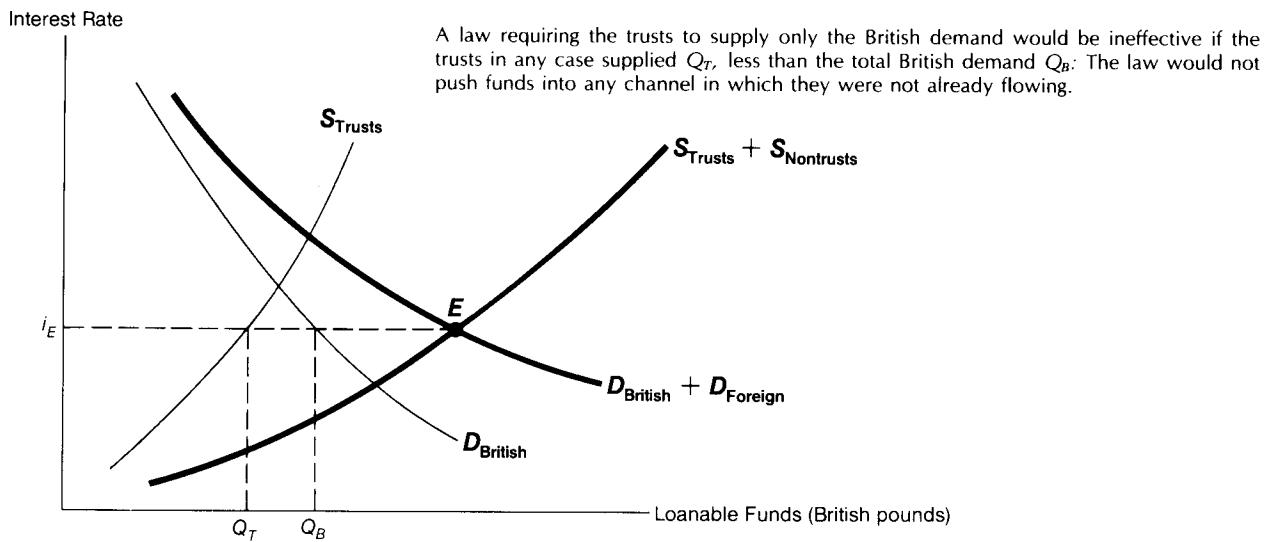
Q: In 1924 J. M. Keynes, a well-known economist of the day, noted that the Colonial Stock Act of 1900 permitted British trusts (holding funds of charities, trade unions, and the like) to invest abroad. Before 1900 they had been required by law to invest in Britain. The effect of the act, he said, was "to starve home developments by diverting savings abroad and consequently, to burden home borrowers with a higher rate of interest than they would need to pay otherwise." Accepting for the purposes of argument that a rise in the interest rate is a

Bad Thing, was Keynes correct? You may wish to know that both before and after the act there were others (private investors, banks, and so forth) besides the trusts making investments at home. Think of the interest rate as the price (so many percent per year) and the "loanable funds" the quantity (so many million British pounds of money for investment).

A: Keynes was wrong. Consider the situation after the trusts are permitted to invest where they please. The

² John Maynard Keynes (his father, John Neville, was also an economist), 1883–1946, First Bursar of King's College, editor of the *Economic Journal*, advisor to the British Treasury, patron of the arts, first Baron of Tilton, architect of half a century of economic thinking and economic policy, was, to use his own description of Ricardo, "the greatest mind that found economics worthy of its powers" [this in R. F. Harrod's, *The Life of John Maynard Keynes* (New York: Harcourt Brace & World, 1951), p. 467]. "Keynes" rhymes with "brains." The argument and quotation in the question come from his "Foreign Investment and National Advantage," *Nation and Athenaeum* 35 (1924): 586.

Figure 6.6
A Requirement to Supply a Particular Demand Is Not Constraining If Others Also Supply It



nontrust British investors must be included in the market. The equilibrium rate of interest occurs when the supply of loanable funds from the trusts plus the supply of funds from the nontrusts (both British) equals the demand from British users of funds plus the excess demand of foreign users of funds. That is,

$$S_{\text{TRUSTS}} + S_{\text{NONTRUSTS}} = D_{\text{BRITISH}} + D_{\text{FOREIGN}}$$

The diagram corresponding to this equation adds horizontally the supply and demand curves of the usual shape (see Figure 6.6). The heavy curves are the summed supply and demand curves that determine equilibrium. Notice that, as the diagram is shown, at the equilibrium interest rate, i_E , the British quantity demanded of funds, Q_B , is greater than is the quantity supplied by the trusts, Q_T . Now suppose (to run history backward) that the trusts are required by law to invest in Britain. What happens? Nothing of consequence. Since the trusts can place all their funds in Britain at the existing interest rate without oversupplying the Brit-

ish market (Q_B is greater than Q_T), nothing happens to the equilibrium interest rate when they obey the law and bring their funds home. Some of the nontrusts formerly in Britain now move to foreign loans, trading places with the trusts. But no gap between British and foreign interest rates can develop. The nontrusts still investing in Britain are free to move to foreign loans and would do so if such a gap developed, raising British interest rates and lowering foreign interest rates until the gap disappeared. The presence or absence of the Act, therefore, has no important effect, and no effect at all on British interest rates, contrary to Keynes's assertion. Only if the trusts supplied more funds than British demand would absorb at the equilibrium interest rate would the presence or absence of the Act matter. That is, only if the diagram were redrawn to make Q_T larger than Q_B would Keynes be correct. But Q_T greater than Q_B contradicts the observation that nontrusts supplied some British demand before the Act.

Other Uses of Summing Supplies and Demands

The choice of the equation and diagram that best suits the problem at hand is an acquired skill. For example, when the geographical separateness of two sets of demanders and suppliers is important, the curves can be separated. The American television market is separated from the Japanese (America imports as well as produces television sets). The equilibrium condition is that the television

sets supplied by the Japanese in excess of Japanese demand find a home in America fulfilling demand in excess of American supply. The equation that says this is: $S_J - D_J = D_A - S_A$. It is merely a rearrangement of the assertion that world (Japanese and American) demand equals world supply, $D_A + D_J = S_A + S_J$. The diagram corresponding to the rearranged equation is shown in Figure 6.7.

The American and Japanese diagrams share a price axis because the markets have the same price. To accommodate the shared axis, the Japanese curves run in the opposite direction, away from zero to the left instead of the right. In equilibrium the excess supply from Japan equals the excess demand in America. At P_E Japanese exports equal American imports, as indicated by the slashed segments.

T or F: If an effective quota is imposed on American imports of television sets from Japan, then the American price will rise and the Japanese price will fall. American consumers of television sets will be hurt; Japanese consumers will be benefited.

A: Before the quota the quantity of American imports is equal to the excess demand in America. To have any effect the quota on the number of television sets that can legally be imported must be less than the current quantity of American imports. If the quota is effective, then the segments Excess Demand in the American half of the figure and Excess Supply in the Japanese

half become equal to the quota. When the imposed quota is effective the Excess Demand (the quantity of American imports) and the Excess Supply (the quantity of Japanese exports) must be smaller than they are initially. This is possible only if American excess demand is choked off by a higher American price and Japanese excess supply choked off by a lower Japanese price. (Look at the diagram; only under these circumstances will a smaller segment fit between both curves.) A gap will form between the high American price and the low Japanese price. American consumers get a higher price, Japanese consumers a lower price. True.

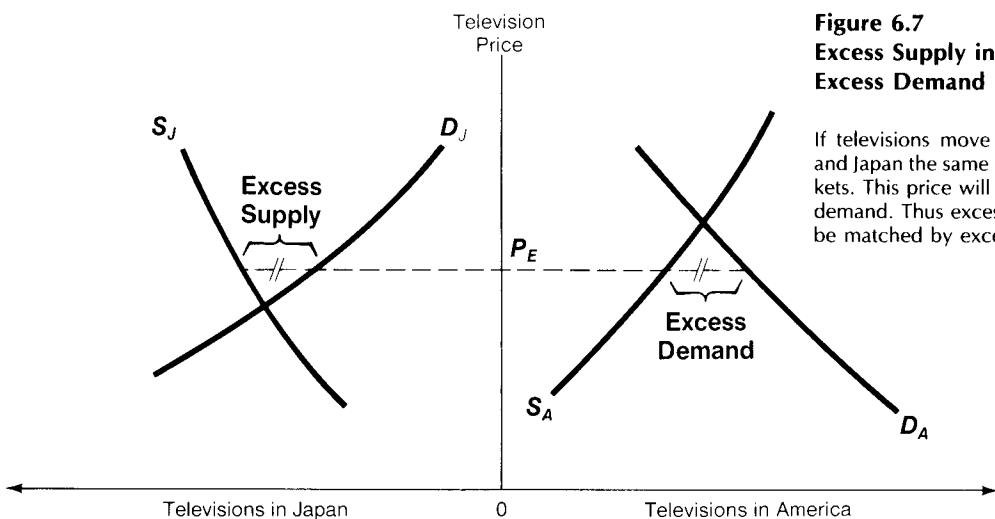


Figure 6.7
**Excess Supply in One Market Equals
Excess Demand in Another**

If televisions move costlessly between America and Japan the same price will prevail in both markets. This price will equate total supply and total demand. Thus excess supply in one market must be matched by excess demand in the other.

**Why Price Taking Is
a Consequence of
Summing**

The groups of suppliers and demanders don't have to be similar in size. The small consumer in a big market shows why he should take prices as given. Consider the situation of Peter Lindert, one consumer of bread among 200 million. The supply curve of bread facing him is the total supply (suppose that it is 10 billion loaves per year) minus the summed demand curves of the other 199,999,999 consumers. This assertion is, again, merely a rearrangement of the assertion that total supply equals total demands. $S_{\text{TOTAL}} = D_{\text{LINDERT}} + D_{\text{OTHERS}}$ can be written as $D_{\text{LINDERT}} = S_{\text{TOTAL}} - D_{\text{OTHERS}}$. The right-hand side is the *supply curve facing Lindert*. A supply curve of bread is a schedule of the amounts of bread one can get by offering the market various prices. Suppose that Lindert offers 0.10 cent over the market price for loaves of bread. If he buys up everything the market offers him at this new price, the price will rise by 0.10 cent. Suppose that such a rise causes each other consumer to reduce his annual consumption of bread by a quarter of a loaf. Therefore, if Lindert really does buy all that the market offers him at the new price, he will be able to buy what's left over from the cut of $\frac{1}{4}$ in the loaves per person, namely $\frac{1}{4}(199,999,999) = 20$ million more loaves of bread. That is, the logic of adding and subtracting supply and demand curves implies that *Lindert faces a virtually horizontal supply curve*. Consuming now, say 50 loaves a year at 40 cents each, he can if he wishes consume the colossal quantity of 20,000,050 loaves by raising the price he offers very slightly, to 40.1 cents each. Clearly, he should take the price as given.

The same point applies to less extreme cases. One apartment renter among 1000, or 1 automobile buyer among 10,000, faces a supply curve so nearly horizontal that nothing is gained by keeping in mind its slight upward slope. One demander among many faces a given price at which she can buy virtually all that she wishes without affecting the price. By similar logic, one supplier among many is also a price taker. In other words, the logic of supply and demand, which assumes price taking, itself justifies the assumption in the case of large numbers of participants in a market. All this comes from manipulating the obvious equation that "the sum of all supplies must equal the sum of all demands."

Summary

The summing of supply and demand curves is essentially an application of the Law of One Price in conditions of price taking. Each arrangement of the condition for equilibrium—supply equals demand—has a corresponding diagram. The diagrams are among the most useful tools of the economist's craft. They make it clear, for example, why the initial allocation of a good does not matter for its final allocation, precisely where large numbers of participants fit into the theory of supply and demand, and how the theory is used.

EXERCISES FOR SECTION 6.3

1. Exports of wool cloth from Britain to America increased a lot in the eighteenth century. In fact, they increased faster than British wool cloth exports elsewhere. Does this necessarily mean that British exports would not have increased without the American market? (*Hint:* Compare the Indian cotton problem.)
2. The auto industry in the world does not have many major companies. General Motors,

for instance, takes up a pretty big share of the markets in which it sells. *True or false:* It's *not* sensible in this situation for General Motors to think of itself as a price taker, facing a nice, flat demand curve at a given price. (*Hint:* Look at the role played by Lindert's small share of the whole demand curve for bread.)

3. *True or false:* In view of Exercise 2, the auto industry does not have a supply curve. (*Hint:* Look at the logic of non-price takers not having a *demand* curve.)

PROBLEMS FOR SECTION 6.3

- **1.** Devise a diagram that embodies a constant price differential (due to, say, transport costs) between the price of the television exporter (Japan) and the television importer (America).
- **2.** Helen Louise Stueland has a demand curve for stocks of clothing: The lower the price, the larger her wardrobe. Draw a diagram exhibiting how she determines each year how much new clothing to buy in view of her existing stock, supposing that half the stock wears out each year.
- **3.** If Britain has a fixed exchange rate (that is, the number of U.S. dollars per British pound is fixed) then inflation abroad can be imported. Since Britain imports wheat and autos, a rise in the world price of wheat or autos will raise British prices. It is usually asserted that the strength of this effect depends on the ratio of imports to national expenditure, that the effect will be stronger in Britain (where imports divided by income equal 0.26) than in America (where it is only 0.06). But America and Britain import roughly the same list of commodities. What is different is merely the share of total British consumption of, say, autos provided by imports. *True or false:* The Law of One Price suggests that the strength of imported inflation would be the same in both countries.
- **4.** In 1907 the British iron industry accounted for three-fourths of world exports of pig iron but only one-sixth of world production. Which figure is the best summary measure of Britain's "importance" in the market for pig iron?

True or False

- **5.** The ownership of housing in Iowa City before trade begins does not affect the ownership after trade.
- **6.** A high television tariff (a tax on television sets shipped from Japan to America) could stop Japanese exports to America entirely.
- **7.** The proof of a horizontal supply curve of bread facing a single demander would break down if bread were not assumed to be in fixed supply.

6.4 Extensions of Supply and Demand

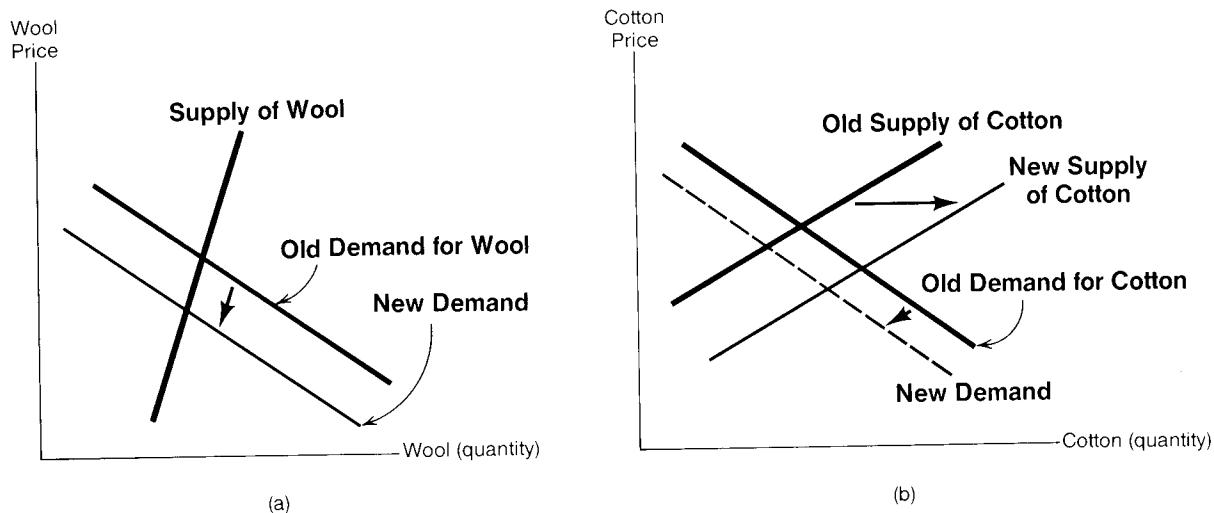
What to Read For	How can "substitute" and "complement" be defined in terms of the effect of price on quantity demanded? How is this definition related to the earlier definitions of "substitute" and "complement" given in terms of the shapes of indifference curves? What are related markets? When things like wool and mutton are jointly supplied, why is the demand curve for the whole thing (sheep) the vertical sum of the separate demand curves? Why is a supply curve the vertical sum of the willingness to supply each input into the good?
-------------------------	--

Substitutes and Complements Require Two Diagrams

Supply and demand is the economist's first and best tool. It can be applied, for example, to related markets, such as the markets for cotton and wool cloth or the markets for cotton cloth and the buttons, needles, and sewing services to make it into clothing. These markets are related through their demand functions. The demand function for cotton contains, of course, the price of cotton. The price enters with a superscript minus sign attached to it to signify that the quantity demanded goes down when the price goes up: $Q_{\text{Cotton}}^{\text{DEMAND}} = f(P_{\text{Cotton}}^-)$. The negative sign is represented by the downward slope of an ordinary demand curve. A fuller statement of the demand function for cotton would include among other things the price of substitutes (wool, P_w) and complements (sewing services, P_s): $Q_c^D = f(P_c^-, P_w^-, P_s^-)$. Notice the signs of the superscripts. These tell how the ordinary demand curve shifts in or out in response to changes in the prices of substitutes and complements. The definitions of "substitute" and "complement" given earlier depended on the shape of indifference curves; the definition here, more direct and useful, depends on the sign of the effect of price on quantity demanded. The two definitions are connected. More expensive wool causes substitutions in favor of cotton; more expensive sewing services, on the other hand, cause the price of clothing to rise and the quantity demanded of clothing and (therefore) cotton to fall.

Figure 6.8
One Way to Think About Related Markets: Markets for Wool (a) and Cotton (b)

An outward shift in the supply curve of cotton causes a fall in the price of cotton. Because cotton and wool are substitutes, a fall in the price of one causes a downward shift in the demand curve for the other. Thus the original shift in supply causes a downward shift in the demand curve for both goods.



Q: The supply curve of cotton moves out. *True or false?*

The less sensitive to price is the quantity of wool supplied, the smaller will be the increase in the quantity demanded of cotton.

A: The assertion is depicted in Figure 6.8. The heavy curves are the supply and demand curves before the disturbance. After the supply curve of cotton in the right panel moves out to New Supply of Cotton, the price drops. Since cotton is a substitute for wool in the production of clothing (signified in the initial demand curve for wool by the plus sign attached to P_C) the fall in the price of cotton pushes in the demand curve for wool. This is indicated by the light demand curve in the left panel. Because the supply curve of wool slopes upward—instead of being flat—the price

of wool will therefore be driven down. And because wool is a substitute for cotton, the fall in the wool price will now in turn push the demand curve for cotton downward. As required by the assertion, the increase in the quantity demanded of cotton will be smaller. Notice that the feedback from the market for wool to the market for cotton would be cut if the supply curve of wool were flat, that is, if the quantity supplied of wool were very sensitive to its price. In that case the price of wool would be pegged by its flat supply curve, and the rise in demand would be unable to affect the price. The flatter the supply curve, in other words, the smaller the fall in the price of wool and the smaller the consequent fall in the quantity demanded of cotton, as asserted.

Joint Supply Requires Vertical Summation

A demand curve is a schedule of the amount people are willing to buy at a given price. At \$10 a head, Nicholas von Tunzelmann is willing to buy 40 sheep, buying (that is) out to the point at which the pleasures and profits he gets from an additional sheep are worth just \$10 to him. But looked at the other way, *the demand curve is also a schedule of the price that people are willing to pay at a given quantity*. At 40 sheep, von Tunzelmann is willing to pay \$10 for another one. Recall the earlier jargon: The price along a demand curve for sheep is the *marginal valuation* of a sheep at various quantities.

Suppose, then, that a sheep provides two things in rigidly fixed proportions: mutton (sheep's meat) and wool. And suppose that, having 5 sheep's worth of wool and mutton, von Tunzelmann values the wool from an additional sheep at \$15 and the mutton from the same sheep (slaughtered when its wool gives out) at \$20. What is von Tunzelmann's demand price at 5 sheep? It is $\$15 + \$20 = \$35$, the amount he is willing to pay for the wool plus the amount he is willing to pay for the mutton. That is, he is willing to pay \$35 for an additional sheep. If he has a large number of sheep—say, 40—he is willing to pay only a smaller amount for the wool and mutton of still another sheep, since at 40 sheep he is already well clothed and fed. That is, his demand curves for wool and mutton slope downward and therefore so does his demand curve for sheep, which is the *vertical sum* of the wool and mutton curves in Figure 6.9.

In other words, when things like wool and mutton, movies and air-conditioned comfort, lumber and wood shavings, vocational training and general education, or my protection from conquest and thine are jointly supplied, the demand curve for the whole thing (sheep, movies in hot weather, timber, college, or national defense) is the *vertical sum* of the separate demand curves.

The contrast between vertical and horizontal sums can be brought out by noting that the demand curve for sheep by Nicholas and Carol von Tunzelmann together is the *horizontal sum* of their separate demand curves for sheep, each of which is the *vertical sum* of their individual demands for sheep's worths of mutton and of wool. The order in which the horizontal or vertical summation is performed does not matter. Carol's demand for mutton at various prices could be summed horizontally with Nicholas's and then their willingness to pay for a given sheep's worth of wool and mutton summed vertically, or the vertical

Figure 6.9
The Demand for Sheep Is the Vertical Sum of the Demand for the Corresponding Wool and Mutton

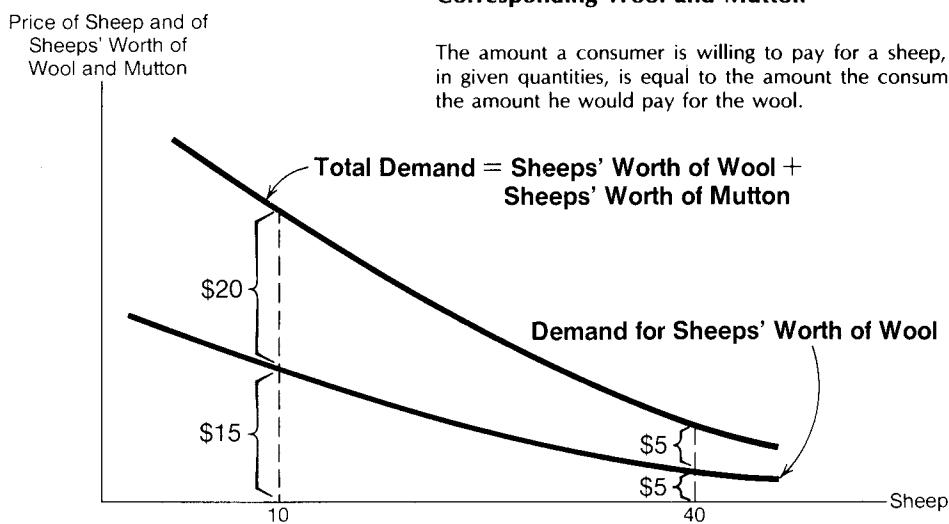
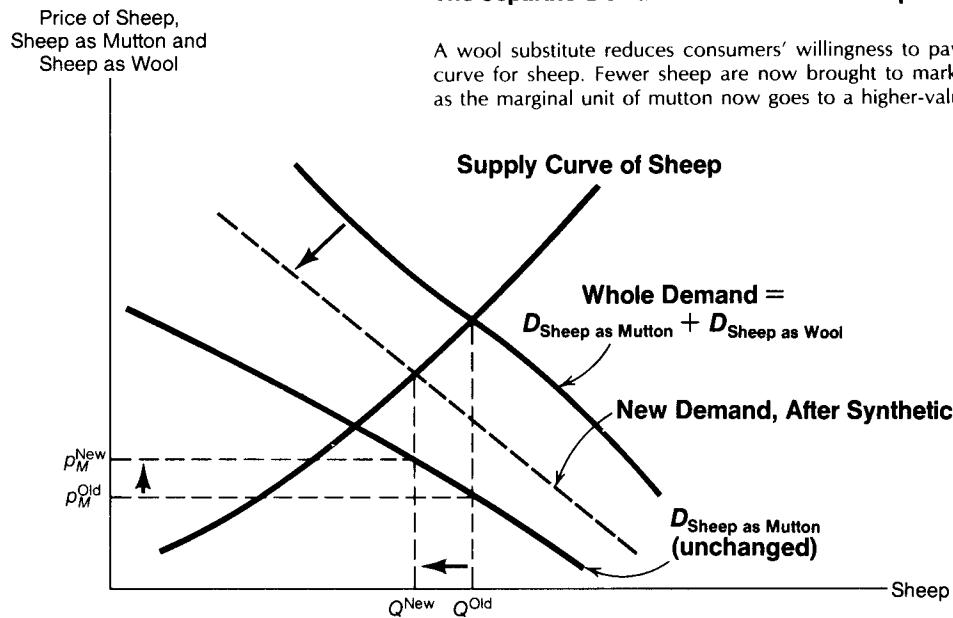


Figure 6.10
The Separate Demand Prices Must Add Up to the Supply Price



sum could precede the horizontal. Either gives the same result. The demanders of wool and mutton need not be the same people. As long as they bought the same sheep, Nicholas could demand only mutton, Carol only wool. Their demand curve for sheep would still be the vertical sum of the demands for sheep as mutton and sheep as wool. The critical point is simply that a single sheep provides double pleasure.

Facing the market's vertical sum of demands is an ordinary supply curve of sheep. Just as the equilibrium price implies an equilibrium distribution of the total quantity among demand curves summed horizontally, so too the equilibrium quantity implies an equilibrium distribution of the total price among demand curves summed vertically. That is, the prices at which wool and mutton sell separately are determined. This equilibrium has remarkable consequences.

T or F: The invention of a synthetic substitute for wool will raise the price of mutton.

A: The assertion sounds unlikely. But wait. The willingness to pay for wool will fall when the substitute arrives. The demand curve for sheep *as wool* will fall, reducing the desirability of raising sheep and reducing the quan-

tity supplied. The price of sheep as mutton will have to rise to induce consumers to consume the smaller quantity of sheep. Therefore, true. Figure 6.10 makes the same point as the words. The total demand price falls, which reduces the quantity supplied. At the lower quantity the price of sheep as mutton is higher.

Uses of Joint Supply

The idea of vertical addition is easily generalized. For example, the characteristics supplied jointly by a single product can be more than two. The price of a typical automobile, for example, is the sum of the values attached to its horsepower, weight, length, standard equipment, and so forth.³ Again, one can generalize the idea to joint demand, in parallel with joint supply.

T or F: A rise in the cost of houses will lower the price of furniture.

A: True. The whole product—call it furnished accommodation—becomes more expensive and the quantity de-

manded will fall. Since houses and furniture are demanded jointly, the quantity demanded of furniture will fall. Therefore, its supply price will fall (see Figure 6.11).

The applications of joint supply or demand are many. Consider the following.

Q: The price of slaves in the United States before the Civil War, it is said, consisted of not only the return to the business use of slaves (in house or field) but also a return to the prestige value of slaves. In other words, a slaveowner's social position is said to have depended on the number of slaves he owned, independent of the income he extracted from them. Suppose the national supply curve of slaves to be vertical at the number of slaves that existed (it was: slave importation was ended in 1807). Show that the price of slaves

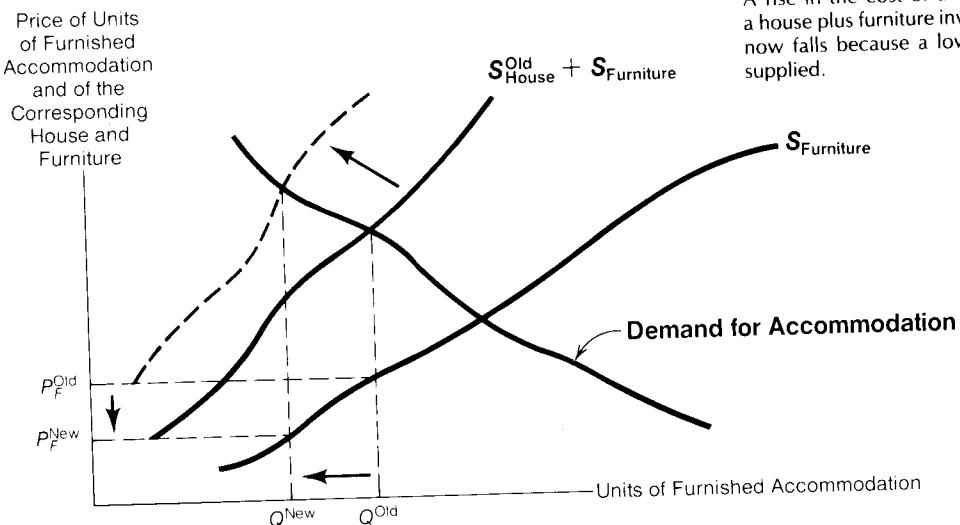
would be bid up above their business-only value. If you were told that in fact slaves sold for their business-only value, *not* higher, what would you conclude about the importance of prestige in the demand for slaves?

A: Because business value and prestige value are supplied jointly by the same slave, the two demand curves are summed vertically (see Figure 6.12).

The Business Price is the business-only value. It is what the slave would sell for if his or her only value

³ See Zvi Griliches, "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," *Government Price Statistics, Hearings*, U.S. Congress, Joint Economic Committee, January 24, 1961 (Washington, D.C.: GPO, 1961), pp. 173–196, reprinted in Arnold Zellner, ed., *Readings in Economic Statistics and Econometrics* (Boston: Little, Brown, 1968).

Figure 6.11
Joint Demand for Houses and Furniture



A rise in the cost of a house shifts the supply curve of a house plus furniture inward. The price paid for furniture now falls because a lower quantity of furniture is now supplied.

was to work in the owner's field or household. The Whole Price adds on the feelings of feudal dignity and political power that slaveowners are alleged to have enjoyed as well. But the allegation is false, for slaves *did* sell at the Business Price. Prestige was apparently not an important factor in the demand for slaves. The

broader significance of the point is that slaveowners would appear to have been capitalistic and calculating in their attitude toward slaves—not, as both critics and apologists of the system have sometimes argued, feudal and uncalculating.⁴

A final and important application of joint supply is to what are called *public goods*. These are supplied to all consumers when they are supplied to any. The usual example is national defense. If \$50 billion is spent on defense, that defense is provided to each resident of the country, whether or not the resident paid anything for it or wishes to pay anything for it. Henry David Thoreau spent a night in the Concord jail because he did not wish to pay taxes to protect peace-loving Americans against wicked and aggressive Mexicans; nonetheless he got the protection. Television signals or knowledge of economics are available to anyone with a television set or a library card. A fire station provides protection to all the houses in its neighborhood simultaneously. All these are supplied jointly to all consumers. A rational government deciding how much to spend on fire protection would sum the demands of all consumers vertically, as though on fire protection were selling sheep. The socially desirable amount is the amount at which the marginal

⁴ The argument and its extensions are controversial. It was first made by A. H. Conrad and J. R. Meyer, "The Economics of Slavery in the Antebellum South," *Journal of Political Economy* 66 (April 1958): 95–130; it was remade by Robert W. Fogel and Stanley L. Engerman, *Time on the Cross: The Economics of American Negro Slavery* (Boston: Little, Brown, 1974), pp. 70–73, and *Evidence and Methods—A Supplement* (Boston: Little, Brown, 1974), pp. 62ff. An attempt to unmake it is P. A. David and P. Temin, "Capitalist Masters, Bourgeois Slaves," *Journal of Interdisciplinary History* 5 (Winter 1975): 445–457.

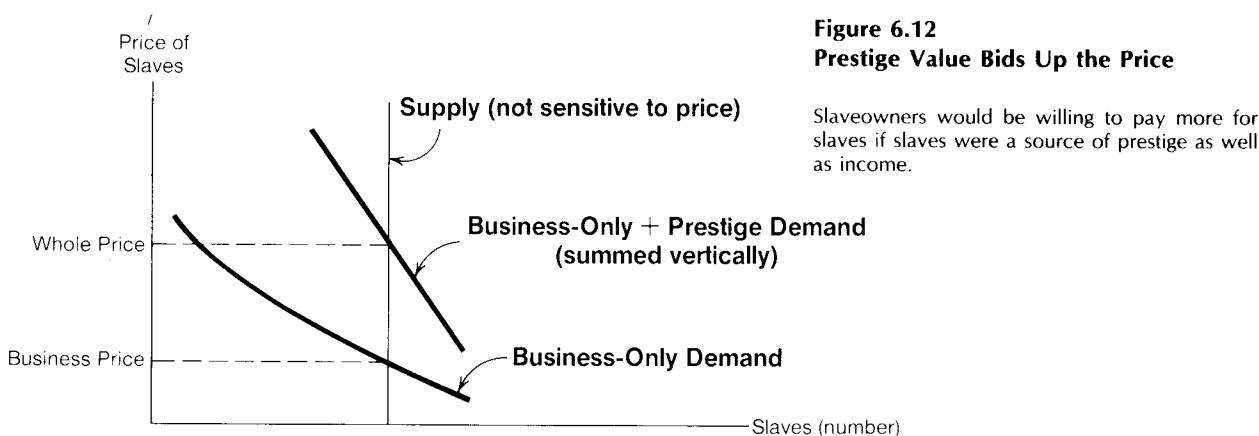


Figure 6.12
Prestige Value Bids Up the Price

Slaveowners would be willing to pay more for slaves if slaves were a source of prestige as well as income.

cost of fire protection is equal to the summed values of the marginal benefit. And likewise for knowledge or national defense.

Summary The idea of related markets, like that of equilibrium and of adding up supply and demand curves, is simple: Write down as a set of simultaneous equations the supply and demand curves in all the closely related markets. The idea has uses in analyzing oil and coal; cotton and wool; gasoline and automobiles; or trains, buses, and airplanes. The idea that a demand curve is a curve of willingness to pay leads naturally to the notion that it is the *vertical* sum of the willingness to pay for all characteristics of the good. A sheep provides wool and mutton, each having its own demand. So too a supply curve is a vertical sum of the willingness to accept each input into the good. A knife consists of a blade and handle, each with its own supply.⁵ These ideas are applicable to all manner of cases, from slavery to national defense. The rule of application is nicely symmetrical. If two things are demanded jointly in product Z , add up at given Z worths their supply curves; if two things are supplied jointly in product X , add up at given X worths their demand curves.

EXERCISES FOR SECTION 6.4

1. Using Figure 6.8 as a guide, examine the sequence of events when the supply of cotton moves from Old to New, supposing that:
 - a. The Supply of Wool curve is perfectly flat. Does the demand curve for cotton shift at all?
 - b. The Supply of Wool curve is perfectly vertical.
 - c. The demand for wool does *not* depend on the price of cotton.
2. In olden days, when air conditioning was uncommon, movie theaters with air condition-

⁵ Incidentally, the wool-mutton and blade-handle examples are both inventions of Alfred Marshall, *Principles of Economics*, 8th ed. (first published in 1920; New York: Macmillan, 1948; seventh printing, 1959), Book V, Chapter VI. Marshall (1842–1924), a mathematical physicist by training, was the guru of British economics from the 1880s on. He perfected the analysis of supply and demand, completing the marginalist revolution.

ing did very well in the summer. If Lee Alston values the average movie at \$2, what can you say about what he'll pay in the summer?

3. Match the pairs into vertically summed, jointly supplied goods, and explain:

One Half	The Other
1. Wool	A. Large-horsepower engine
2. Lothian's geometry course	B. Straw
3. Frank Lewis's security from thieves	C. Assurance of oil supply in the Persian Gulf
4. Comfort of a large car	D. Calories
5. Theory of joint supply	E. Mutton
6. Roads to the grocery store	F. Marglin's instruction in typing
7. Defense of Western Europe	G. Libecap's enjoyment of lawful and orderly traffic
8. Protein	H. Instruction in opportunity cost
9. Wheat	I. Sports news
10. Letters to the editor	J. Roads for commuting to work

PROBLEMS FOR SECTION 6.4

- 1. Suppose that instead of valuing slaves for both business and prestige simultaneously, slaveowners divided into two distinct classes, one (Cavalier Fops) valuing slaves *exclusively* for prestige, the other (Grasping Capitalists) *exclusively* for business. What now is the appropriate demand curve? If you take measurements of the business value of slaves from business-like (and record-keeping) plantations, what relation will you find between the business value and the market price of slaves? Will this finding shed light on the prevalence of Cavalier Foppishness among slaveholders?
- 2. In England (and in other places) before the twentieth century, the ownership of land brought prestige and power. A successful London merchant would abandon trade for a landed estate, his son would ride to hounds, and his grandson would sit in Parliament. Investments in land earned a money return of 2% annually when comparably secure investments elsewhere earned 5%. Why?

True or False

- 3. In the example of cotton and wool cloth in the text, the lower demand curve for cotton does not end the story; a lower demand curve implies a lower price, which lowers still further the demand curve for the substitute, wool; this in turn lowers the demand curve for cotton again; and so on in a vicious cycle.
- 4. By symmetry with the parallel assertion about substitutes (cotton and wool cloth), the less sensitive to price the quantity supplied of sewing services (a complement to cotton cloth), the larger will be the increase in the quantity demanded of cotton cloth coming from an outward move in its supply curve.
- 5. If the supply of sheep is perfectly vertical, then the invention of a synthetic substitute for wool will have no effect on the price of mutton.
- 6. A rise in the demand for beef will reduce the price of dog food, soap, shoes, and leather jackets.

- **7.** If an additional year of schooling increases the student's future income, properly measured, in excess of the costs of tuition and forgone employment, but the student finds the additional year excruciatingly unpleasant, then the student will quit school prematurely from the point of view of maximum money income.
- **8.** The prestige value of slaves can enter the utility function of slaveowners (along with the business value of slaves) yet play no part in determining the market price of slaves.

Measuring Supply and Demand

7.1 Elasticities of Supply and Demand: The Essential Ideas

What to Read For

How sensitive is consumption to changes in income? To changes in price? What is the elasticity with respect to income? What is the elasticity with respect to price? What is an Engel curve? What shapes do curves of infinite, zero, or unit elasticity have? How do you know if a point on a demand curve is an elastic or an inelastic point? Why is an elasticity of 1.0 significant?

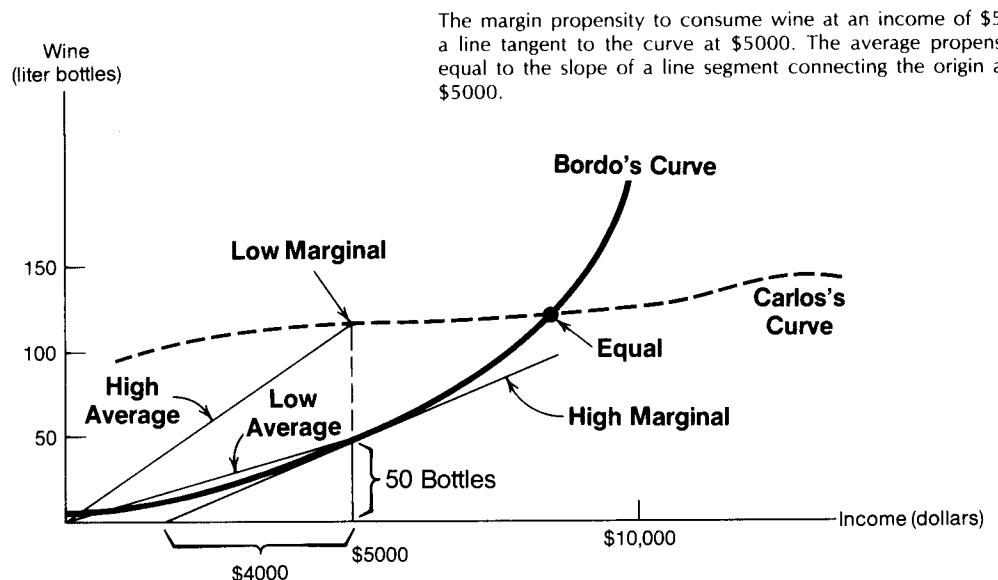
Average and Marginal Propensities to Consume

The argument so far has been qualitative, not quantitative. That is, the scale of diagrams and the exact shape of curves have not mattered. We have made statements such as: The farther a family lives from the center of New York, the cheaper its housing space and the more it will consume. The more unlikely a fire, the less a householder will pay for insurance. The higher the price of wheat, the more a Kansas farmer will supply. The higher a Californian's income, the more she will consume of a normal good, such as drinkable wine. These statements all indicate whether consumption rises or falls. But they do not answer the question of *how much* consumption will change. It is now time to pay attention to how much.

Diagrams, if taken literally, contain quantitative opinions. Consider Figure 7.1. Suppose that the solid curve represents Michael Bordo's wine consumption plotted against his income. The dashed curve is Ann Carlos's wine consumption plotted against her income. The solid indicates that income has much influence on Bordo's consumption of wine (ignore the light tangent lines for a moment). The rise in wine consumption for a given rise in income is large along the solid curve. It is small along the dashed curve.

The qualitative opinion that a higher income causes higher wine consumption does not distinguish between these two *quantitative* opinions. It says that there is some sensitivity of wine to income, but not how much. The how muchness depends upon the shape of the Wine-Income curve. It is measured by a slope

Figure 7.1
The Slope of the Wine-Income Function Measures the Sensitivity of Wine to Income



The margin propensity to consume wine at an income of \$5000 is equal to the slope of a line tangent to the curve at \$5000. The average propensity to consume at \$5000 is equal to the slope of a line segment connecting the origin and the consumption path at \$5000.

of the Wine-Income curve, that is, by the slope of the line tangent to it. The sensitivity of consumption at \$5000 of income along the solid curve, for example, is measured by the rise over the run of the tangent line. Look at it—it is 50 bottles divided by \$4000, or, whipping out your calculator, 0.013 bottle per dollar.

The sensitivity of consumption to changes in income, or the *marginal propensity to consume* wine (to use other words), is lower along the dashed line. At \$5000 of income, for example, it is nearly zero. That is, the dashed Wine-Income curve is at this point nearly flat. On the other hand, at \$5000 the *average propensity to consume* wine is higher along the dashed line. The average is the slope of a ray from the origin out to the point at which the average is being measured, and High Average has a higher slope than does Low Average. Symbolically, if income is I and wine W , the average propensity is W/I and the marginal $\Delta W/\Delta I$ (the *change* in wine divided by the *change* in income). At \$5000 Ann Carlos drinks a lot of wine. She has a high average propensity to consume but she is insensitive to changes in income (low marginal).

T or F: At the point labeled Equal in the diagram, both Carlos and Bordo have the same average and marginal propensities to consume wine.

A: False. They have the same average but not the same marginal propensity. That is, the ray from the origin to Equal is the same for both curves, but the slope of the solid curve at Equal is higher.

Similar thinking applies to more important problems:

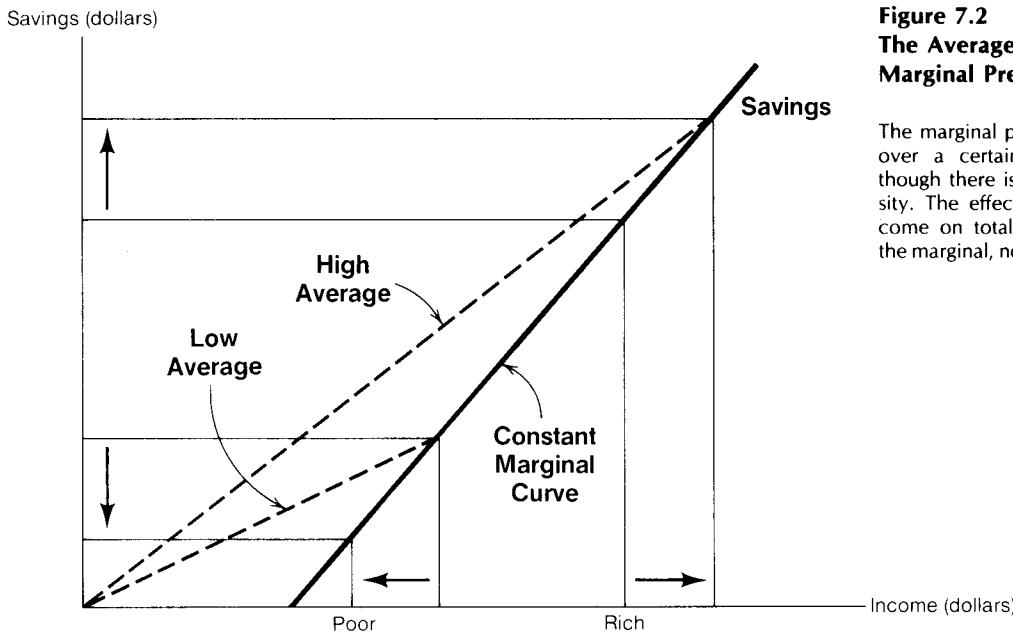


Figure 7.2
The Average Describes, the
Marginal Predicts

The marginal propensity may be constant over a certain range of income even though there is a falling average propensity. The effect of a redistribution of income on total savings is determined by the marginal, not the average, propensities.

T or F: If the rich have a higher average propensity to save than do the poor, a redistribution of income away from the poor and toward the rich will raise total saving.

A: The crucial question here is *how much* total saving by the poor will decline and total saving by the rich will rise if income is redistributed. Therefore, what matters is the marginal, not the average, propensity to save. A dollar taken from the poor reduces their savings according to their marginal propensity to save out of the dollar, giving the dollar to the rich increases their savings according to their marginal propensity. The average propensity is irrelevant, as can be seen in Figure 7.2. The average propensity to save of the person with Rich income is higher than that of Poor (look at the slopes of the dashed rays from the origin), yet the reshuffling of the dollar has no effect at all because the marginal propensity is the same. So, false.

Q: What do you think of this argument that increasing inequality of income increased savings in eighteenth-century England:

It is generally recognized that more saving takes place in communities in which the distribution of wealth is uneven than in those in which it approaches more closely to modern conceptions of what is just . . . the rise of new institutions, including that of the National Debt, intensified the disparities that had been handed down from earlier generations. . . . In this way, increasingly, wealth came into the hands of those whose propensity was to save, rather than to spend.¹

A: The argument is wrong. The fact that the rich save more than do the poor as a proportion of their incomes is not enough to make the case. The marginal, not the average, propensity to save is what predicts behavior.

¹ T. S. Ashton, *The Industrial Revolution 1760–1830* (London: Oxford University Press, 1948), pp. 7–8.

The Income Elasticity The average and marginal propensities are enough to make comparisons inside the world of wine drinkers or savers. The point of speaking quantitatively, however, is to use our intuitions about wine drinking compared with, say, bread eating, that is, to compare different products. One intuition, for example, is that bread eating is less sensitive to changes in income than is wine drinking. Bread is, in the vernacular, a "necessity," wine a "luxury" (at least outside France and its spiritual colonies). But to say that bread is "less sensitive" to income than is wine is meaningless if the measure of sensitivity is the marginal propensity: 0.013 liter of wine per dollar is neither more nor less than 1 loaf of bread per dollar. The difference in units makes it nonsensical to speak of moreness or lessness. The marginal propensity depends on the choice of units, itself arbitrary. If one measured wine in quarts (1 liter = 1.057 quarts) and income in \$5 units (\$5 = 1 fin), the slope of the wine-income function would change from 0.013-liter bottles per dollar to 0.068-quart bottles per fin.

What is wanted is some way of using the marginal propensity to spend on wine that rids it of this ambiguity. The way to do it is to divide the marginal propensity by the average propensity. The average propensity is measured in the same units (liters per dollar, quarts per fin, and so on) and is therefore made large or small by the choice of units in the same way as is the marginal propensity. Dividing the marginal by the average corrects, so to speak, for the arbitrary choice of units.

The result is the economist's measure of the sensitivity of one variable to another, the *elasticity* of wine consumption with respect to income:

$$\frac{\text{Marginal propensity}}{\text{Average propensity}} = \frac{\Delta W / \Delta I}{W/I} \\ = \text{elasticity of } W \text{ with respect to } I$$

The units of W and I cancel in the division. Liters divided by liters leaves a number: not of liters, just a number with no units. The elasticity has no units. It is simply a number, such as 1.0 or -3.6, with no units of weight, volume, money, or whatever attached.

That this is so is clear if the definition is rearranged a little:

$$\text{Elasticity of } W \text{ with respect to } I = \frac{\Delta W/W}{\Delta I/I}$$

This alternative definition reads: "Elasticity is the proportionate change in wine consumption, $\Delta W/W$, divided by the proportionate change in income, $\Delta I/I$, that causes the change in wine consumption." In other words, elasticity is the percentage change in wine consumption that is *caused by* a certain percentage change in income. The proportionate change in wine consumption is unaffected by the units in which the wine is measured: A rise in W from 10 liters to 11 liters is the same as a rise from 10.57 quarts to 11.63 quarts, namely, a 10% rise. A proportionate (or percentage) rise is unitless, and therefore so too is the ratio of two proportionate rises.

Elasticity, then, is a unitless measure of sensitivity. The elasticity of wine with respect to income is denoted by η_{WI} , the subscripts reading "of W with respect to I ." The Greek letter η , called "eta," is often used to symbolize an elasticity (so is ϵ , epsilon). The definition in its various forms is

$$\eta_{WI} = \frac{\text{marginal}}{\text{average}} = \frac{\Delta W / \Delta I}{W/I} = \frac{dW/dI}{W/I} = \frac{\Delta W/W}{\Delta I/I}$$

$$= \frac{d \log W}{d \log I} = \frac{\text{rate of change of } W}{\text{rate of change of } I}$$

These equations are worth pondering (don't worry if you don't know what " $d \log W$ " means; it's interesting only to students of calculus).

A curve plotting the amount purchased of something against income is called an *Engel curve*. Figures 7.1 and 7.2 were Engel curves for wine and savings. Figure 7.3 illustrates some elasticities of Engel curves for food. Notice especially the three simple cases: (1) The elasticity of food with respect to income is equal to zero. In this case, food purchases are completely insensitive to changes in income and the Engel curve is flat. (2) The elasticity is equal to 1.0. Food consumption is *proportional* to income; that is, the percentage increase in food consumption is exactly equal to the percentage increase in income. The Engel curve is any ray through the origin. (3) The elasticity is equal to infinity, with food purchases jumping from zero a little below Threshold Income to some large amount a little above Threshold Income. The Engel curve is vertical. The more curved curves can be thought of as combinations of these three extremes. An extreme version of the curve marked $\eta > 1$ would approach the shape of $\eta = \infty$, which is one way of recognizing that its elasticity is indeed greater than 1.0. Another way is to notice that at Point the curve is rising faster than a ray through the origin, which means that the elasticity must be greater than 1.0.

Notice that if the vertical food axis is multiplied by the price of food, the Engel curve tells how money expenditure on food depends on income (which is just total money expenditure).

Tor F: That the share of food expenditure in total expenditure has fallen while income has risen over the last century and a half in the United States indicates that

(setting aside changes in the price of food relative to other goods) the income elasticity of demand for food is less than 1.0.

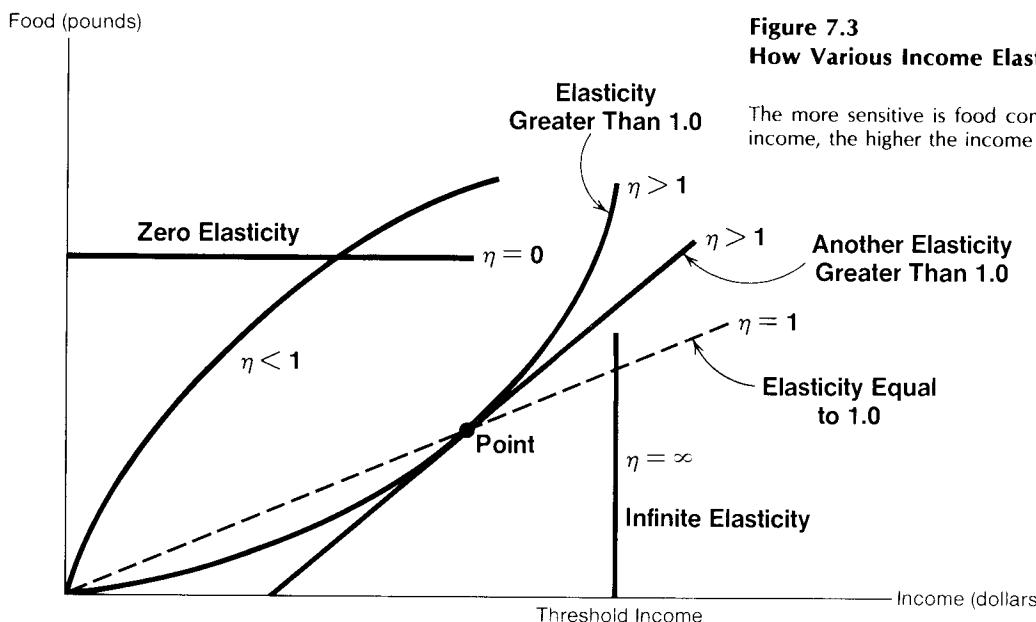


Figure 7.3
How Various Income Elasticities Look

The more sensitive is food consumption to changes in income, the higher the income elasticity of food.

A: If food expenditure rose by the same percentage as did income, then the elasticity would be 1.0. But the falling share means that in fact food expenditure rose

by a *lower* percentage than did income, which implies in turn that the elasticity is less than 1.0. This is what was to be demonstrated. So, true.

The significance of income elasticities is apparent in the low ones over the last 50 years: Nonautomotive travel, for example, had a low income elasticity, with the consequent bankrupting of rail passenger travel; rented housing also had a low income elasticity, condemning city centers densely populated with apartment buildings to decay. In both of these examples, demand was not very sensitive to changes in income.

The Price Elasticity

An elasticity can be defined for any dependent variable with respect to any independent variable. Food consumption depends on price as well as on income. Food supply also depends on price. We therefore can also speak of the sensitivity of demand and supply to changes in price.

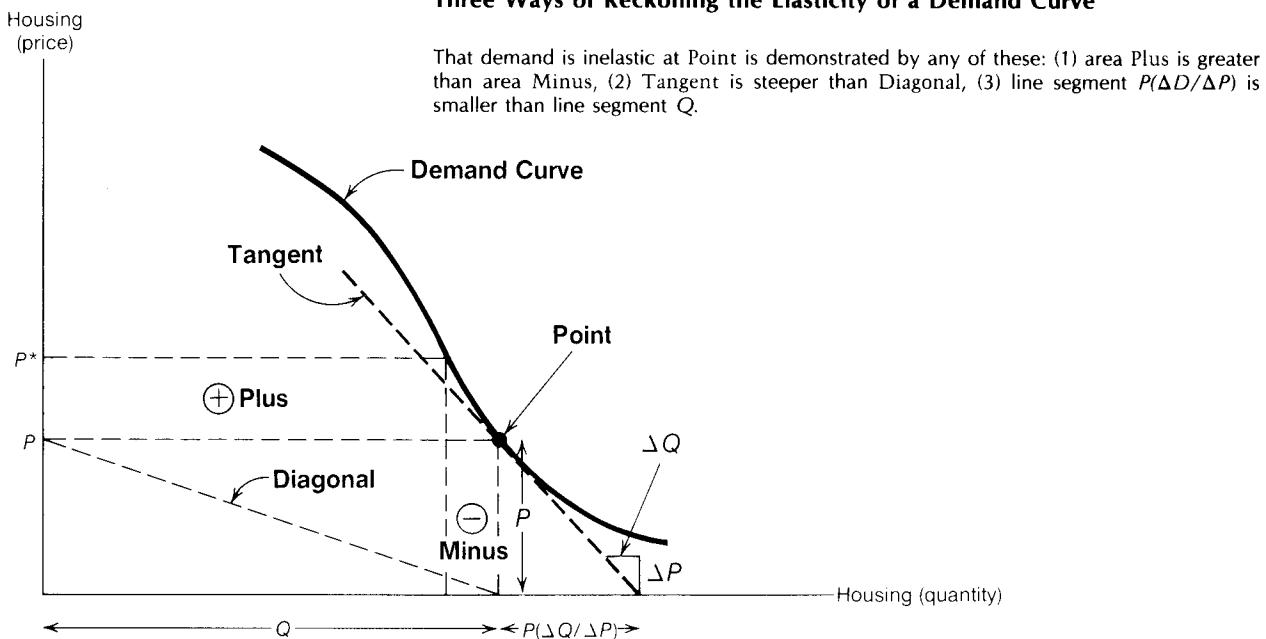
The definitions of income elasticity of Engel curves translate directly to the *price* elasticity of supply curves. The only modification necessary is to allow for the convention of putting the independent variable (price) on the vertical instead of the horizontal axis. The dependent variable (quantity supplied) is on the horizontal axis. Thus, a straight ray through the origin is a supply curve with a price elasticity of 1.0; a line horizontal at some price has infinite price elasticity; a vertical line zero elasticity; and so forth. Think it through. One essentially takes Figure 7.3 and puts "price" on the vertical axis and "quantity supplied" on the horizontal axis.

The price elasticity of demand is a little less straightforward because it is always negative. By the Law of Demand, consumption varies inversely with price, regardless of the sensitivity of demand to price changes. As the price of housing goes up by $100(\Delta P/P)\%$ the quantity demanded goes *down* by $100(\Delta Q/Q)\%$, the elasticity being $\Delta Q/Q \div \Delta P/P$.

If the percentage rise in price were the same as the percentage fall in quantity, the elasticity would be -1.0 . You can see that the total expenditure on housing (that is, PQ) would remain unchanged. That is, a 10% rise in price that causes a 10% fall in quantity will leave price multiplied by quantity at whatever it was before: $\$10 \times 100 \text{ ft}^2 = \$11 \times 90.91 \text{ ft}^2$. The special significance of an *income* elasticity of 1.0 arises from its implication of a constant share in total expenditure for the item. Likewise, the special significance of a *price* elasticity of -1.0 arises from its implication of a constant money expenditure on the item. The *unit* (-1.0) elasticity is a border between rising expenditure as the price rises (elasticity less than -1.0 in absolute terms, for example, -0.55 or zero) and falling expenditure as the price rises (elasticity greater than -1.0 in absolute terms, for example, -3.5 or $-\infty$).

The point Point in Figure 7.4 can be seen to be one of *inelastic* demand (that is, elasticity less than 1.0 in absolute terms) in three ways. First, notice that the area of rectangles fitted under the demand curve are revenues (price times quantity). The rise in price from P to P^* obviously causes the rectangle of total expenditure to rise, which implies inelastic demand. That is, the subrectangle marked with a minus sign is evidently smaller than the one marked with a plus sign, making the loss of revenue from the lower quantity less than the rise in revenue from the higher prices.

Figure 7.4
Three Ways of Reckoning the Elasticity of a Demand Curve



Notice, second, the line Tangent at Point. The height P (price) multiplied by the slope of the Tangent, $\Delta Q/\Delta P$, gives the horizontal segment marked off on the housing axis $P \times \Delta Q/\Delta P$. Divide this distance by the segment marked as Q (which is what it is: the quantity purchased at Point). The result is: $(P\Delta Q/\Delta P) \div Q$, which is $(\Delta Q/Q)/(\Delta P/P)$, which is the price elasticity at Point. Because the segment Q is here larger than the other segment, the elasticity is less than 1.0 (in absolute terms). (Notice that since the price elasticity of demand is always negative, we can for simplicity ignore the negative sign and speak "in absolute terms.") Third, notice that the definition of the price elasticity can be written in the marginal/average form as $(\Delta Q/\Delta P) \div (Q/P)$. The slope of the line Diagonal is evidently Q/P (slope here being looked at *from the price axis*), which is in the case portrayed of greater slope than the line Tangent (which has the slope $\Delta Q/\Delta P$). The elasticity is therefore in absolute terms less than 1.0. Notice that the slope of the Tangent by itself doesn't tell you everything; the slope of a curve isn't quite enough by itself to show its elasticity.

Uses of Price Elasticity

The classification of goods into low and high price elasticity follows from the notion that elasticity measures sensitivity. If the amount demanded of a good is very sensitive to price, as it would be (for example) if it had good substitutes, then the elasticity of its demand curve is high. Cigarettes as a whole have a low price elasticity, but individual brands—holding constant the price of substitute brands—have high elasticities.

The simpler applications of this idea revolve around price elasticities greater than or less than 1.0 (from now on the tiresome but important qualification "in absolute terms" will be left off):

Q: Headline: "Detroit Buses in Red, Seek Fare Hike, No Schedule Change." If they think raising fares will put them in the black, what are the managers of the Detroit bus company assuming about the price elasticity of demand for bus trips? (*Hint:* Remember that an elasticity of 1.0 is a dividing line.)

A: They are assuming that it is less than 1.0, for only in this case will the fare increase raise money revenues. Since costs do not change ("No Schedule Change"), the bus company can concentrate exclusively on the revenue side, making it big. The direction in which to change fares to make it big depends on the elasticity.

Likewise, farmers cheer or groan at a harvest blight depending on whether the demand curve facing farmers is inelastic (less than 1.0) or elastic (greater than 1.0). Since farmers in medieval Europe faced a local market in which the demand for grain was inelastic, they were overcompensated for harvest blight by the rise in price. With improved transportation over wider weather regions in early modern times, they lost this advantage, because high prices no longer went along with low crops.

The price elasticity need not be the same everywhere along a demand curve. The straight-line demand curve has all the possibilities, from 0 to ∞ elasticity, the elasticity of 1.0 (remember: -1.0) occurring halfway along it. Any of the three diagrammatic ways of exhibiting elasticities can show this. The method of diagonals of the price-quantity rectangle, for example, arrives at the result given in Figure 7.5. Notice that the line Diagonal corresponding to the point Midpoint is necessarily parallel to the demand curve itself, implying by the diagonal method that the elasticity is 1.0. At higher prices the demand curve has, viewed from the price axis, a higher slope than does the Diagonal, implying that $\Delta Q / \Delta P$ is greater than Q/P and that, therefore, the elasticity is greater than 1.0. If you compare the rectangles in the diagram, you can convince yourself that the Midpoint, $\eta = 1.0$, is the point of largest revenue.

T or F: The owner of the only bridge across a river facing a straight-line demand curve for trips will charge as high a price as possible for each trip, supposing the trips themselves cost the owner nothing.

A: False. The owner has no costs, so revenues are profits. Therefore, the owner sets the price at the point of unit elasticity, which will maximize revenue (= price times quantity), not at "as high a price as possible." The

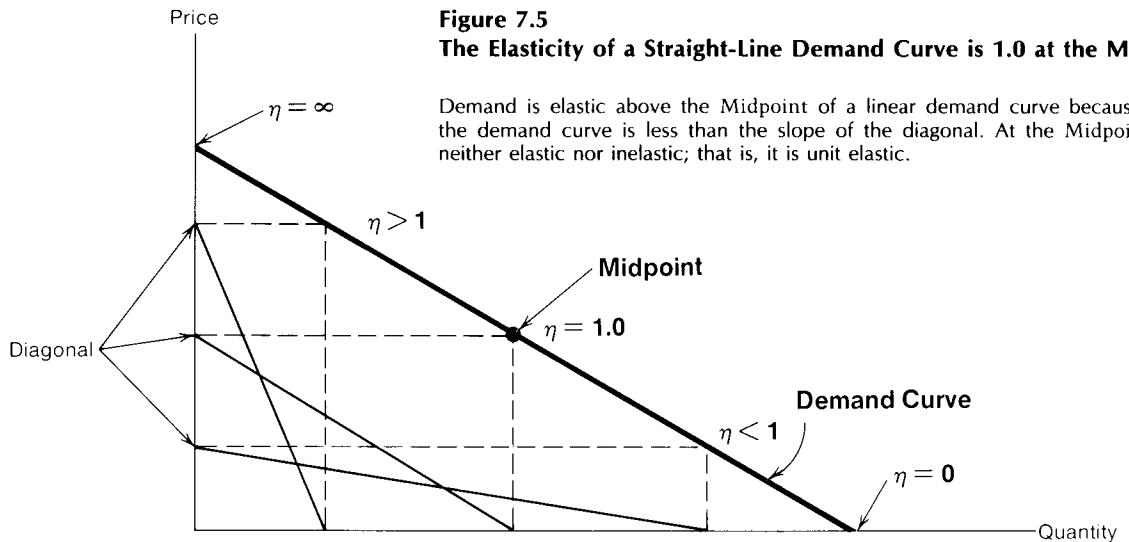


Table 7.1**Impressive Jargon for Elasticities**

Income Elasticities			Price Elasticities (defined positive)		
Good Is:	η_{HI} Is	As Income Rises The Share of H in Total Expenditure:	Demand Curve Is:	η_{HP} Is:	As Price Rises the Expenditure ($P \times H$) on H :
Normal good	>1	Rises	Perfectly elastic	$=\infty$	Falls
Luxury	$=1$	Remains	Elastic	>1	Remains
Unit elasticity		constant	Unit elasticity	$=1$	constant
Necessity	Between 1 and 0	Falls	Inelastic	<1	
Inferior good	<0		Perfectly inelastic	$=0$	Rises
			Example of Giffen paradox	<0	

man in the street thinks of a monopolist as pursuing a "high profit margin." A self-interested monopolist, however, pursues high profits *in total*, not high profits

per trip or per telephone call or per office visit. The idea of price elasticity keeps this in view.

Elasticities, then, are ways of quantifying responsiveness or sensitivity. Table 7.1 gives the jargon for income, I , and price elasticities, P , of demand for housing, H .

Summary Elasticity means "sensitivity." The elasticity of rented housing consumption with respect to real income is the percentage change in rented housing divided by the percentage change in real income causing the change in rented housing. Alternatively, it is the marginal propensity to spend on rented housing divided by the average propensity along an Engel curve. The price elasticity of demand is defined similarly (although its natural negative sign is frequently suppressed). The uses of both income and price elasticities often turn on whether they are larger or smaller than 1.0.

EXERCISES FOR SECTION 7.1

1. Here are some points on Bordo's Engel curve for wine:

Income	Wine Consumed (1-liter bottles)
\$10,000	150
10,100	153
20,000	400
20,200	404
30,000	500
30,300	501

What is his income elasticity of demand for wine in the neighborhood of \$10,000? \$20,000? \$30,000?

2. In the numbers of Exercise 1, what is the marginal propensity to consume wine at \$10,000? At \$20,000? What are the average propensities? In view of these, what are the two elasticities?

3. *True or false:* The elasticity of any straight-line supply curve going right through the origin is 1.0, no matter what its slope is, which shows that slope alone is not a measure of elasticity.

4. If the elasticity of demand facing Iowa growers of corn is 3.0, will they be happy or unhappy at a bad crop confined to Iowa? (That is, the corn crop is good in Illinois, Nebraska, Indiana, and the like, but bad in Iowa.)

5. If the elasticity of demand facing all U.S. growers of corn is 0.75, and all U.S. growers suffer a bad harvest, how will they feel (all the way to the bank)?

6. Explain in terms of the percentage changes in quantity and price why the elasticity of demand for a straight line is infinite at its topmost point (where it cuts the price axis).

PROBLEMS FOR SECTION 7.1

True or False

- **1.** If people are granted copying money in fixed amounts and if no one without a grant can use the machines, then the price elasticity of demand for copying facing the machines will be 1.0.
- **2.** The partial destruction by earthquake of a city's housing will reduce money expenditure on housing.
- **3.** The demand curves facing sellers of Kent cigarettes, an hour of unskilled labor, and a 1979 Ford Mustang serial number 368590 are all highly elastic.
- **4.** The supply curves facing Gerry Gunderson, ordinary consumer, are perfectly inelastic.

7.2 Constant Elasticities

What to Read For

What does a formula for a demand curve look like that has unchanging elasticity throughout its length? What is the relationship between elasticities and the exponents in multiplicative functions? What is the significance of the rate of change transformation in thinking about curves with constant elasticities? If you know the elasticities of a market's supply and demand curves, can you infer the elasticities facing individual suppliers or demanders? What is the procedure, exactly?

Exponents Are Elasticities

A demand curve in the form of a straight line, such as $Q = 50 - 0.5P$, has varying elasticity at various different points—1.0 at the midpoint, higher at higher prices, lower at lower. Since little else is known about a demand curve than what roughly its elasticity might be, you will often want to assume that the elasticity is not varying but constant. The general form of a demand function with a constant price elasticity is $Q_{\text{DEMAND}} = DP^{-\eta}$, in which D is a constant serving to position the demand curve and the exponent η is the (absolute value of the) elasticity of quantity demanded with respect to price.

It is not instantly obvious that constant elasticity implies this exponential form.² The special case of unit elasticity will make it plausible: When $\eta = 1.0$, $Q = DP^{-1} = D/P$, or $QP = D$; that is, with unit and constant elasticity the total expenditure for any price is a constant, D , which is the assertion just proven in a problem. Another particular case worked out in detail makes it still more plausible. For example, if the demand curve for gallons of gasoline per week is $Q = 7.75P^{-0.5}$, how much does gasoline consumption fall if the price of \$1.60 a gallon is raised 10%? The otherwise useless square root key on your pocket calculator comes into play: $7.75/1.60^{0.5} = 6.13$ gallons a week at \$1.60 a gallon falls to $7.75/1.76^{0.5} = 5.84$ gallons when the price is \$1.76 a gallon. That is, a 10% rise in price has caused a (roughly) 5% fall in quantity because $(6.13 - 5.84)/6.13 = 0.05$ roughly, just as the elasticity of -0.5 would imply: 5% is 0.5 of 10%.

A more complete demand curve for, say, apartments in Boston would include terms for consumers' income, Y , and for the price of substitutes such as houses, P_H :

$$Q_A = 65.3P_A^{-2.0}P_H^{1.2}Y^{0.8}$$

You should get into the habit of interpreting the exponents in multiplicative functions of this sort as elasticities.

T or F: If the demand function for apartments is indeed $65.3P_A^{-2.0}P_H^{1.2}Y^{0.8}$, then a 10% rise in income and all prices will have no effect on the quantity demanded of apartments.

A: True. The exponents are elasticities (that is, percentage changes in the quantity demanded of apartments for each percentage change in the variables). Thus, a 10% rise in P_A causes a $(-2.0)(10\%) = 20\%$ fall in

Q_A ; a 10% rise in P_H causes a $(1.5)(10\%) = 15\%$ rise in Q_A ; and a 10% rise in Y causes a $(0.5)(10\%) = 5\%$ rise in Q_A . The 20% fall just offsets the 15% + 5% = 20% rise. Notice that the result is familiar. The demand function here satisfies as it should the condition that a doubling of money income and all prices—which does not alter the budget line—must not alter the consumer's equilibrium.

Exponents in multiplicative functions, then, are elasticities. Conversely, constant elasticities can always be expressed in multiplicative form.

T or F: Jeffrey Williamson, with an income of \$5000 per year, spends 10% of his income on saving, for which his income elasticity of demand is 2; therefore, if his income rises to \$6000 per year, he will spend 12% of his income on saving.

A: True. An answer requiring nothing more than arithmetic is as follows. He spends 10% of his income, or 500 units of savings at \$1 a unit, initially. His income goes up by $(\$1000/\$5500) = 18\%$, taking the midpoint between \$5000 and \$6000 as the base (this choice of

² It can be proven with second-year calculus; see, for example, R. G. D. Allen, *Mathematical Analysis for Economists* (London: Macmillan, 1960), p. 418. And first-year calculus suffices to show the converse, that is, that the exponential form implies constant elasticity, as follows. The definition of elasticity in its derivative form is $P dQ/Q dP$. For the demand function $Q = DP^{-\eta}$, the derivative with respect to P , dQ/dP , is $-\eta DP^{-\eta-1}$. Multiply this by P and divide it by Q to arrive at an expression for the elasticity:

$$\frac{P dQ}{Q dP} = \frac{P}{Q} (-\eta DP^{-\eta-1}) = \frac{P}{DP^{-\eta}} (-\eta DP^{-\eta-1})$$

Notice that when the expression for Q is inserted, everything cancels out except $-\eta$. That is, the elasticity is the constant $-\eta$. Q.E.D. You'll find this section tough if you are not enrolled in a calculus course. Try to read around the math to see the essential points, especially the points toward the end.

base is arbitrary: \$5000 or \$6000 would be good, too). His income elasticity is 2, so the percentage increase in his consumption of saving will be $2(18\%) = 36\%$. So he will now consume $(1.36)(500) = 680$ units of

saving. Taking the midpoint again, he will devote $\$680/\$5500 = 12.36\%$ of his income to saving, which is 12%, close enough.

COMMENT To come to the conclusion that the assertion is false because the result of such a calculation is not 12% exactly is to let arithmetic dominate reason. Depending on the (arbitrary) choice of base for the calculations of the percentage rates of change, one can arrive at a variety of results centered around 12%.

Alternative Solution

A neater solution exploits the implicit assumption that the income elasticity is constant, that is, that Williamson's consumption of saving is $S = kY^2$, a multiplicative form. The scaling constant, k , can be determined by solving for it when $S = \$500$ and $Y = \$5000$ (the initial values). In these units $k = 0.00002$, so $S = (0.00002)Y^2$, and the result that the new S will be \$720 (exactly 12% of 6000) can be inferred by setting $Y = \$6000$ and solving for S .

The Rate of Change of a Constant Elasticity Expression Is Linear

An important point about constant elasticity functions such as $Q_A = DP_A^{-2.0}P_H^{1.2}Y^{0.8}$ or $S = kY^2$ is that they are "linear in logarithms." That is, if you take the logarithm of both sides of the saving function, say, you get $\log S = \log k + 2 \log Y$. If you know that on logarithmic graph paper a plot of something growing at a constant percentage rate is a straight line, it will come as no surprise that the "rate of change transformation" of $S = kY^2$ has the same linear form, namely, $S^* = k^* + 2Y^*$, in which an asterisk (*) signifies "rate of change of." Another way of seeing that this is true is to divide the last expression by Y^* , holding k constant (i.e., setting the rate of change k equal to zero). The result on the left-hand side is S^*/Y^* , which is the definition of the income elasticity of saving with respect to income. The result on the right-hand side is 2, which you know to be the income elasticity. Therefore, the original equation of rates of change was a true one.

The point is that constant elasticity functions can be made into linear equations by taking rates of change. This makes them very useful because expressible in the simplest mathematical form. For example, a popular choice of a functional form for the demand for money balances is

$$M_d = \frac{kPy^\alpha}{i^\beta}$$

because it is easy to express it in a linear form suitable for statistical manipulation:

$$M_d^* = P^* + \alpha y^* - \beta i^*$$

where y is real income, P the price level, and i the interest rate. The numbers α and β are elasticities of the quantity of money demanded with respect to real income and the interest rate (for example, $\alpha = M_d^*/y^*$ holding P and i constant). Note that k drops out: The rate of change of a constant is nothing.

The applications of these ideas to real problems depend on a number of simple but important facts about rates of changes, some of which have been used already. The algebra to prove them, which appear in the appendix to this section, is routine. (Some of these facts are accurate only for small rates of change, for 10% changes but not for 100% changes.)

These facts can be illustrated using variables to which they may apply, such as price (P), quantity (Q), and quantity demanded by an individual demander (q). $(P^{-\eta})^* = -\eta P^*$ for η constant; $(PQ)^* = P^* + Q^*$; $(C/Q)^* = C^* - Q^*$; and, finally,

$$(q_1 \pm q_2)^* = \left(\frac{q_1}{q_1 \pm q_2} \right) q_1^* \pm \left(\frac{q_2}{q_1 \pm q_2} \right) q_2^*$$

The last fact serves to establish the obvious proposition that the elasticity of the market demand curve at any price is equal to a weighted average of the elasticities of demand of the individual demanders. Begin with the assertion that the total quantity demanded at any price is the sum of the demands by, say, the two individuals in the market: $Q = q_1 + q_2$. The rate of change of this sum will be, then,

$$Q^* = \left(\frac{q_1}{Q} \right) q_1^* + \left(\frac{q_2}{Q} \right) q_2^*$$

Now price elasticities have the form Q^*/P^* , that is, the percentage change in quantity demanded divided by the percentage change in price. In the search for an expression in elasticities, therefore, divide both sides by P^* (the percentage change in the one market facing each individual) to get

$$\frac{Q^*}{P^*} = \left(\frac{q_1}{Q} \right) \frac{q_1^*}{P^*} + \left(\frac{q_2}{Q} \right) \frac{q_2^*}{P^*}$$

If the individual rates of change, q_1^* and q_2^* , are interpreted as the rates of change caused by the change in price, then the equation is

$$\eta_{QP} = \left(\frac{q_1}{Q} \right) \eta_{q_1 P} + \left(\frac{q_2}{Q} \right) \eta_{q_2 P}$$

That is, as asserted, the elasticity is a weighted average.

Q: The elasticity of the demand for slaves in southern cities was fairly high, that is, 0.86 (there were good substitutes for slaves in city occupations); the total elasticity of demand for slaves, in city and countryside together, was 0.08; 96% of slaves lived in the countryside. *True or false:* Therefore, the elasticity of demand for slaves in the countryside was very low, that is, 0.48 (there were no good substitutes for slaves in country occupations).

A: City and country can be viewed as two demanders. According to the formula just derived, the elasticity is η in

$$0.08 = (0.96)(\eta) + (0.04)(0.86)$$

Solving gives $\eta = 0.048$.³ Therefore, true.

Simple though it is, then, the formula gives useful meaning to the “importance” of a particular demander. But think before applying it:

T or F: If the elasticity of demand for each of several brands of cigarettes taken separately is 10, then the elasticity of demand for the several brands taken as a group is also 10.

A: False: It is cruel to deceive you, but the deception makes the point that common sense must dominate mathematical formalism. The formula just given refers to the demand of demanders of cigarettes, not, as in this equation, the demand *facing* suppliers of one brand. Plainly, if the price of Marlboros rises with other prices constant (which is what the elasticity of demand for Marlboros “taken separately” must mean), people will substitute other cigarettes (whose price has not changed) for Marlboros. On the other hand, if the prices of all cigarettes move up simultaneously (which is what is relevant for the “elasticity of demand for the several brands taken as a group”), there will be less or no inter-brand substitution. Evidently, this latter elasticity of demand must be much smaller than 10 (say, 0.10).

³ See Claudia D. Goldin, *Urban Slavery in the American South 1820–1860: A Quantitative History* (Chicago: University of Chicago Press, 1976), p. 101.

The Elasticity of Demand Facing One Seller

The formula relevant to the last problem and to others like it is less simple than the first. By way of introduction, consider the situation of a single large firm in the market for, say, coal, and suppose that you wish to know the elasticity of demand facing the firm. One reason for wishing to know it, for example, is that the elasticity is a reasonable measure of the firm's *market power*. If the elasticity is very high, the firm has little power to raise the price it faces; if it is very low, the firm has much power. In one sense, any firm has the power to raise its price. The owner of a coal mine is free to raise the price the firm charges the wholesaler to \$1000 a ton, although the owner will find that the quantity he sells at such a price is zero. The point is that a small firm with little market power cannot raise the price it charges without losing all its business. In other words, a firm has "power" when it faces a less than complete fall in quantity demanded when it raises the price. Power is facing a less than perfectly elastic demanded curve.

The question is, then, what is the elasticity of demand facing a single coal-selling firm? The answer begins with a simple equilibrium condition, namely, that the quantity demanded from the firm equals the total quantity demanded in the market minus the amount supplied by other firms:

$$q_i = Q_D - Q_S$$

Applying the formula for the rate of change of a sum or difference, the rate of change of this is

$$q_i^* = \left(\frac{Q_D}{q_i} \right) Q_D^* - \left(\frac{Q_S}{q_i} \right) Q_S^*$$

Now suppose that the firm in question—Peabody Coal, say—raises the price of coal. The other firms could behave in either of two ways: either keep their own price unchanged or match Peabody's price. If they keep the price unchanged, Peabody will lose all its business (nobody will buy coal at \$16 a ton from Peabody when they can get it for \$15 elsewhere), and without recourse to the formula, it is obvious that the elasticity facing Peabody is infinite. If the other firms match Peabody's price, however, Peabody will lose some but not necessarily all its business, and the formula comes into play. Since the other firms must in fact get a higher price to be enticed to serve Peabody's former customers, the price-matching behavior is the natural one. The new, higher price of coal established on Peabody's initiative decreases the quantity demanded by consumers and increases the quantity supplied by other producers. How much? By the amount of the elasticities of demand $\eta = Q_D^*/P^*$, and of supply, $\epsilon = Q_S^*/P^*$. The algebra comes to the same conclusion. Dividing the formula just derived for the rate of change of Peabody's quantity demanded, q_i^* , by the rate of change of the price, P^* leads to

$$\eta^i = \left(\frac{Q_D}{q_i} \right) \eta - \left(\frac{Q_S}{q_i} \right) \epsilon$$

If the elasticity of demand—which is naturally negative—is defined to be positive, then the formula is

$$\eta^i = \left(\frac{Q_D}{q_i} \right) \eta + \left(\frac{Q_S}{q_i} \right) \epsilon$$

The formula expresses an important economic idea, namely, that for given market elasticities (η and ϵ) the elasticity of demand facing one supplier is higher the higher is the output of other suppliers relative to the one in question.

T or F: If Peabody Coal supplies 20% of the coal purchased, if the elasticity of supply of other coal sellers is 1.0, and if the elasticity of demand for coal is 0.5, then Peabody Coal faces an elasticity of demand of 6.5.

A: Look at the formula, inserting the numbers given:

$$\eta^i = \left(\frac{1.0}{0.2}\right)(0.5) + \left(\frac{0.8}{0.2}\right)(1.0) = 6.5$$

In the absence of a conspiracy among coal suppliers, this arithmetic is reason for being skeptical of assertions that even quite large companies can "administer" prices. Peabody Coal might by itself raise prices 10%, but it would have to be willing to suffer a consequent loss of $(10)(6.5) = 65\%$ of its business. True.

Writing Down Elasticities of Excess Supply or Demand

The general task in such problems is to find unknown elasticities from known ones and market shares. The technique of solution takes rates of change of the equilibrium condition that supply equal demand. You will find it useful to learn how to leap directly from the equilibrium condition to the elasticity equation, that is, from $q_i = Q_D - Q_S$ to $\eta^i = (Q_D/q_i)\eta + (Q_S/q_i)\epsilon$. When you have made the leap, check the place you have landed by making sure that the signs in the equation make sense, as they do in the coal problem. If the elasticity of demand is defined positively, the equation says that rises in the elasticity of both demand and supply raise the elasticity facing the residual supplier, which is as it should be, because these raised elasticities both signify a raised amount of demand left to Peabody Coal at a higher price. This, in turn, is in accord with the diagrams of excess supply and demand (see Chapter 6).

Q: The elasticity of demand for grain is generally thought to be low among poor peasants, say, 0.25 numerically. The elasticity of total supply of grain from poor peasants is also thought to be low, say, 0.10. The peasants are said to be set in their ways, responding weakly to the incentive of price. Furthermore, the peasants are often only slightly involved in the outside market, exporting only 5%, for example, of their total supply. *True or false:* If all this is true, a 1% increase in the price of grain will result nonetheless in a nearly 7% increase in the amount exported, making the peasants look, to

the careless eye, very responsive to the incentive of price.

A: *True.* The supply to the rest of the world by the peasants must equal the supply in total produced by the peasants minus the demand by them for their own grain. Skipping to the figures,

$$\epsilon_{\text{TO REST OF THE WORLD}} = \frac{20}{1} (0.10) + \frac{19}{1} (0.25) = 6.75$$

In this sense the peasants are indeed responsive to price, but not in their underlying attitudes.

Such are the practical uses of elasticities of excess supply and demand. A theoretical use can cap the argument. It is that small demanders and suppliers face given prices of supply and demand. Richard Weisskoff, a single buyer of an apartment in Cambridge, for example, might take 1% of the whole supply in a month. The elasticity of supply he faces, therefore, can be derived from the identity $S_{\text{TO WEISSKOFF}} = S_{\text{TOTAL}} - D_{\text{OTHERS}}$. It is $\epsilon_w = (100/1)\epsilon_T + (99/1)\eta_o$. The elasticities of supply and demand could be very low—0.05 and 0.2, say—yet still lead to a very high elasticity facing Weisskoff: $(100)(0.05) + 99(0.2) = 24.8$. In other words, if Weisskoff raised his offer per room as little as 1%, he would be supplied with about 25% more rooms than he could buy, if he wanted them, at 1% above the old price. If he lowered his offer 1%, he would be supplied

with 25% fewer rooms. He faces a virtually flat supply curve, a given price. This is the algebraic version of the geometry of price taking. A single small buyer among many other buyers and a single small seller among many other sellers face a price over which they have no influence. The smaller the share, the closer is the supply or demand curve facing the buyer or seller to being utterly flat.

Summary

Constant elasticity is the simplest assumption consistent with a little knowledge about the shape of a curve. Its mathematical form is especially simple: for example, $Q = SP^\epsilon$, which could be a supply curve with price elasticity ϵ and location S . Such forms are easy to manipulate, because their logarithms or rates of change are linear: For example, the rate of change Q^* equals $S^* + \epsilon P^*$. The algebra of rates of change—and in particular the rate of change of a sum—leads to a way of connecting known and unknown elasticities. The technique is to write down the equilibrium condition in quantities (quantity supplied equals quantity demanded), take the rate of change of this expression, and divide by the (market-wide) rate of change of price. Since a price elasticity is a rate of change of a quantity divided by the rate of change of the price causing the change in quantity, the resulting equation is in elasticities. The technique is widely applicable to problems of excess supply and demand, such as, at the most general level, the problem of the elasticity of supply facing a single demander with a small share of the market. The elasticity of supply facing the demander is higher the smaller is his share, providing an algebraic proof of the geometry of price taking.

APPENDIX TO SECTION 7.2: The Algebra of Rates of Change

The rate of change of Q is the change in Q divided by Q in some base period, taken here (arbitrarily) to be its initial value. (When Q rises from 100 to 110, the $\Delta Q/Q$ [signified by Q^*] is $10/100 = 10\%$ instead of $10/110 = 9.1\%$.) The method of deriving the truths about rates of change of exponential, multiplicative, and summed expressions is illustrated best with the rate of change of PQ :

$$\begin{aligned}(PQ)^* &= \frac{\Delta(PQ)}{PQ} = \frac{(P + \Delta P)(Q + \Delta Q) - PQ}{PQ} \\ &= \frac{PQ + P\Delta Q + Q\Delta P + \Delta P\Delta Q - PQ}{PQ} \\ &= \frac{P\Delta Q}{PQ} + \frac{Q\Delta P}{PQ} + \frac{\Delta P\Delta Q}{PQ} \\ &= \frac{\Delta Q}{Q} + \frac{\Delta P}{P} + \left(\frac{\Delta P}{P}\right)\left(\frac{\Delta Q}{Q}\right) = P^* + Q^* + P^*Q^* \\ &\cong P^* + Q^*\end{aligned}$$

when P^* and Q^* are small, because when they are small their product, P^*Q^* , is very small. Thus, the rate of change reduces multiplication to addition (and division to subtraction). In the same way, it reduces exponents to multiplication, and addition to weighted averaging. For example, the exponential expression Q^n has a rate of change

$$(Q^\eta)^* = \frac{(Q + \Delta Q)^\eta - Q^\eta}{\Delta Q} = \left(\frac{Q + \Delta Q}{Q} \right)^\eta - 1$$

$$= (1 + Q^*)^\eta - 1$$

The expression $(1 + Q^*)^\eta$ is similar to the expression for the value of a dollar placed at compound interest at $i\%$ for N years, $(1 + i)^N$. For small rates, compound interest is about the same as simple interest, which is $1 + Ni$. That is, for small rates of change, Q^* , the expression $(1 + Q^*)^\eta - 1$ approximately equals $1 + \eta Q^* - 1 = \eta Q^*$. That is, for small Q^* , $(Q^\eta)^* = \eta Q^*$, as was to be shown.

Deriving the rate of change of a sum (or difference) is even simpler:

$$(q_1 + q_2)^* = \frac{(q_1 + \Delta q_1 + q_2 + \Delta q_2) - (q_1 + q_2)}{\Delta q_1 + \Delta q_2}$$

$$= \frac{\Delta q_1 + \Delta q_2}{q_1 + q_2} = \frac{\Delta q_1}{q_1 + q_2} + \frac{\Delta q_2}{q_1 + q_2}$$

Dividing top and bottom of the first term in this last expression by q_1 and similarly for the second term gives

$$(q_1 + q_2)^* = \left(\frac{q_1}{q_1 + q_2} \right) \left(\frac{\Delta q_1}{q_1} \right) + \left(\frac{q_2}{q_1 + q_2} \right) \left(\frac{\Delta q_2}{q_2} \right)$$

$$= \left(\frac{q_1}{q_1 + q_2} \right) q_1^* + \left(\frac{q_2}{q_1 + q_2} \right) q_2^*$$

which was to be demonstrated. Notice that this result, unlike the others, is exact, whatever the size of q_1^* and q_2^* .

Notice, too, that all the results are derivable with the calculus, using in particular the idea of total differentials. Thus, for the exponential function Q^η ,

$$\text{Rate of Change} = \frac{dQ^\eta}{Q^\eta} = \eta \frac{Q^{\eta-1} dQ}{Q^\eta} = \eta \frac{dQ}{Q} = \eta Q^*$$

The more primitive method of Δ 's is worth retaining, though, because it keeps the exact results for large changes (instead of infinitesimal changes, dQ) in view when they are needed.

EXERCISES FOR SECTION 7.2

1. If the demand curve for apartments is $Q_A = 65.3 P_A^{-2.0} P_H^{1.5} Y^{0.5}$, how much will the quantity demanded, Q_A , rise if the price of apartments, P_A , falls 10%? If the price of houses that substitute for apartments, P_H , falls 1%? If the income of the consumers rises 5%? If all happen at the same time? *IF all variables rise 10%?*

2. Put the exponential expressions into linear, rate-of-change form; and put the linear expressions into exponential form (use asterisks to signify rates of change):

- a. $Q_A = 65.3 P_A^{-2.0} P_H^{1.5} Y^{0.5}$
- b. $Q = kP^{-N}$
- c. Percentage change in wine purchased = percentage change of income.
- d. The percentage change in wine purchases is always twice the percentage change of income.
- e. $Q = kP^{-1}$

3. Suppose that the poor have a price elasticity of demand for housing of 0.25 (expressed positively) and now own half the housing; the rich have an elasticity of 0.75 and own the other half. What is the elasticity of the whole demand for housing?

4. If Peabody as described in the text supplied 1% of the coal, what elasticity of demand would it face? If it had 100%?

PROBLEMS FOR SECTION 7.2

1. Suppose that you wanted to find how much of the 15% or so per year growth of cotton cloth production in the United States 1815 to 1833 was attributable to technological change (the coming of the power loom and the factory) and how much to demand (the rise of population and incomes and the transport improvements that made the demand of the West available to the factories of the East). The problem is evidently one of supply and demand. Suppose that the demand curve can be represented as $Q = DP^{-\eta}$, in which a rising D represents an outward-moving demand. The supply curve can be represented as $P = (1/T)Q^{1/\epsilon}$, in which a rising T represents a falling price at which cloth producers would offer a given quantity—this being one natural way of measuring technological change. (The peculiar inversion of the elasticity, ϵ , is merely a ploy to make ϵ be the elasticity of supply, even though here the price and not the quantity is taken to be the dependent variable.) Suppose, finally, that the elasticity of demand is about 2.5, the elasticity of supply about 20, the rate of change of demand, D^* , about 10% per year, and the rate of change of technology 2.5% per year. Solve the problem by stating the demand and supply functions in rate-of-change form, then solving the resulting linear equations for Q^* .

2. The balance of payments is the annual net flow of money out of (deficit) or into (surplus) a country. In other words, according to the "monetary approach to the balance of payments," a country's balance of payments is conveniently viewed simply as the excess supply of money (deficit) or the excess demand for money (surplus) by citizens of the country during a year. Suppose that there are two countries in the world, the United States of Great Britain (USGB) and the Federal Republic of Nippon (FRN), having the same form of their demand curves for money, namely,

$$M_D^i = k_i P y_i^\epsilon$$

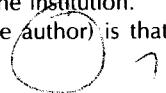
where i names the country, M_D^i is the quantity of money demanded, k_i is a constant pertaining to the i th country, P is the world price level (assumed to be the same as each country's price level by the working of arbitrage in commodity markets), y_i is the i th country's real income, and ϵ is the income elasticity of the demand for money.

a. If the world's supply of money is fixed and the world starts from an equilibrium distribution of the supply, show that if the USGB grows slower than the FRN (but both are growing some), P will fall, the USGB will have a persistent deficit in its balance of payments, and the FRN will have a persistent surplus.

b. Show that, if the conditions of (a) hold, except that each year the USGB produces new money (the only new money produced in the world) in an amount that raises the world stock faster than the rate of growth of the demand for money, then there will be world inflation, and the surpluses and deficits in the annual balance of payments in the USGB and the FRN will be larger than in (a).

- 3. "Wherever American slavery touched urban conditions, it was in deep trouble. The free atmosphere and cosmopolitan exposure of cities made slaves increasingly difficult to control, and the economic life of the cities, with their growing factories staffed largely by white immigrant laborers, became increasingly poorly suited to the institution."

The keystone of the evidence for this view (developed by the author) is that between



1820 and 1860 the total slave population grew by 180%, while the slave population of the ten largest southern cities grew by only 90%.

a. Write down the total demand curve (city and country combined) for slaves in terms of the price, P , the shift terms in urban, D_u , and rural, D_r , areas, and the elasticity of demand in urban, η_u , and rural, η_r , areas. If the total quantity of slaves, Q , was inelastically supplied (as it was), write down the equilibrium condition.

b. Explain why the availability of white immigrant labor in the cities and the lack of substitutes for slave labor in the countryside (free men would not tolerate the gang labor so efficient in cotton production) suggests that in absolute values the urban elasticity, η_u , was greater than the rural elasticity, η_r .

c. The hypothesis that cities were increasingly hostile to slavery can be interpreted as the assertion that the rate of change of D_u was less than that of D_r . Show that, if the slave price were rising, the same observation of slower growth of slave population in the cities than in the countryside could be explained by $\eta_u > \eta_r$ and that, if the price were falling, this pattern of elasticities would imply a faster growth of the population in the cities.

d. If the relative decline of urban slavery could be explained by the rates of change of D_u and D_r , one would expect the decline to be the same from one decade to the next. It was not. The urban slave population oscillated violently. Show that the data in Table 7.2 for St. Louis (duplicated less dramatically in other slave cities) are consistent with the alternative explanation of the relative decline of slavery in the cities, namely, $\eta_u > \eta_r$ in absolute values.

4. The supply curve of slaves in the American South from 1830 to 1860 was perfectly inelastic to price (the slave trade from Africa had been closed) but was moving out at 2.3% per year (the growth rate of the slave population). The elasticity of demand for slaves was very low, around 0.08. The price of slaves rose about 2.7% per year. *True or false:* If all these assertions are true, then the demand curve for slaves 1830–1860 was growing a little faster than was the supply curve (but only a little faster).

True or False

- 5. The only supply curve with a constant elasticity of 1.0 is a straight line through the origin at a 45° angle.
- 6. If Brazil supplies a third of the world's demand for coffee and if the elasticity of demand for coffee is 0.1, then the elasticity of demand facing Brazil is at least 0.3.
- 7. If the United States consumes a fifth of the world's oil production and if the elasticities

Table 7.2
Changes in Slaves and Prices, St. Louis, 1820–1860

Dates Compared (census years)	Rate of Change (%) of Slave Labor Force in St. Louis	Rate of Change (%) of Slave Prices in St. Louis
1820–1830	+55	−3
1830–1840	−41	+32
1840–1850	+79	−25
1850–1860	−47	+84

Source: Claudia A. Goldin, *Urban Slavery in the American South 1820–1860: A Quantitative History* (Chicago: University of Chicago Press, 1976), Table 25, p. 83, and Table 23, p. 70.

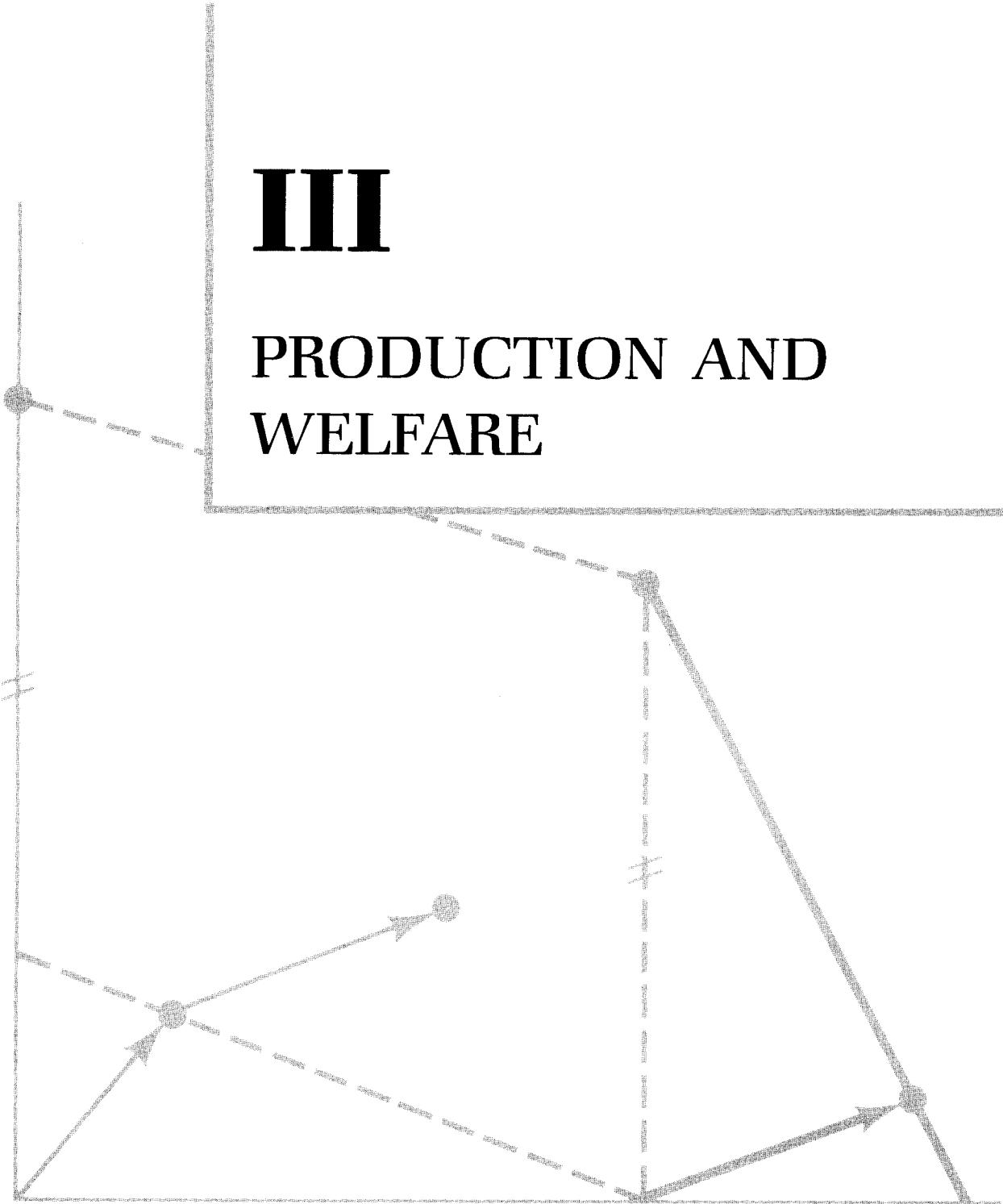
of world supply and of non-American demand are both around 0.3, then the elasticity of the supply curve facing American consumers is 2.7.

- o 8. If the price elasticity of demand for coal is 2.0 and the price elasticity of demand for all fossil fuels (coal, oil, natural gas) is 1.0, then the price elasticity of demand for noncoal fossil fuels must be less than unity.
- o 9. If steel is produced only in Pittsburgh, if the price of steel at Pittsburgh is not affected by the quantity of steel demanded (horizontal supply curve), and if railway services are used in fixed proportion to the quantity of steel shipped to all markets as a whole, then the elasticity of demand for railway services in the shipping of steel with respect to the price of the railway services is equal to the elasticity of demand for steel multiplied by the ratio of costs of railway services used to ship steel to the final market price of steel.
(Hint: Work through step by step the effects of a 1% increase in the price of transport on the demand for transport and thence on the demand for transport itself.)
- 10. If the cost of transporting ships is a negligible percentage of their price and if the output of Japanese shipyards is growing faster than the output of German shipyards, then the German supply curve is moving out faster than the Japanese.
- 11. For a certain product in a certain locale, the lower the ratio of the elasticity of demand to the elasticity of supply (ignoring signs), the higher the ratio of local consumption to local production. Illustrate the truth or falsity of this proposition with automobiles, housing, and wheat in both the producing and consuming locales.
- 12. Other things equal, one would expect the elasticity of supply of food to the world market to be *lower* in a region with mostly subsistence farmers (who eat a lot of their own food) than in a region with mostly commercial farmers (who do not).

steel

III

PRODUCTION AND WELFARE



8

Production Possibilities

SIT DOWN
 It's real opportunity cost
 that matters.
 Problem → P. 157
 End with efficiency
 of selling elections.
 What is real cost
 of stealing?

8.1 Production and Specialization

What to Read For **What is the specialization theorem? How is opportunity lost related to specialization? What is the division of labor? Comparative advantage? How is free trade related to specialization? Why is free trade good? Why does a society's production possibility curve bulge outward? Why does its supply curve of, say, food slope upward?**

Look at the True Opportunity Cost

The theory of demand, just completed, asks how people behave given a budget line; the theory of supply, now begun, asks how the budget line came to be given. In the theory of demand, the premise that there is no such thing as a free lunch is imposed from the outside; in the theory of supply, it is the result.

The central idea in the theory of supply is opportunity cost, namely, that supplying one thing costs *not* supplying another thing. Opportunity cost is merely a restatement of scarcity or, in other words, of the budget line. But a close attention to opportunity cost characterizes the most economic of economic arguments. Few noneconomists, for example, would take the following proposition seriously:

T or F: From the social point of view, and given equal power to deter crime, fines are better than prison sentences.

A: Fines merely transfer money from the criminal to the state (or to the victim). They do not reduce the society's ability to consume. Prison sentences, by contrast, do reduce the society's ability to consume: The guards, the prison walls, and the time of the prisoners themselves are diverted from employment in which they

could be making other things. A guard on a prison wall could be employed as a farmer or a traffic cop. He would be so employed under a system of fines, and society would be getting his output. Social output would be higher. Therefore, true. Note, by the way, that fines are not necessarily less severe or less crime deterring than is prison. A \$15,000 fine might deter the same amount of armed robbery as a two-year sentence in prison. If it does not, it can be raised to \$20,000 or \$30,000 or whatever does the job.

The Specialization Theorem: Make the Opportunity Cost Smallest

The most important application of opportunity cost is the *specialization theorem*, namely, that production of all goods can be larger if people are permitted to specialize in production and to trade their products. Free trade, as was shown earlier, leads to a best point of consumption inside a given Edgeworth box; the specialization theorem asserts that free trade leads also to the largest possible Edgeworth box. A federated United States, a European Common Market, a world without import tariffs, or an economy without ghettos, segregation, apartheid, passports, entry restrictions, or other barriers to trade will produce more.

The reason is simple. If a baker cannot trade with a carpenter, the baker must build her own house. Each room the baker builds will have a high opportunity cost in loaves forgone. The baker's hour produces many loaves but few rooms, and to build a room therefore results in the sacrifice of many loaves of bread. The carpenter's bread baking, likewise, will have a high opportunity cost in rooms unbuilt. If they can trade, however, each can specialize in what each one does best—to use the jargon, in his or her *comparative advantage*. Comparative advantage matters because the decision to specialize depends on what you do best compared with the other things you could do. The carpenter may be in fact a better baker than the baker, but as long as his hour of baking has a higher cost in forgone housing than does the baker's, the carpenter will continue to specialize in carpentry. After each has produced his or her own good in the best way (that is, at the lowest cost in output of other things forgone), each can consume both goods by trading for the other.

When this argument is applied to people, it is called the *division of labor*; when it is applied to nations, it is called *specialization by comparative advantage*. It applies to all levels of society, from the family to international trade. If a family is to produce the largest output, its members must specialize by comparative advantage. John is quicker at chopping a pile of wood than is little Bobby, his son. He has an *absolute advantage*. Does that mean John should do everything? No. If the opportunity cost of John's chopping is starting the furnace and replacing a tire on the car, while the opportunity cost of Bobby's is only playing marbles and sweeping the back stairs, it is best to assign Bobby to the woodpile. This division of labor will produce the chopped wood with the least valued sacrifice of other goods (started furnaces, played marbles, and so forth).

Proof of the Theorem in the Case of Foreign Trade

A diagram makes the argument clear. It shows how to connect individual and social opportunity costs. To take a historically important example, consider the possible combinations of food and manufactures that the Old World and the New World could produce in the nineteenth century (see Figure 8.1).

The possible combinations are bordered by a social budget line, called, you remember, the production possibility curve. Its slope is here constant (a *straight-line* production possibility curve), representing constant marginal opportunity cost. The Old World, for example, sacrifices the same amount of food for each additional unit of manufactures produced.

Were trade prohibited between the Old and New Worlds, each would pick by the usual indifference contours some point such as E or E' along each curve. At E and E' both the Old World and the New World are producing (in order to consume) both goods. Were trade allowed, however, the two worlds would be in effect merged into one. The production possibility curve for the merged

① Suppose we're selected on planet Earth thru the 4th dimension [attack ins. he does it better]

Figure 8.1
The Production Possibility Curves of the Old (a) and New (b) Worlds

The Old World produces manufactures at a lower opportunity cost in terms of food foregone than does the New World. The cost is the slope of the production possibility curve.

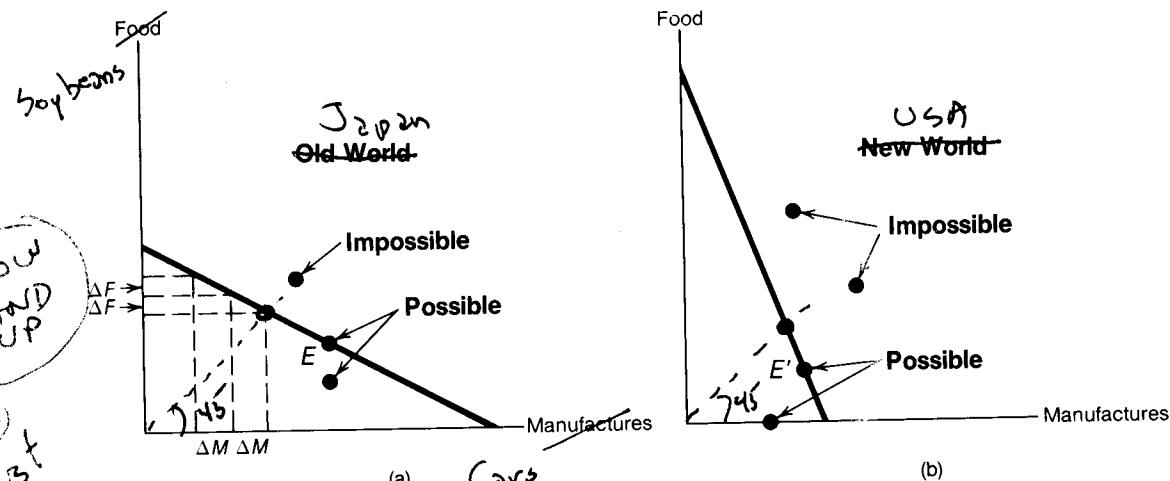
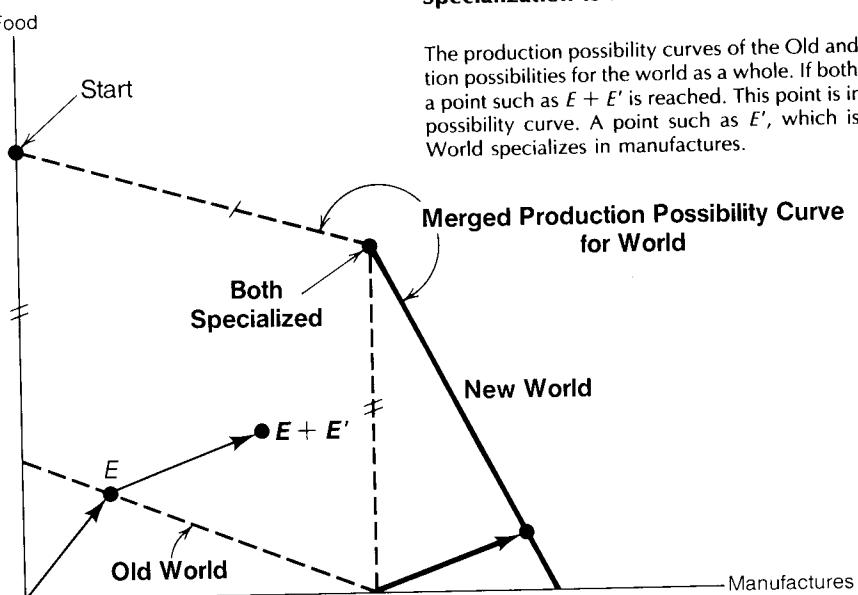


Figure 8.2
Specialization Is More Efficient Than Nonspecialization

The production possibility curves of the Old and New Worlds are added to give the production possibilities for the world as a whole. If both sides produce both food and manufactures, a point such as $E + E'$ is reached. This point is inefficient because it is not on the production possibility curve. A point such as E' , which is efficient, can be reached only if the Old World specializes in manufactures.



world is constructed as follows. Think of producing successively larger amounts of manufactures as cheaply as possible in food forgone. The early production, clearly, will be located entirely in the Old World, because the production there will have the lowest cost in food forgone. Only when the Old World is completely specialized in manufactures will the production of still more manufacturers require the New World to produce some. Moving to the right along the manufactures axis, in other words, first the Old World's and then finally the New World's production possibility curve is used to produce manufactures (see Figure 8.2).

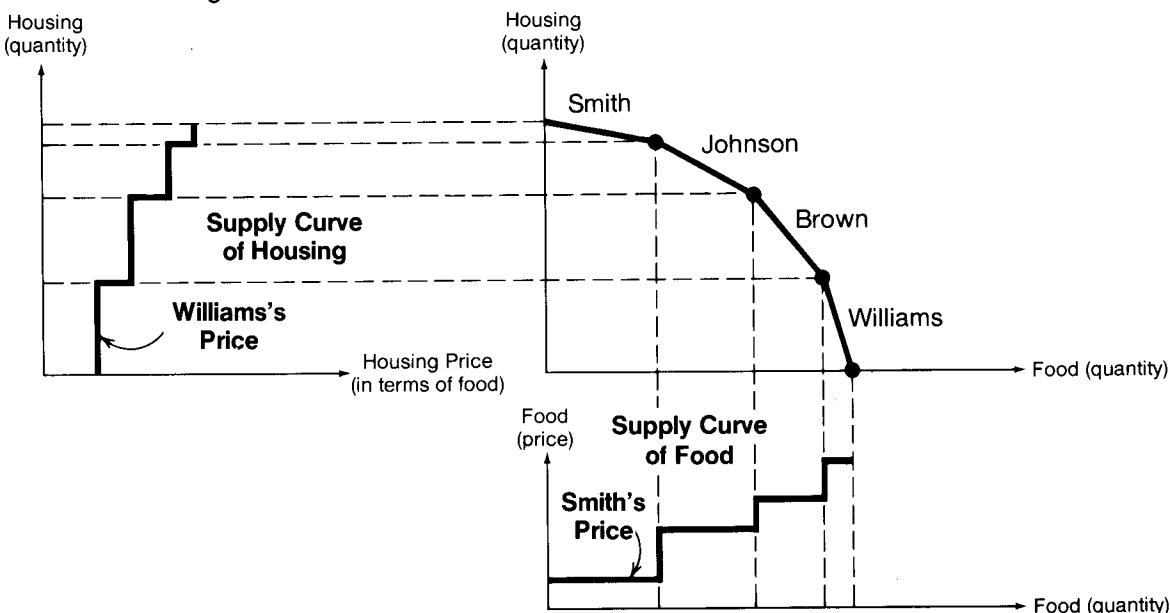
When no manufactures are produced anywhere, the merged world is at the point Start, both places producing only food. As the capacity of the Old World to produce manufactures is used up, the world moves along the dashed line (along which the New World still produces only food); after the capacity is used up, the world moves along the solid line. The outmost line, dashed and solid, is the world's production possibility curve with trade.

Now the proof of the specialization theorem is simple. Notice in the previous diagram the points E and E' . These are the points of production (and consumption) before trade. The point $E + E'$ in the new diagram adds the two bundles together. It is *inside* the world's production possibility curve. All such sums of nonspecialized points are inside the curve. Specialization, and the trade that

You can see that if you think like every farmer does this is going to be smooth GTO 166

Figure 8.3
A Bulging Production Possibility Curve Implies Upward-Sloping Supply Curves

The concave production possibility curve says that, as more food is produced the cost of food in terms of housing increases. This implies an upward-sloping supply curve for food. The same is true of the production of housing.



permits specialization, yields more goods than does self-sufficiency. Prohibiting trade always reduces total income; opening trade increases it.

***Specialization
Implies the
Conventional Bulge
for the Production
Possibility Curve***

The specialization theorem is in one sense obvious. Obviously, removing a restriction by opening trade cannot hurt the world. It is restrictions, after all, that hurt, and removing them must help. The theorem is one way of deriving a convex (that is, bulging outward) shape for production possibility curves. Just as two straight-line production possibility curves in the New and Old Worlds resulted in a (kinked) bulge in the world's curve, so also in a single economy many straight-line production possibility curves of Smith, Johnson, Brown, Williams, Miller, Jones, Davis, Anderson, Wilson, and Taylor would result in a bulged curve for the nation. The bulge is significant because it is an alternative picture of a feature of supply assumed so far with little comment (that is, its upward slope). Figure 8.3 exhibits the case for an economy of Smith, Johnson, Brown, and Williams producing food and housing. The production possibility curve for the economy between food and housing is in the upper right-hand panel. The other two panels are plots of the slopes of the curves. That is, they are the marginal opportunity costs, or the supply curves. From the idea of opportunity cost, then, comes an upward-sloping curve of supply.

Summary

The theory of demand produced an upward-sloping supply curve of food only as an excess supply, subtracting a small own demand for food from a large initial supply. The initial supply was given and unproducible. Eating food, therefore, had socially speaking no opportunity cost in, say, housing forgone. The theory of supply, by contrast, allows production. Producing food does actually have an opportunity cost in housing: For the society as a whole as well as for single consumers, more food means less housing. Opportunity cost is a characteristically economic idea. It motivates the economist's distaste for prisons relative to fines or for subsidies masquerading as production. It is the nub of the specialization theorem, namely, that a division of labor by comparative advantage raises a nation's production possibilities. And the specialization theorem, in turn, is the nub of supply, namely, that at higher relative prices for, say, baked goods, more labor and machinery are brought into the production of bread. The fresh recruits of labor and machinery are less suited to specialization in baking than are the earlier. That is the point. Only the higher relative price permits the higher opportunity cost of the relatively ill-suited labor and machinery to be incurred. Only at higher relative prices is more output forthcoming. As the production possibility curve bulges outward, the supply curve slopes upward.

EXERCISES FOR SECTION 8.1

1. Describe the pattern of specialization that would develop among the following groups, and why specialization would produce more for the group than equal sharing of all tasks:
 - a. Trevor Dick the doctor and Thomas Huertas the auto mechanic, in fixing each other's bodies.
 - b. Clark Nardinelli the .150 hitter who has a good fastball and curve, and Robert Gallman the .310 hitter who can't pitch at all.
 - c. Nardinelli, as described, and a Gallman after an operation on his arm, who now has a terrific fastball and a sensational curve in addition to hitting .310. (The answer is not certain. Pick an interesting possibility.)

d. Mr. and Mrs. Hogendorn, in painting chairs and making curtains. Make the assumptions about the sexual distribution of skills that our society exhibits.

e. The United States and Taiwan in making soybeans and television sets: Suppose the United States can make *both* soybeans and television sets in less time and other resources than Taiwan.

2. Describe the construction of the whole world's production possibility curve, as in Figures 8.1 and 8.2, if the individual production possibility curves of the Old World and the New World are reversed. That is, describe in detail how one would construct Figure 8.2 if the New World had a comparative advantage in manufacturers.

PROBLEMS FOR SECTION 8.1



1. When every few years or so a big company cannot meet its debts, there arises a clamor for subsidies. The Secretary of Transportation is worried at the thought of Penn-Central's bankruptcy; the Secretary of Defense is alarmed at the thought of Boeing's; the Minister for Industry is appalled at the thought of Chrysler-U.K.'s; the President is shocked at the thought of Chrysler-U.S.A.'s. "The bankruptcy of Big Company X," the cry goes, "would be an intolerable disaster for the nation. Imagine life without X! We must intervene at once with \$N million." Evaluate the clamor. Does bankruptcy entail a cost to the society as a whole? *Apply to Farming.*

- o 2. Show that, if the New and Old Worlds had the same slopes for their production possibility curves, then the point $E + E'$ representing production before trade would be on—not below—the world's production possibility curve.
- o 3. Suppose initially that three countries—Japan, Britain, and America—were self-sufficient, producing each along a straight-line of production possibility curve. The maximum tons of food and numbers of machines each could produce by devoting all of its resources to producing one or the other good were as follows:

Maximum Possible Amount of:	Japan	Britain	America
Food (millions of tons)	6	15	30
Machines (millions)	20	20	30

a. What will be the price of a machine expressed in tons of food in each country? Which is the low-cost producer of machines?

b. To take an especially simple case, if consumers in all three countries demanded food and machines in fixed proportions, 1 ton of food with 1 machine, what would be the world's production (= consumption) of the two goods? (*Hint:* The expansion path would be the ray $F = M$ for each country; the point of consumption would therefore be the intersection of $F = M$ and the equations for the production possibility curves.)

c. Sketch at an approximately correct scale the world's production possibility curve if the three countries open themselves to free trade. If the world has the $F = M$ preferences, as given, on which country's segment will production and consumption fall? Describe the pattern of specialization.

d. After free trade and specialization, how much will the world consume and produce of each good? Compare these amounts with those in (b) and draw the moral of the tale.

4. A production possibility curve between food and all other goods is analogous to an indifference curve between the two. The production possibility curve characterizes technol-

ogy, the indifference curve tastes. Experiments in revealed preference are interpretable as assertions that one can or cannot draw an ordinary indifference curve. *True or false:* If two combinations of food and all other goods violate the weak axiom of revealed preference, then from one combination to the next it is conceivable that technology has stayed the same, but inconceivable that tastes have; if the combinations satisfy the axiom, then it is conceivable that tastes have stayed the same, but inconceivable that technology has.

5. After 1859 Japan was opened to trade with the West (earlier it had had limited trade with Holland). That is, Japan was suddenly presented with the world's relative price of, say, food in terms of all other goods. Supposing Japan in 1859 to be so small as to have no effect over world prices and to have a straight-line production possibility curve between the two goods, describe in diagrams the impact on Japan of the opening of trade. In what conditions would the opening of trade have no effect?

True or False

- 6. The \$100 cost of journeying from New York to Washington is in all economic respects identical to a \$100 tax. Were it in fact costless to make the trip, a tax equal to the actual \$100 cost would have the same economic effects as the cost itself.
- 7. From the point of view of social output (if not necessarily of social morality), it is better for members of Congress and government bureaucrats to be bribed directly in cash than to be influenced indirectly by hordes of lawyers and advertisers armed with elaborate offices, researchers, and business luncheons.
- 8. On grounds of efficiency we should not acquire a president by election: The office should be for sale to the highest bidder.

8.2 The Production Function Also Leads to the Production Possibility Curve

What to Read For

What is the production function? Why is it more measurable than a utility function? What is a Leontief production function? An isoquant? Why are isoquants usually curved instead of sharp-cornered? How does the idea that there is more than one way to skin a cat relate to isoquants? How exactly do production functions connect with the production possibility curve? Why does the production possibility curve bulge out?

The Production Function Is Similar to the Hill of Utility

The fundamental idea in the theory of supply is, to repeat, the idea of opportunity cost or its diagrams, the production possibility curve and the upward-sloping supply curve. One way of getting an outward-bulging production possibility curve (and therefore an upward-sloping supply curve) is the specialization theorem. The other way—underneath it all, the same way—is the *production function*.

The word “function” is used here in its mathematical sense of a correspondence between maximum output (what is produced) and inputs (the people, machines, and things used to produce). In plain language, the production function of, say, gourmet food is a book of recipes. Just as various combinations of housing

and all other goods “produce” happiness according to the *utility function*, so various combinations of farmland, tractors, laborers, fertilizer, grocery space, transportation, and so forth produce food according to the *production function*.

Look at the corner of the room, at the floor. Think of the corner itself as the origin. The vertical axis goes from floor to ceiling, measuring food production as a height above the floor. The floor down by the left and right walls (as you look at them now) are the two other axes, measuring amounts of farmland and farm labor. Imagine a hill of food production rising out of the corner. This hill shows the different amounts of food that can be produced from different combinations of farmland and farm labor. Someone climbing the hill would start at a low (or zero) altitude in the corner and would cross higher and higher contour lines representing larger and larger amounts of food as she climbed up the hill, moving out over the middle of the room. The production function, in other words, is analogous to an ordinary hill.

The difference between utility and production functions is that the utility from a bundle of goods is not measurable in any unambiguous way, while the output from a bundle of inputs is measurable quite easily. The output of food can be touched, inspected, tested, and weighed. Utility or joys or happiness cannot. The recipe (or production function) for French bread contains four material ingredients (flour, salt, yeast, and water), one piece of capital equipment (the oven), and one laborer (an inspired baker). Combining these in such-and-such proportions has a measurable result, one loaf. But combining the loaf with cheese and wine in a lunch on the grass produces ecstasy (utility) of an unmeasurable sort, having no natural units and incomparable between two people. To an extent unmatched in utility theory, therefore, economists have actually measured production functions.

Substitutability Is Common in Production as in Consumption

What they have found in the measurements is in accord with common sense, namely, that there is more than one recipe for food, and more than one blueprint for housing, and more than one way in which to make goods out of the resources available to the economy. The same field of wheat can be harvested with 50 laborers, 50 scythes, and no automated harvesters, or 1 laborer, no scythes, and 1 automated harvester. The opposite idea is a production function with fixed proportions, that is, one and only one recipe. It is called a *Leontief production function* in honor of the economist who made the most fruitful use of it. The hill of production in this case has a ridge. In chemical processes the recipes are often fixed, as in the making of water (H_2O) from two atoms of hydrogen and one of oxygen (see Figure 8.4). Given 2 atoms of oxygen, more than 4 of hydrogen does not produce more water. The excess hydrogen is simply unemployed, because all of it that is needed to produce 2 molecules of water is 4 (not 5 or 10 or 8). The production of 2 molecules of water *requires* 4 atoms of hydrogen and 2 of oxygen.

The words “needed” and “requires” suggest that the situation does not involve choices and is therefore not economic. Just so. Corners in utility contours represent the antieconomic idea of “need” in consumption. Likewise, corners in production contours represent need in production. The theory of consumption says that consumers can satisfy themselves with various bundles of goods, not only with the one bundle they “need.” Likewise, the theory of production says that producers can make bread, houses, automobiles, or whatever with various bundles

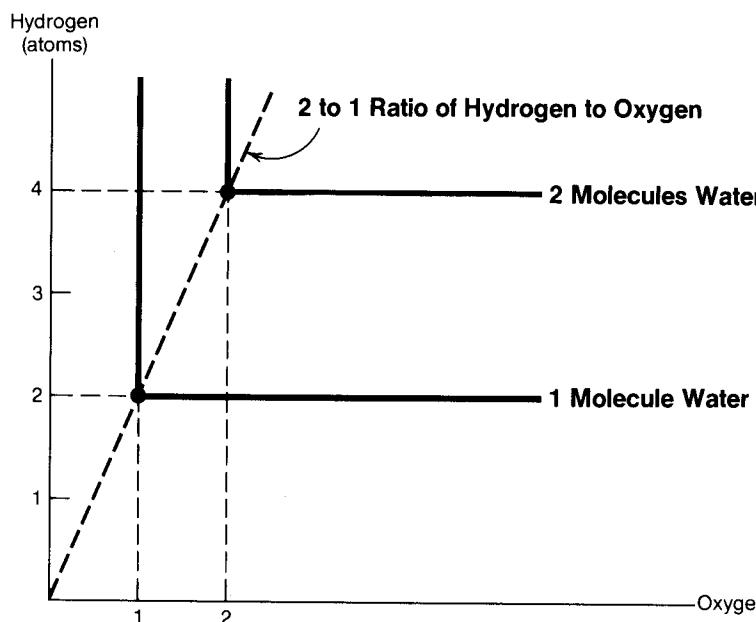


Figure 8.4
In Production Functions with Fixed Proportions the Contour Lines of Output Have Corners

If the ratio of hydrogen to oxygen is greater than 2 to 1, hydrogen is being wasted; if less, oxygen. The waste is just thrown away: It's not needed.

of resources, not only with the unique recipe. The contour lines for most products are not corners, but smooth. The production of most products does not require fixed proportions.

Or, to put it another way, there's more than one way to skin a cat. You can skin the cat in a machine-intensive way by dropping it into an automatic cat-skinning machine and pressing the button. Or you can skin it in a labor-intensive (and machine-light) way by giving it to a team of bloodthirsty folks. The various ways of skinning 1 cat are represented by the contour lines of altitude 1 skinned cat on the hill of production (see Figure 8.5).

Note the contrast in shape with the preceding figure. The curve (contrasted

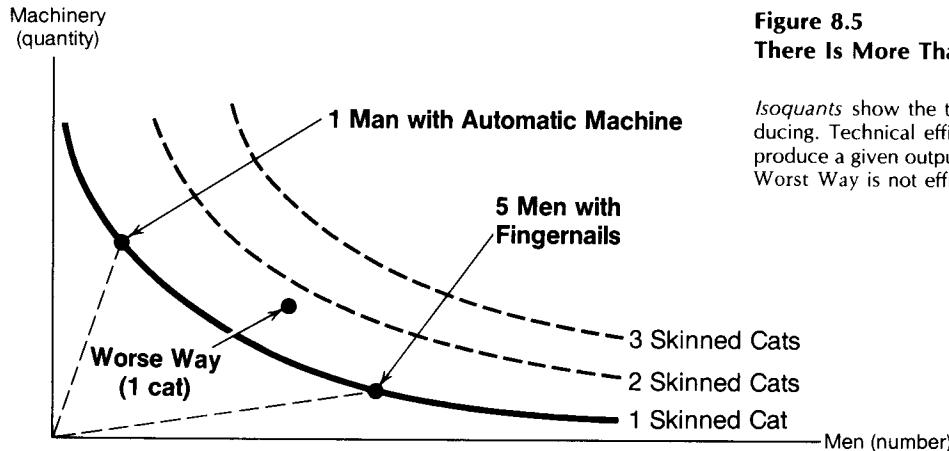


Figure 8.5
There Is More Than One Way to Skin a Cat

Isoquants show the technically efficient means of producing. Technical efficiency requires that the inputs to produce a given output be as little as possible. The point Worst Way is not efficient.

to the corner) says that there are many ways in which a cat can be skinned. The point Worst Way is technically speaking an inefficient way to skin a cat.¹

The Uses of Substitutability

The mere idea of substitutability in production, like the mere idea of substitutability in consumption (the Law of Demand), is powerful and important. Consumers can gratify given tastes in many ways; producers, similarly, can use given technological knowledge in many ways. The idea applies to the production of anything—grain in India, education in America, steel in Britain, industrial output in Russia, all output in Brazil. An example is the enemy's output of soldiers and equipment in the field.

Q: Criticize from the point of view of the economics of production the following argument in favor of "strategic bombing," that is, bombing the enemy's sources of supply at home. "The enemy needs railways (or ball-bearings or oil) to wage modern war; therefore, we can cheaply disarm him, with little risk to our own forces, by bombing from a great height his rail yards (or ball-bearing factories or oil refineries)."

A: The argument supposes that the enemy (successively in this context from the 1930s to the 1970s the Ethiopians, the Loyalists, the Germans, the North Koreans, and the North Vietnamese) *needs* railways. It supposes, that is, that one man in the field *requires* so many miles of operating track and that bombing the track will make it impossible to sustain the man in the field.

But the economic presumption—not always true, but always worth considering—is that on the contrary there are many ways of sustaining a man in the field. To be sure, placing a ton of TNT on the first way of doing it will raise the cost of sustaining the man. The second way (marshaling repair crews to fix the track quickly, rerouting the trains, carrying the equipment on people's backs down the Ho Chi Minh Trail, or whatever) is more expensive. That is why it is the second, not the first, way. But "more expensive" is not "impossible." Although the evidence is in some cases controversial, strategic bombing has in fact seldom had the effects claimed for it. Because there is more than one way in which to skin a cat, the "essential" part of the enemy's war machine destroyed by the bombing was seldom truly essential.

Similarly, the economist is suspicious of the engineer's declarations that San Francisco "needs" a rail system of public transport, or that farmers in the San Joaquin Valley "need" a million gallons of water, or that America "needs" X million barrels of oil. He has seen alternatives, such as buses and cabs in Hong Kong, water conservation in Kansas, and smaller cars and colder houses in Europe. A world of unique recipes for each thing would be simpler, but it is not the world as it is.

Substitutability Implies That the Production Possibility Curve Is a Curve, Not a Point

Such is the direct, practical use of the idea of a production function. It will come up later again and again: As utility functions are to demand, production functions are to supply. The indirect, theoretical use of the idea, however, is what matters for now. Suppose that the United States in 1944 produced two goods, guns and butter, by allocating two resources (or *factors of production*), labor and land, between the two production functions. All the labor and land could be allocated to guns, producing the highest possible level of guns and zero butter, or all of it could be allocated to butter, or various possible combinations in between. Think of the parallel between utility (joys) as a function of inputs of food and all other goods and production (guns) as a function of inputs

¹ The *unit isoquant* (that is, the contour for one cat) represents the ways that use the least machines or people to do the trick. The *New Yorker* once proposed the very Worst Way to skin a cat: Prick a hole in its tail and vacuum out the insides.

of labor and land. The thought suggests that, just as an Edgeworth box can represent the allocation of two goods between the utility of John and of Harry, so too it can represent the allocation of two factors between the production of guns and of butter. It can (see Figure 8.6).

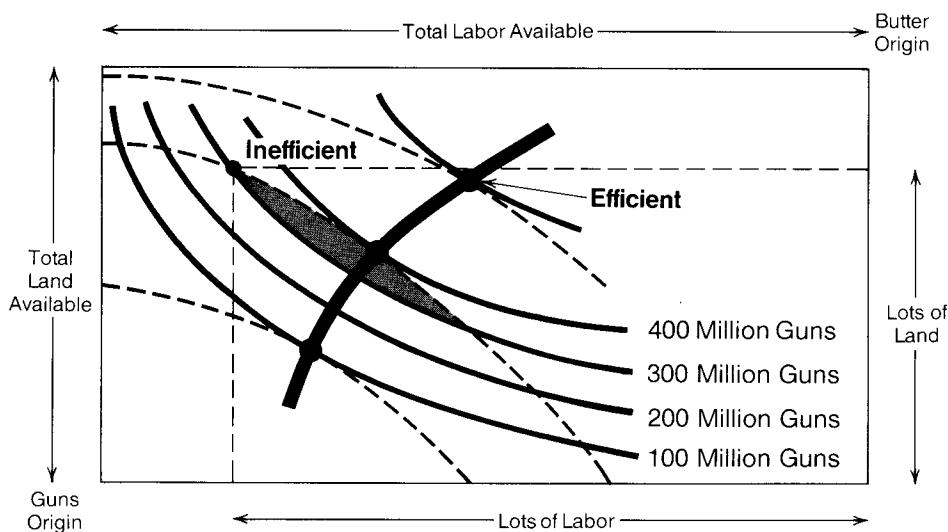
At point Inefficient, for example, lots of labor is allocated to butter production and lots of land to guns, the rest of each factor going to the other product. The analogy with consumption theory suggests further that the line of tangencies between the two sets of contours has especially desirable properties. So it does. Starting from an allocation on the line (the "contract curve," you recall), a central planner of the economy would not get more guns without sacrificing butter. From point Inefficient, by contrast, he could get more guns *and* more butter: He could shift the allocation to any point inside the shaded lens shape. All of which is to say—and this is the payoff—that combinations of guns and butter along the contract curve are points *on the economy's production possibility curve*, that is, efficient points, at which production of both guns and butter cannot be increased.

Figure 8.7 shows the correspondence between the contract curve (for production) and the production possibility curve. The point Inefficient is just that: off the contract curve in the Edgeworth box and inside the production possibility curve.

One more point: The production possibility curve bulges out. By now you're accustomed to seeing that this means the supply curve would slope upward.

Figure 8.6
An Edgeworth Box for Production

At the point Inefficient, the output of guns could be increased to 400 million without reducing the output of butter. This is done by reallocating land and labor to gun production in such a manner that the economy moves along the butter isoquant from the 300 million to the 400 million gun isoquant.



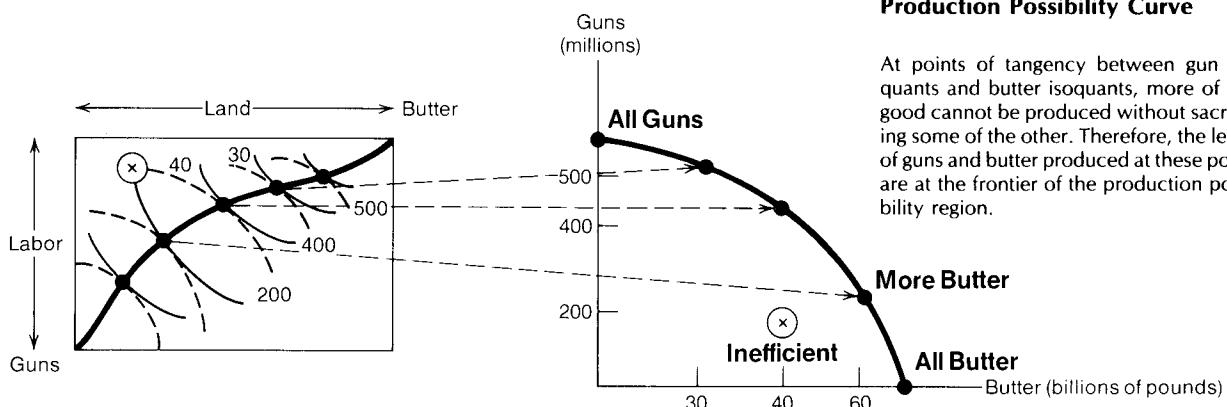


Figure 8.7
Points on the Contract Curve
Correspond to Points on the
Production Possibility Curve

At points of tangency between gun isoquants and butter isoquants, more of one good cannot be produced without sacrificing some of the other. Therefore, the levels of guns and butter produced at these points are at the frontier of the production possibility region.

So the direction of bulge is important. Why does it bulge outward? The answer lies in the difference between the production functions for guns and for butter. Suppose that guns use labor relatively intensely—one might say that Joe Labor had a comparative advantage in gun production relative to Larry Landlord. Imagine the nation starting at the point All Guns and moving toward All Butter. At first the new pounds of butter will take up the land not used very much in guns anyway. The butter folks will take what they wish of the land and labor in the proportions best suited to making butter. The butter will be cheap in guns sacrificed. But as butter takes up more and more of the nation's resources, it will run up against increasingly scarce land. At the extreme, near the point All Butter, the butter folks have to take land and labor as they find them in the economy overall: When you're taking 99% of the labor and land you can't be choosy about the proportions you take it in. The butter is clearly going to be a lot more expensive. So the curve curves as it does: The supply price rises.

Summary

From the marriage of the idea of a production function and the idea of factors of production (as in the preceding section from the marriage of the idea of specialization and of diversity of natural gifts) has come the production possibility curve. No other child could be expected, because the fixed endowments of factors and the fixed production functions that allow one to draw the Edgeworth box for production are merely two kinds of scarcity—a scarcity of resources to make things and a scarcity of (a limit on) the knowledge of how to use resources to make things. Scarcity is the idea underlying the production possibility curve. And a bulging production possibility curve implies an upward-sloping supply curve.

EXERCISES FOR SECTION 8.2

1. From your knowledge of such things, list the inputs into the following products. Don't forget machines and labor. Be specific about the inputs, though not exhaustive. Obviously such questions don't have full and precise answers.

- a. Lumber.
 - b. Books.
 - c. Postal service.
 - d. Police protection.
 - e. New houses.
 - f. Shelter in old houses.
2. Name the substitutes for the following. Bear in mind that not *all* of the input needs to be replaced.
- a. Hand labor in producing roads.
 - b. Doctors in producing health.
 - c. Oil in producing heating of houses.
 - d. Oil in producing transportation.
 - e. Steel in producing cars.
 - f. Professor in producing educated students.

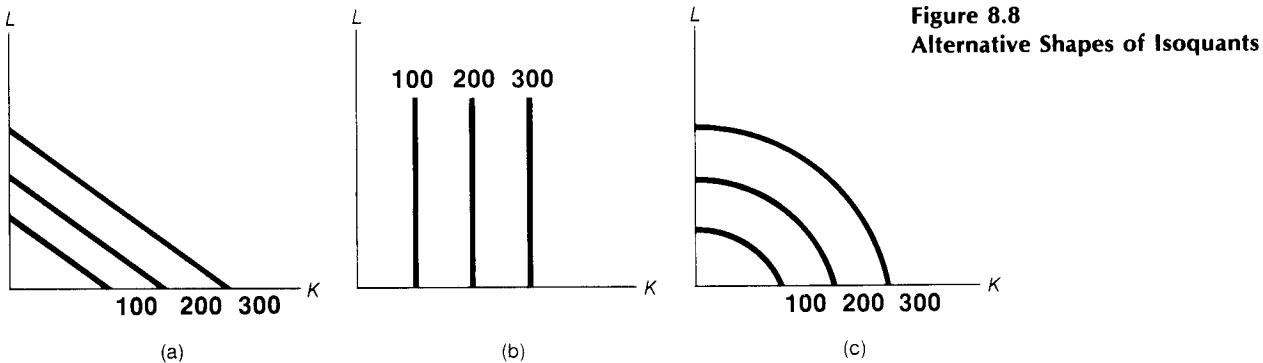
PROBLEMS FOR SECTION 8.2

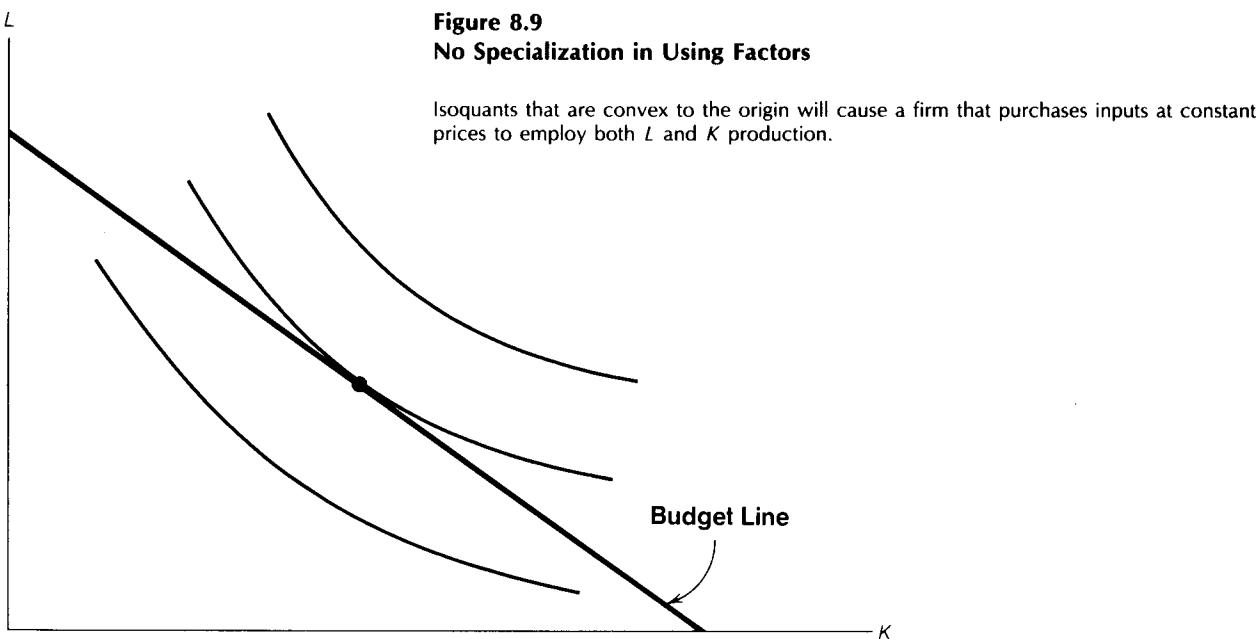
1. Grain could move from Chicago to New York by two routes, water (by way of the Great Lakes and the Erie Canal or down the Mississippi to New Orleans and thence by ship to New York) and rail. By the end of the nineteenth century about 70% of the grain in fact went by rail. *True or false:* Had the railway never been invented, the United States would have lost an amount equal to the value of the transport of grain provided by the railway.

2. Visualize the contour maps in Figure 8.8 (*isoquants* means “equal quantities”) in three dimensions and describe them as hills of production.

3. A company, industry, or whole economy facing a budget line of labor L and capital K would choose the highest contour attainable within the budget line. In view of this, and using the analogous argument in consumption theory, show why the typical shape of output contours is none of the shapes in the preceding problem but rather that shown in Figure 8.9.

4. According to the “need” theory of economics, a continuously depleting resource should rise continuously in relative price, driving the economy to disaster. Oil needed for tractors, which are needed to make flour, which is needed for bread, which is needed to sustain life, should rise continually in price as it depletes. Yet in fact the price of such resources often levels off or falls after an initial rise. The price of Greenland whale bone, for example, rose from £250 per ton in 1835 to about £2800 around 1900, but then fell to £2500 by





1910; the price of whale oil rose from its base of £40 per ton after 1835 but had fallen to £25 by 1910. With cat skinning in mind, why?

- 5. Consider an Edgeworth box in which labor and land are fully employed and are divided, say, 50:50 between guns and butter. How would you represent less than full employment of labor? Is more output attainable from putting the unemployed to work? How would you represent unemployment in the corresponding production possibility curve? What is the result in this diagram of putting the unemployed to work? Is there a free lunch?
- 6. Use the Edgeworth box for production to show that if food and all other goods both use land and labor in rigidly fixed proportions (the proportions being different for the two goods and lying on opposite sides of the proportional endowment of land and labor in the nation as a whole), then only one allocation of land and labor to the two goods (that is, only one point in the box) will fully employ both land and labor.

8.3 How an Economy Works in the Large

What to Read For

What is the economywide demand curve corresponding to the economywide supply? What is the relationship between the supply curve and the economy's production possibility curve? Between the demand curve and the production possibility curve? What's so great about the equilibrium point of supply and demand? How can you use economywide pictures to analyze problems?

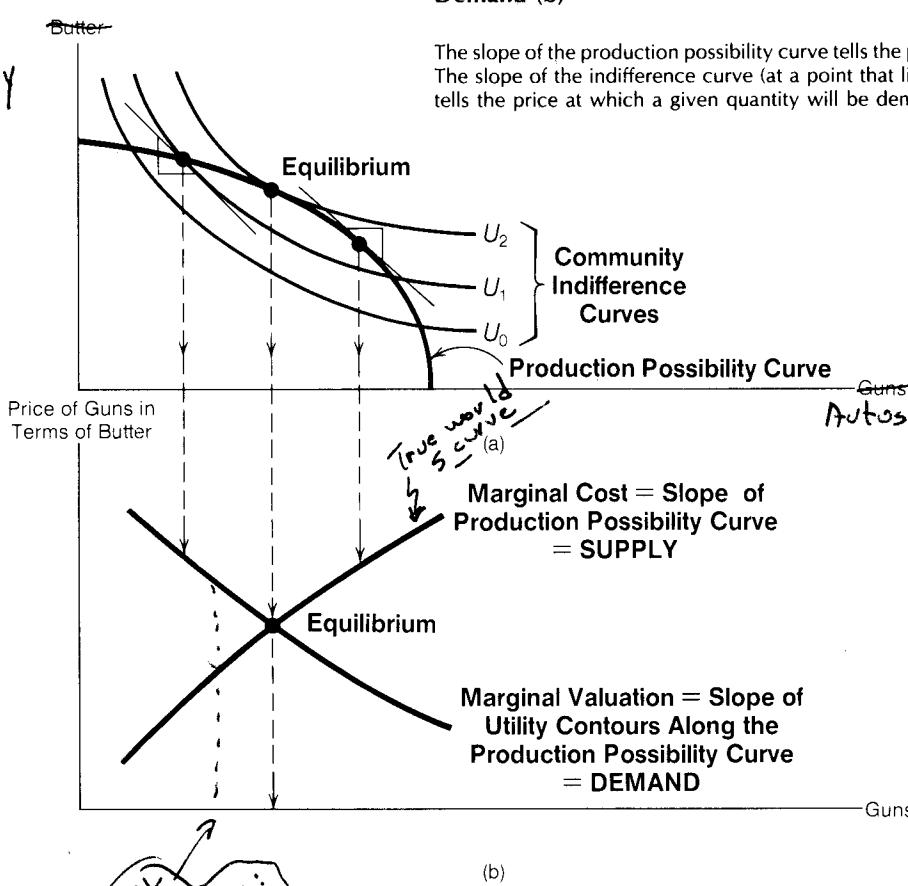
Adding the Law of Supply (the Outward Bulge) to the Law of Demand

Guns become more expensive, then, as more are made. Furthermore, as more are made, they become less valuable. The armies are well supplied, the hunters are armed to the teeth, and an extra gun has a lower marginal valuation than did earlier parts of the supply. Corresponding to the supply of guns curve obeying

This is for the world

Figure 8.10
Production Possibilities and Indifference Curves (a) Lead to Supply and Demand (b)

The slope of the production possibility curve tells the price at which quantity will be supplied. The slope of the indifference curve (at a point that lies on the production possibility curve) tells the price at which a given quantity will be demanded.



the Law of Supply (upward sloping), then, is a demand for guns curve obeying the Law of Demand (downward sloping).

It's hard to resist the impulse to put the supply and demand in the same diagram and to watch where they cross. The supply of guns curve is the marginal cost in terms of butter foregone (that is, the slope of the community's production possibility curve); the demand curve is the marginal valuation (that is, the slope along the production possibility curve of the community's indifference curves). The top panel in Figure 8.10 leads to the bottom panel. Look at this diagram for a while; it will come up again.

The marginal valuation constituting the demand curve must be measured along the production possibility curve because only along the curve are resources fully employed. The marginal valuations come from whatever marginal valuation the amount and distribution of income casts up at the point in question on the production possibility curve.

We're accustomed to think in US terms. But can also think of whole world [e.]

The Equilibrium Is Efficient

Now look at the point Equilibrium. At Equilibrium in the top panel the typical consumer reaches the point that is on the highest indifference curve within the production possibility curve. At that point the community is doing as well as it can. The point is also the one of equal slopes of indifference and production possibility curves. This way of saying it is shown by the other point Equilibrium in the *bottom* panel: Supply (the slope of the production possibility curve) is equal to demand (the slope of the indifference curve along the production possibility curve).

Allowing for production, in other words, leaves unaltered the conclusion of earlier chapters: When markets work well, they lead to efficiency, as by an invisible hand. Much of the rest of the book discusses how markets might actually work badly. But for now be of good cheer. The selfish interests of the butter producer lead him and his colleagues to sell butter at its true opportunity cost in terms of guns. The selfish interests of the butter consumer, likewise, leads him to buy at its marginal valuation. Marginal valuation is equal to marginal cost, and the opportunities for mutually advantageous exchange between producers and consumers, as between different consumers, are exhausted. The economy is driven by self-interest to be efficient both in consumption and in production.

The Uses of the Economywide Diagram

The model is applicable directly to questions of how economies behave.²

Q: It is said that two decades of weather favorable to British agriculture before the middle of the eighteenth century affected the output of nonagricultural goods, especially manufactures. But it is not clear how. One historian has argued that good harvests cheapened food, therefore enriching city dwellers. Therefore it increased the demand for manufactures—and advanced the Industrial Revolution. Another has argued that good harvests cheapened food, therefore impoverishing farmers (the demand for food was inelastic, implying that a lower price gave lower total expenditures on food). Therefore, it decreased the demand for manufactures—and retarded the Industrial Revolution. Arbitrate this dispute.

A: One decrease plus one increase equals zero. Both are partly correct, but both are taking a part of the nation (each a different part) as the whole. City dwellers were enriched, farmers impoverished. But the enrichment of one is the impoverishment of the other, offsetting in total effect. To the extent that one is made better off (paying less), the other is made worse off (receiving less). These analyses of the effects of good harvests,

in other words, concern the shifting of something in Britain's left hand to its right. What is required is an analysis of the effect of her increased output overall (arising from the good harvests), not shifts in distribution.

Q: Good weather, like technological improvement, raises the production function of agriculture: At each bundle of inputs the resulting output is higher than it was before.

1. How does it affect the production possibility curve between agriculture and all other goods?
2. How does it affect the corresponding supply curves of agriculture and all other goods?

A: 1. The simplest way to answer (1) is to look at the two ends of the production possibility curve (see Figure 8.11). At the point Exclusively Agriculture the economy devotes all its resources to agriculture, and the higher productivity in agriculture has its maximum effect. At Exclusively All Other Goods, on the other hand, the higher productivity in agriculture has no effect, because

² The example is taken from Richard Ippolito, "The Effect of the 'Agricultural Depression' on Industrial Demand in England, 1730–1750," *Economica* n.s. 42 (August 1975): 298–312.

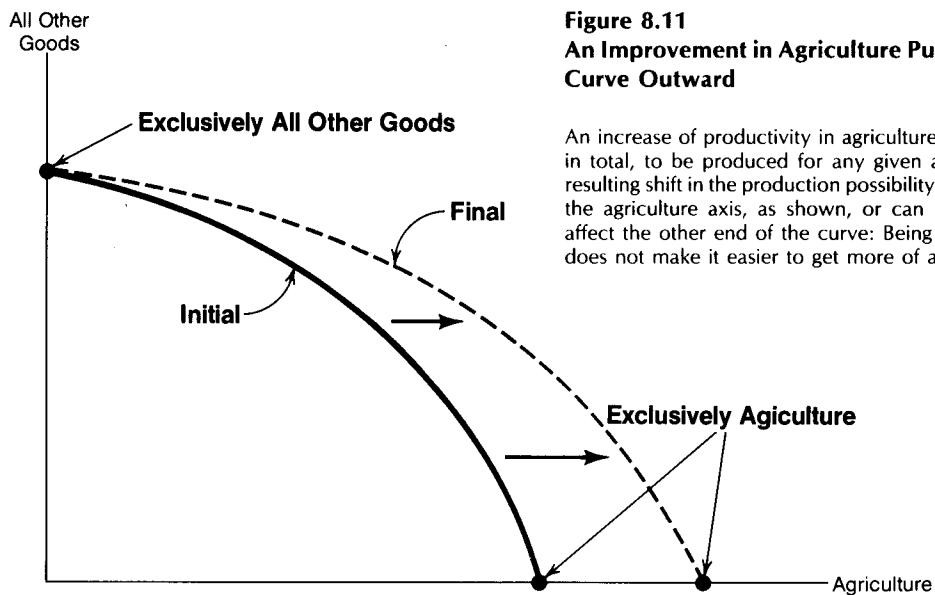


Figure 8.11
An Improvement in Agriculture Pushes the Production Possibility Curve Outward

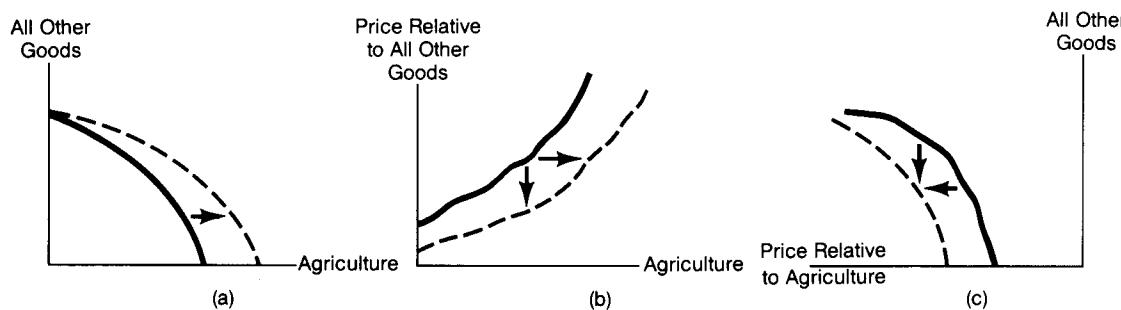
An increase of productivity in agriculture permits more agricultural products, in total, to be produced for any given amount of all goods produced. The resulting shift in the production possibility curve can be equiproportional along the agriculture axis, as shown, or can be irregular. In any case it doesn't affect the other end of the curve: Being able to get more agricultural goods does not make it easier to get more of all other goods.

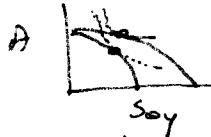
there is no agriculture to be affected. The final production possibility curve is a horizontal stretching of the initial curve. A horizontal stretching of the curve occurs because, for example, a 20% rise in the agricultural output reaped from given inputs, due to good weather, would have a larger absolute effect the larger was agricultural output to begin with.

2. The answer to (2) depends on the effect of the stretching on the slope of the production possibility curve. Such smooth stretching—as opposed to rough stretching, which could leave the final curve in some irregular shape—affects the supply curves of agriculture and all other goods as follows. Look at Figure 8.12 and turn the book clockwise on its side to see the left panel.

Figure 8.12
Improvement in Agriculture (a) Raises the Supply of Agriculture (b) and Lowers the Supply of All Other Goods (c)

The outward shift in the production possibility curve along the agriculture axis affects the supply curve as follows: More agricultural goods are forthcoming at any given price; fewer all other goods are produced at that price. So agriculture gets cheaper (of course).





The lower supply of all other goods (that is, the higher supply price for a given amount) is an inevitable result of the cheapening of agriculture. To put it another way,

the new, higher productivity of resources in agriculture implies that employing them in all other goods has a higher marginal opportunity cost.

Toward? ?



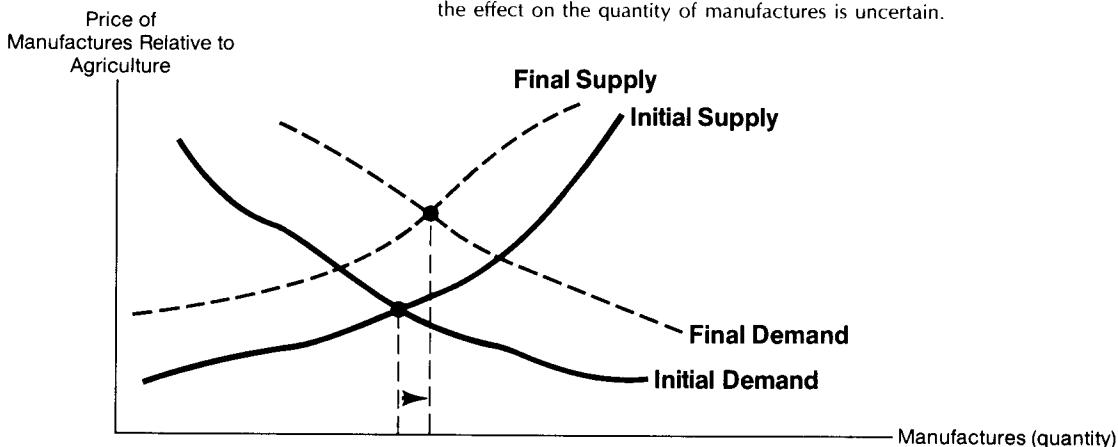
Return, then, to the historical question: What *was* the effect of good harvests on the demand for British manufactures and the Industrial Revolution? Suppose that manufactures are the All Other Goods. The effect of the good harvests on the whole demand curve is plain. Good harvests plainly make the nation richer. Therefore, if manufactures are normal goods, the income effect will be positive and the demand curve for manufactures will move outward. The effect on the quantity demanded in equilibrium, however, is ambiguous. The demand curve has moved outward, *but the supply curve has moved inward*, since manufactures now have a higher opportunity cost (see Figure 8.13). In other words, the income effect works to increase the amount of manufactures demanded; but the substitution (or price) effect works to decrease it. The supply and demand curves move in opposite directions.

The net effect pictured here is a small rise. But the effect could easily be a fall or a larger rise. The answer to the historical question depends on the relative strengths of the income and the substitution effect. And in any event the substitution effect serves to moderate the income effect. It can be shown, in fact, that the net effect was trivially small. Looking at the good harvests from a national perspective reveals that they had little effect on the coming of the Industrial Revolution.

Such are the uses of supply and demand applied to entire economies. You will sometimes hear it said that supply and demand are ideas applicable only to little markets in isolation. The jargon is that the model of supply and demand is a *partial equilibrium* rather than a *general equilibrium* one. General is not always better than partial equilibrium because the costs of complexity may exceed

Figure 8.13
The Uncertain Effect of Bigger Harvests

When the demand schedule shifts outward due to an increase in income and the supply schedule shifts inward due to an improvement in factor productivity in the other good, the effect on the quantity of manufactures is uncertain.



the benefits of accuracy and completeness. The main point here, however, is that the model of supply and demand can be given a very general interpretation, as a picture not merely of local events in the market for, say, shoes, but of the national links between shoes (or consumer goods or manufactures) and all other goods.

Summary

Getting the most happiness out of a production possibility curve, it turns out, is the same as equalizing supply and demand. In other words, the point of equality of supply and demand is efficient. Supply and demand have economy-wide interpretations: supply as the slope of the economy's production possibility curve, demand as the slopes of utility contours along the production possibility curve. Supply and demand, in other words, apply to the entire economy as they do to single markets.

EXERCISES FOR SECTION 8.3

1. Go through the analysis of guns and butter with another pair of products, say education and all other goods; or leisure and all other goods.
2. What if the better British weather in the text problem led also to an improvement in manufacturing productivity? (For instance, many manufactured goods used agricultural inputs: Beer used barley and hops, iron used wood for charcoal, woolens used wool, and so forth.)

PROBLEMS FOR SECTION 8.3

- 1. In rich countries over the last century, the birthrate has fallen. Children can be viewed as an item of consumption in which technological change has been small relative to that in all other goods over the last century. Show that if the income elasticity of demand for children is small (say, zero, to take the simple extreme), then the small rate of technological change in child rearing implies a fall in the number of children.
- 2. Suppose that Britain in 1750 faced a given international price of agricultural relative to manufactured goods. At the given price it could buy or sell any amount without altering the price. Exhibit in a production possibility diagram how Britain would choose its point of production and its (different) point of consumption. Identify exports and imports.

True or False

- 3. If the British economy in the eighteenth century were open to trade (see Problem 2), it is more likely that good harvests would have raised manufacturing output.

The Economics of Welfare and Politics

9.1 The Economics of Ethics

What to Read For **What is positive economics? What is normative economics? Should economists make moral judgments about the economy? What is laissez-faire? Why might laissez-faire be a Good Thing? How do you show this with a utility possibility curve? What is utilitarianism? What is the bliss point and how is it related to the laissez-faire point? How can you show tastes for equality? What is an egalitarian?**

Why Bother with Happiness?

People are not satisfied with merely understanding how the economy operates; they want also to judge the operation morally, to say whether it is good or bad and to convince Bud McGrath next door of its goodness or badness. Consider the following example. The invention of a wheat that is so short and stout that it can bear a large number of grains without breaking will have certain predictable effects. Big farmers who are willing to buy the fertilizer necessary to get the larger number of grains will in the first instance make more money out of the invention than small farmers. The price of wheat will fall; India and Mexico, formerly importers as well as producers of wheat, will become exporters; and so forth.

Economists could if they wished confine their work to making such predictions of outcomes, leaving to moral philosophers or politicians or men in the street the task of judging the outcomes, good or bad. There is a case to be made for evading in this way the responsibility for offering moral judgments as an economist, because (the case goes) an economist has no special wisdom about morality. Two reasonable people could agree on all the predictions of the outcomes of the "green revolution" in wheat yet disagree on whether the revolution was, on balance, a good or a bad thing. In this view the economist refusing to go beyond prediction merely recognizes her comparative advantage. Economics could convince reasonable people of the truth of the "positive" ("what is") predictions; but economics cannot, it is said, end a "normative" ("what should be") disagreement. As Lionel Robbins put it in 1932:

If we disagree about means, then scientific analysis can often help us to resolve our differences, [but] if we disagree about ends it is a case of thy blood or mine—or live and let live according to the importance of the difference. . . . [For example] if we disagree about the morality of taking interest (and we understand what we are talking about) then there is no room for argument.¹

*Moral arguments tend to...
work thru stories.*

Robbins and most economists after him, however, went much too far in distinguishing arguments about ends from arguments about means. For one thing, although arguments about what the end *should* be may be less easy to decide in some ultimate sense than are arguments about what the means in fact *are*, this is irrelevant to deciding arguments well short of the ultimate. For another, the materials for scientific argument are on close inspection similar to those for moral or artistic argument. That murder is evil is no more or less true than that the law of demand is reasonable. Science is shot through with artistic and social judgments; morality is shot through with facts and logic.

Laissez-faire

The economist's brand of moral philosophy is called *welfare economics*. The central controversy in welfare economics concerns laissez-faire, a French idiom meaning "let things drift" or "refrain from interference." It has come to mean the policy of removing the government from economic affairs: no Federal Communications Commission, no health inspectors, no minimum wage, no antitrust laws. Much of welfare economics consists of arguments for and against laissez-faire.

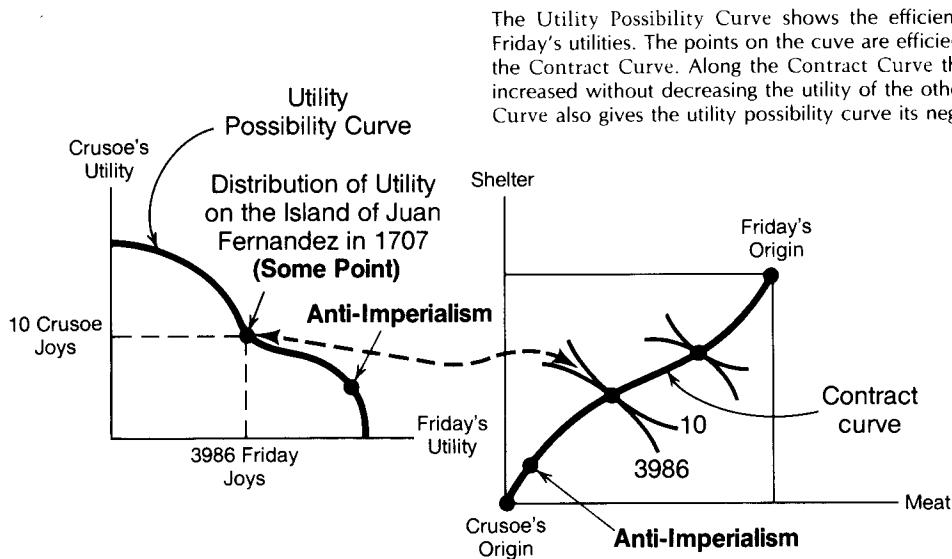
The argument in its favor is by now familiar to you. Start with the premise that envy is not to be indulged, that the childish instinct to throw the candy down the sewer if not all can have some is to be resisted. The goal is to make each person as happy as possible without hurting other people. To a noneconomist it would seem that such a goal would require the close attention of the Prince, rushing about with his inspectors and police to enforce the laws that order happiness (laws such as minimum wages and protective tariffs) and prohibit hurt (such as factory safety regulations and laws against price gouging). Wonder of wonders, however, if people are merely permitted to exchange and if the exchanges take place under certain favorable conditions, the goal will be achieved with laissez-faire. The Prince won't have to lift a finger. As has been shown repeatedly in earlier chapters, at the conclusion of trade the distribution of goods will be efficient. That is, all the mutually advantageous exchanges will have taken place. To revert to earlier language, people will be on the contract curve, along which no one can be made better off without making someone else worse off. Unexpectedly, then, the invisible hand will achieve easily what the Prince and all his guards, counselors, sheriffs, and servants could achieve only with difficulty.

Laissez-faire Arrives on the Utility Possibility Curve

The technical way of putting this is to say that well-functioning markets will put society on the *utility possibility curve*. The curve is to the Edgeworth box in which two goods are allocated between Crusoe and Friday as the production possibility curve is to the other Edgeworth box in which two inputs are

¹ Lionel Robbins, *An Essay on the Nature and Significance of Economic Science* (London: Macmillan, 1932), p. 134, sentences reordered. Quoted in A. K. Sen, *Collective Choice and Social Welfare* (San Francisco: Holden-Day, 1970), pp. 62–63, which is a useful (though advanced) discussion of the point.

Figure 9.1
Laissez-faire in a Perfect World Achieves Some Point on the Utility Possibility Curve



The Utility Possibility Curve shows the efficient combinations of Crusoe's and Friday's utilities. The points on the curve are efficient because they are derived from the Contract Curve. Along the Contract Curve the utility of one party cannot be increased without decreasing the utility of the other. This property of the Contract Curve also gives the utility possibility curve its negative slope.

allocated between guns and butter (the only two products in the society). This is shown in Figure 9.1. The curve slopes downward like the production possibility curve. But it has no particular direction of bulge. Along the curve marked as the Contract Curve Crusoe cannot become better off without Friday becoming worse off, but because the measure of utility can vary from person to person the axes can be stretched to achieve any bulge.²

A particular endowment of goods and bargaining skill will lead to a particular point on the Contract Curve and—what is the same thing—a particular point on the Utility Possibility Curve. Any point on the curve is efficient, in the technical sense of making one person as happy as is possible without reducing the happiness of the other. As usual, then, there are infinitely many social arrangements that are efficient. That is, there are infinitely many different points on the utility possibility curve.

One could stop here, declaring good any society that achieved by free exchange Some Point on the utility possibility curve. Any point off it can be improved upon by mutually advantageous exchange; any point on it cannot. Such a society would not necessarily be noble or equal or just. But, to make the most modest

² Strictly speaking, the curve can slope upward too. If Crusoe, for instance, values Friday's utility, then a move along the contract curve apparently making Friday better off at the expense of Crusoe might in fact make Friday *and* Crusoe better off. Or Friday may be so badly fed that his work for Crusoe suffers, in which case both could be made better off by giving Friday more food. See J. de V. Graaff, *Theoretical Welfare Economics* (Cambridge: Cambridge University Press, 1957), pp. 59–63, for this and other points.

claim, it would be efficient. Or, to make the claim that is sometimes made, it would be free.

Another initial endowment of goods and power between Crusoe and Friday would lead to another point, point Anti-imperialism, say, with Crusoe worse off and Friday better off. The ethic of laissez-faire does not rank the two points. In its fullest form, it merely requires that the process of getting a distribution of utility be *mutually agreeable* (peaceful, uncoercive, voluntary). It doesn't require that any particular distribution be gotten.³ If Crusoe were a superb basketball player, for example, and Friday an avid fan, there would be nothing unfair by this standard about Crusoe's earning a million dollars a year by charging admission to Friday (and Friday's fellow fans) to see him make slam dunks. The final distribution of income and even of utility might be grossly unequal. But laissez-faire would not permit a redistribution of incomes that had themselves been achieved by just means. Justice in this view lies not in the distribution of income but in how the distribution is achieved.

Utilitarianism Chooses a Point on the Utility Possibility Curve

Over the past century and a half, however, an increasingly popular ethic alternative to laissez-faire has wished to go beyond life, liberty, and the pursuit of happiness. The ethic is that of utilitarianism, the doctrine that there is a social good achievable by balancing one person's happiness against another's. In other words, utilitarianism declares that the various distributions of utility between Crusoe and Friday can be ranked. Frequent talk of "the general will," "national interest," and "what's good for America" makes the declaration sound trivial. But it is in fact a controversial and significant moral step. The significance is that as soon as one takes it—as soon as one accepts the idea that Crusoe's utility can be weighed against Friday's and the two added up to get a number which is to be made as large as possible—then *the positions achieved by free exchange no longer have a special moral claim*. The government might now be justified in seizing a rich person's property to improve the lot of the poor. Or for that matter, it might be justified in seizing the poor person's property to improve the lot of the politically powerful and rich. The justice in it would be the achievement of a higher number for *social* utility. That Friday and his fellow fans freely chose to accept Crusoe's offer of a year of slam dunks in exchange for a million dollars is no longer very important. What is important is that Crusoe ends up rich and Friday poor, which is good or bad depending on how one values Crusoe's utility against Friday's.

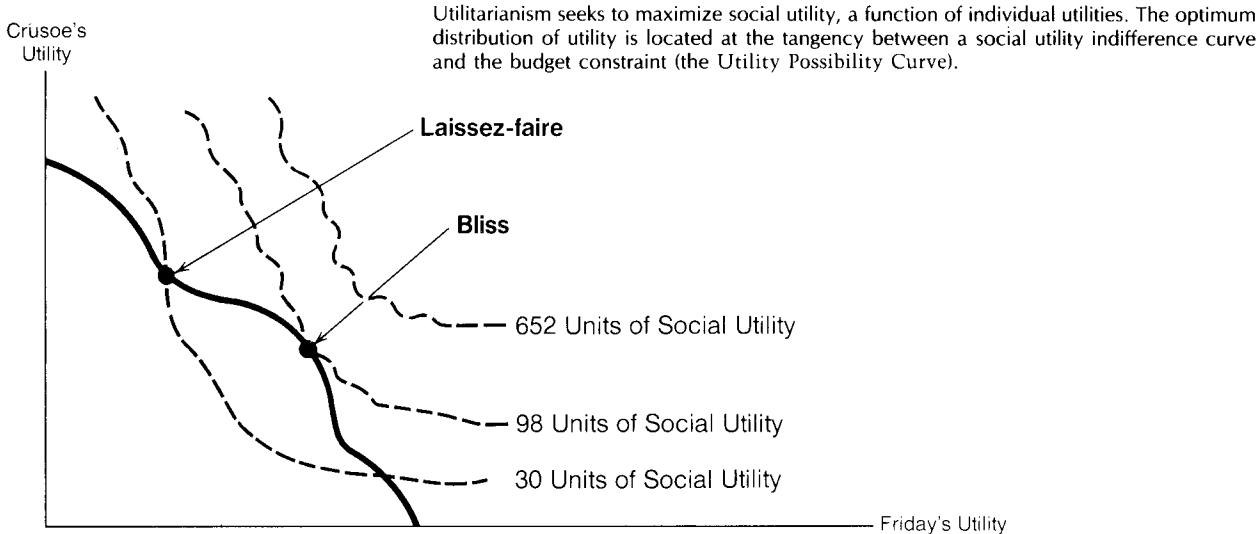
In other words, the diagram of utilitarianism adds contours of a social indifference curve to the diagram of the utility possibility curve, valuing Crusoe's utility against Friday's (see Figure 9.2). The best point is Bliss, the highest attainable contour within or on the utility possibility curve. The point Laissez-faire would only by a wildly improbable accident correspond to Bliss.

Shapes of Social Indifference Curves

The analogies with the theory of a single consumer's utility are exact. The case of perfect complementarity, for example, is the social indifference curve for an *egalitarian*: Any of Crusoe's income in excess of Friday's income is worth-

³ As given, for example, in Robert Nozick, *Anarchy, State and Utopia* (New York: Basic Books, 1974), from which the basketball example is drawn.

Figure 9.2
The Idea of Utilitarianism Can Be Represented by Social Indifference Curves



less, contributing nothing to social utility (see Figure 9.3). Notice that the axes measure income, not utility.⁴

Society's move from the point Poor to the point Inequal Rich has no effect on social utility by this standard. As the leading modern proponent of such indifference curves put it, society is "perfectly just when the prospects of the least fortunate are as great as they can be."⁵ At Inequal Rich they are not. A move to the perfect equality of Equal Rich improves the prospects of the least fortunate (Friday) and makes the increase in income socially useful.

The egalitarian curve is from one perspective generous: It is a generous millionaire who wishes to give away his millions. From another perspective it is envious: It is an envious pauper who begrudges every dollar of someone else's enrichment. The opposite indifference curves—either nongenerous or nonenvious depending on one's perspective—are analogous in the theory of consumption to perfect substitutability. According to these, a dollar of Crusoe's income is exactly *equal* in social worth to a dollar of Friday's (see Figure 9.4). Point Better is better than point Worse because total income in dollars is higher, even though the move impoverishes Crusoe. That is, such indifference curves say that things are better when income is higher *regardless of who gets it*.

⁴ The replacement of utility by income and the giving of moral weight to income depending on who earns it involves a logical flaw. Put briefly, it is that larger income for Professor North is not always better for him. If he gets the larger income by working much harder, then his net happiness may in fact fall. The larger income makes him happier but the harder work makes him less happy. "Net happiness" (that is, utility measured in dollars) must be on the axes, not income. This devastating criticism has not prevented people from using moral weights on income.

⁵ John Rawls, "Distributive Justice," in P. Laslett and W. G. Runciman, eds., *Philosophy, Politics and Society*, 3rd ser. (1967), reprinted in Edmund S. Phelps, *Economic Justice* (Harmondsworth, England: Penguin, 1973), p. 328.

Figure 9.3
Egalitarian Social Indifference Curves Have Corners at the Line of Equality

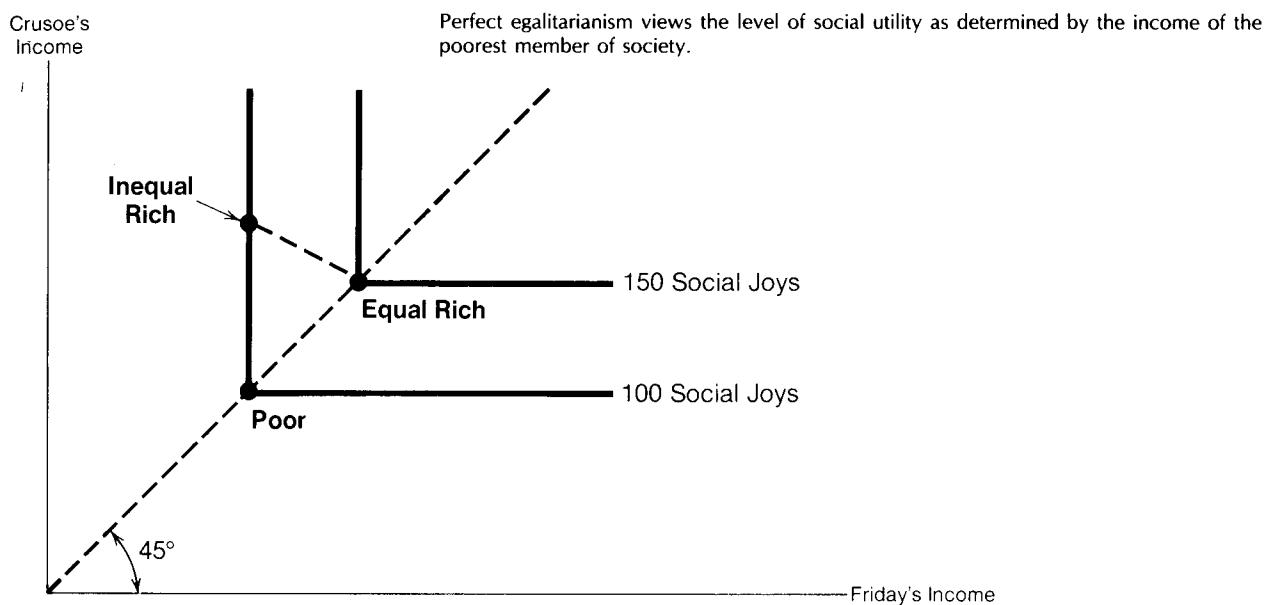
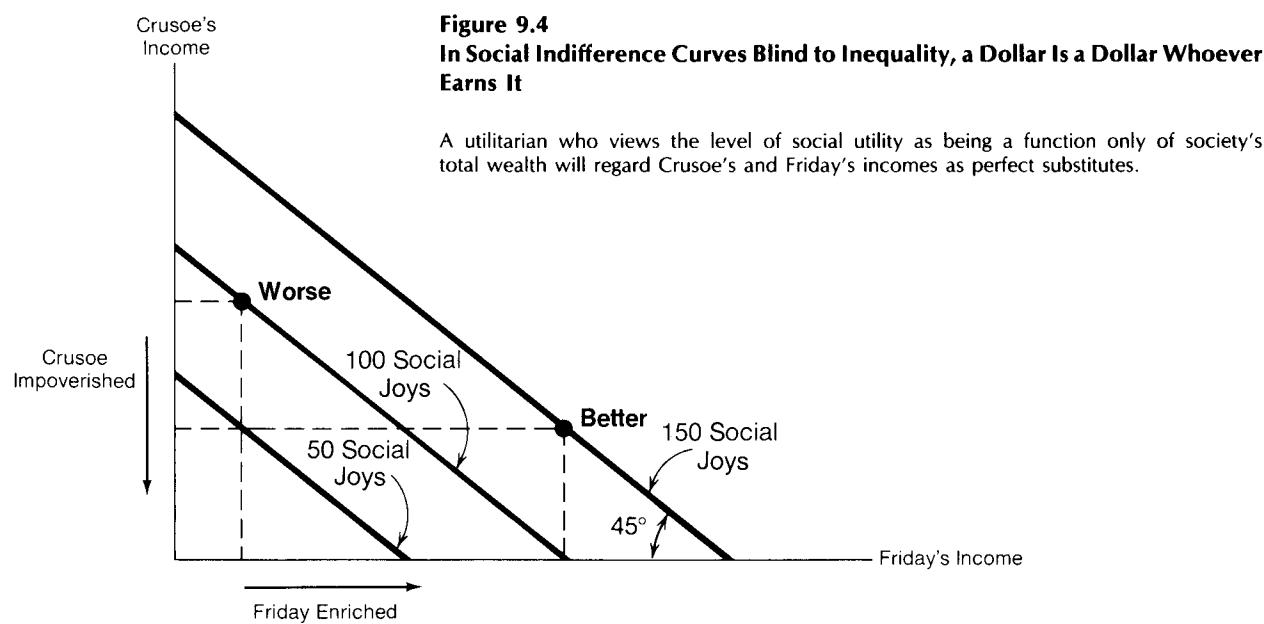


Figure 9.4
In Social Indifference Curves Blind to Inequality, a Dollar Is a Dollar Whoever Earns It



Laissez-faire Demands More Than a Dollar-Is-a-Dollar

At first glance, the dollar-is-a-dollar curve appears to be identical to laissez-faire, for if markets work perfectly then laissez-faire will also lead to maximum income. All the resources in the society could be owned by slothful dolts uninterested in running the economy in a productive fashion but would come to be managed by whomever could use them the most efficiently. Valuing resources the most is achieving the most income. Therefore, laissez-faire, like the dollar-is-a-dollar curve, recommends maximum income.

Laissez-faire, however, requires further that income be achieved by free exchange, not by theft or taxation. By contrast, the dollar-is-a-dollar indifference curves would register a social gain even if higher income were achieved by forcing people into collective farms by violence. The very idea of social indifference curves, to repeat, is hostile to laissez-faire, because such curves value "end states," not methods of achieving them. Laissez-faire concerns itself with how people behave (lawfully, uncoercively, and so forth), not with how they end up (poor, unequal, and so forth).

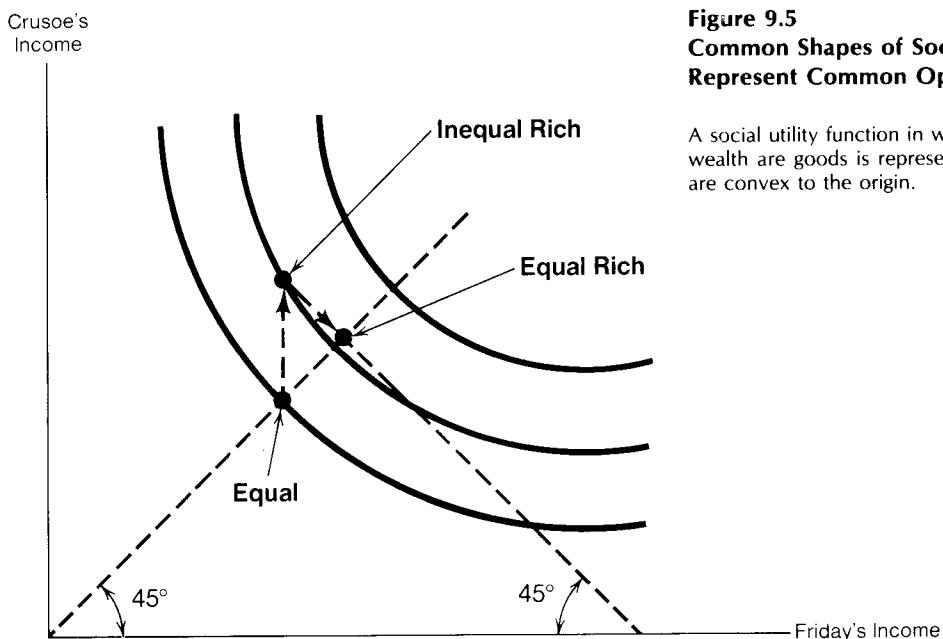
Typical Curves

The social indifference curves that most people nowadays carry in their heads are described as follows.

Q: Draw typical indifference curves between incomes of Friday and of Crusoe for an observer who to some degree values equality (but is not a perfect egalitarian).

A: See Figure 9.5. Because the observer presumably does not care whether it is Friday or Crusoe who gets the

higher income, the curves are symmetrical around the 45° line. A move equal to Inequal Rich now *is* socially useful; but so is the subsequent move to Equal Rich. Both equality and enrichment have value. For any symmetrical, convex set of indifference curves, both are true.



**Figure 9.5
Common Shapes of Social Indifference Curves
Represent Common Opinion**

A social utility function in which both equality and total social wealth are goods is represented by indifference contours that are convex to the origin.

Summary "Welfare" economics is the economist's own little contribution to moral philosophy. The economist could if he wished abstain from offering an opinion on the happiness of society, just as an engineer abstains from offering an opinion on the desirability of a bridge that she is asked to build. The economist's opinion, it is often claimed, is merely an opinion, a matter of taste. Yet "normative" (what-should-be) economics, like art criticism or ethics, makes the taster more wise about tastes. For example, a taste for laissez-faire (that is, "leave alone," do not use the government to interfere) implies that one favors allowing people to trade to an efficient point. This is a point on the utility possibility curve, getting the most utility possible for Crusoe given some level of Friday's utility. Laissez-faire makes no choices between one efficient point and another. Indeed, it forbids the government from intervening to make such choices.

An alternative taste is embodied in social indifference curves, that is, a characterization of how one feels about Crusoe's utility relative to Friday's. The assertion that comparisons like this should be made is called *utilitarianism*. It includes a wide range of opinion on the social good, from *egalitarianism* to a dollar-is-a-dollar-regardless-of-who-earns-it. The analogy with the contrasting shapes of individual indifference curves for complements and substitutes is exact. But the more fundamental contrast is between laissez-faire and utilitarianism, between economic morality based on entitlement (the justice of how a distribution was achieved) and economic morality based on end states (the justice of what a distribution looks like in the end, however achieved).

EXERCISES FOR SECTION 9.1

1. Which of the following assertions are normative, positive, or mixed normative and positive? Why?
 - a. Thou shalt not kill.
 - b. Falling bodies accelerate at 32 feet per second per second of fall.
 - c. "Miracle wheats" raise yields.
 - d. Miracle wheats are good for India.
 - e. The Law of Demand applies to the market for autos.
 - f. The Law of Demand is pretty, and therefore persuasive.
 - g. The Law of Demand is persuasive.
2. *True or false:* Once a society is on its utility possibility curve no one can be made better off.
3. *True or false:* Exponents of laissez-faire want the highest possible income for society.

PROBLEMS FOR SECTION 9.1

- o 1. Suppose that John D. Rockefeller achieved his immense wealth by providing good service at a low price, not (as you may have heard) by underselling the competition and then raising prices to monopoly levels. Suppose that he passed on his wealth to his children and grandchildren by legal and ethical means. Fabulously wealthy as they are, should the present Rockefellers be made less wealthy?

True or False

2. An end-state theory of justice could imply that it was morally admissible to compel Crusoe to perform slam dunks for Friday and his friends or to force Crusoe to pay taxes for the support of Friday.

Sources of Injustice

9.2 National Income and Its Ambiguities

(A) Start w.
Chicago headline
p. 187

What to Read For Can we measure increases in happiness? What is *real income*? How is its change measured? What about differences of income from one place to another? What is the *index number problem*? Is there a solution to the problem? What are Laspeyres and Paasche indexes?

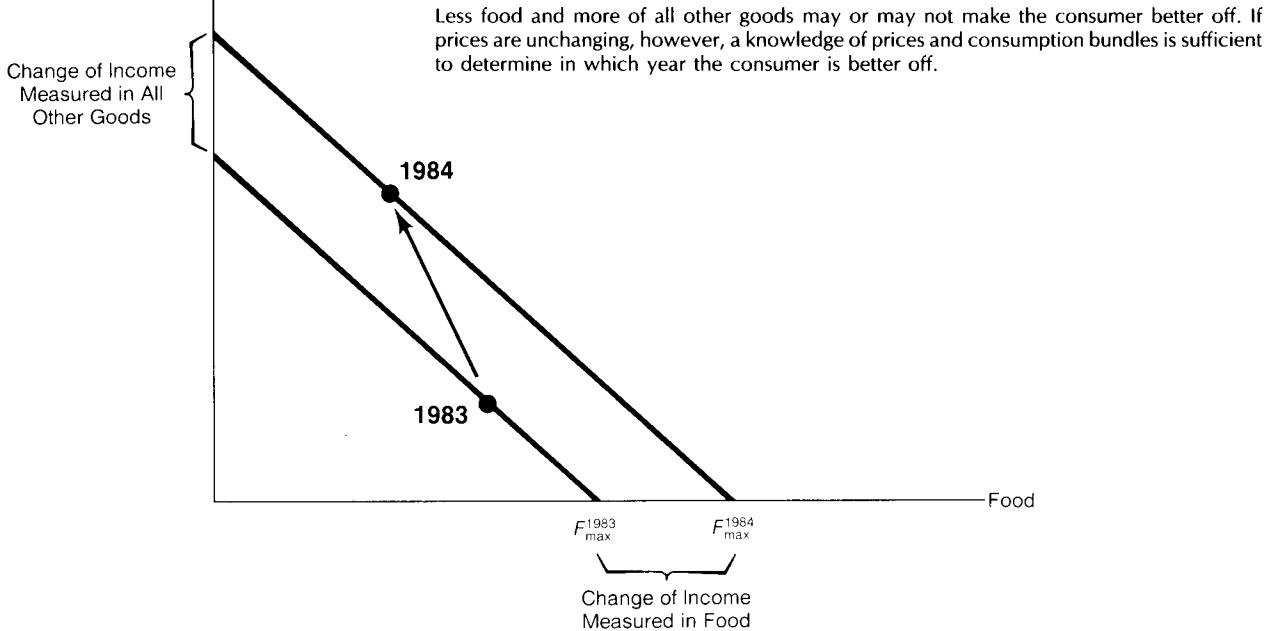
The Problem of Valuing a Shifting Bundle of Goods

Social indifference curves asserting that a dollar-is-a-dollar-no-matter-who-earns-it are the simplest imaginable ones. Therefore, they play a special role. They achieve simplicity by merely sidestepping the difficult ethical problem that economies are made up of more than one person. They amount to reducing the measurement of society's happiness to the happiness of the typical person.

Consider, therefore, the question of whether David Galenson is better off in 1984 than in 1983, given the knowledge of the bundles of goods he consumes each year. If his utility function for food, housing, automobiles, books, and so forth were known, then one could simply insert the two bundles into the function and read off which gave more joys. But in practical terms his utility function is not known; only his observed bundles and the market prices he faces are known. If his consumption of every single good has increased between 1983 and 1984—gasoline, hot dogs, pencils, and everything else—the question of

Applies to all kinds of things. Every - thing faces an Aggrega- tion Problem

Figure 9.6
Ambiguity and Its Resolution in Measuring Real Income



whether he is better off in 1984 is not very challenging. Obviously he is better off.

But if his consumption of some goods has increased and of others decreased, the question is not so simple. Refer to Figure 9.6. The question is, what has happened to his *real income*, that is, to his ability to get goods?

In terms of food his real income has decreased from 1983 to 1984 since the amount of food has decreased; in terms of all other goods it has increased. Simply adding together the tons of the numbers of foods and all other goods to solve the problem of weighting the increase in the one against the decrease in the other would be silly. Two oranges plus one living room couch equal nothing in particular. Strangely, this is how such problems are commonly solved in other contexts, such as that of measuring the amount of crime. The FBI's "index crimes" are all crimes of a serious nature and the number of them (one murder = one auto theft) divided by the number of people is "index crimes per capita." That's not a sensible way of adding up crimes.

**The Solution:
Relative Prices**

Economists have the advantage over other observers of society that they set in front of their noses a more sensible way of weighting one thing with another to form an aggregate (of crime or power or, in this case, income): the relative price of the two. The solution is attractive because prices do measure the marginal value consumers put on, say, one unit more of food relative to all other goods, because consumers are on their budget lines. Weighting goods by their prices can be viewed, in fact, as a crude (very crude) approximation to a utility function. And in any event it permits the two goods to be expressed in one intelligible unit of real income.

If the relative price of food and all other goods has not changed between 1983 and 1984 the procedure is easy. One can simply express the bundle in terms of one of the goods, say, food, by adding the amount of food consumed to the product of the price of all other goods in terms of food and the amount of all other goods consumed, as was done in Figure 9.6. In the case pictured, income has increased over the year. The 1984 bundle is (revealed) preferred to the 1983 bundle, and real income therefore has increased. Indeed, one can say *how much* real income has increased. It has increased by the distance bracketed on the food axis (or, in terms of all other goods—which can be translated back into food by multiplying by the price—by the distance bracketed on the all-other-goods axis).

In the simple case of one person, then, the answer to the question "Can you detect an increase in happiness?" is "yes." The logic is that of revealed preference. The question of the amount that happiness has increased is not, however, solved by this measure of the amount in dollars. That Galenson has 30% more dollars or the goods those dollars buy in 1984 than in 1983 does not mean he is 30% happier, rising from, say, 100 to 130 joys. But the weighting of goods consumed by their prices does at least measure whether and by how much the budget line has moved outward. That's as good as we can do.

**Applications of the
Measurement of
Real Income**

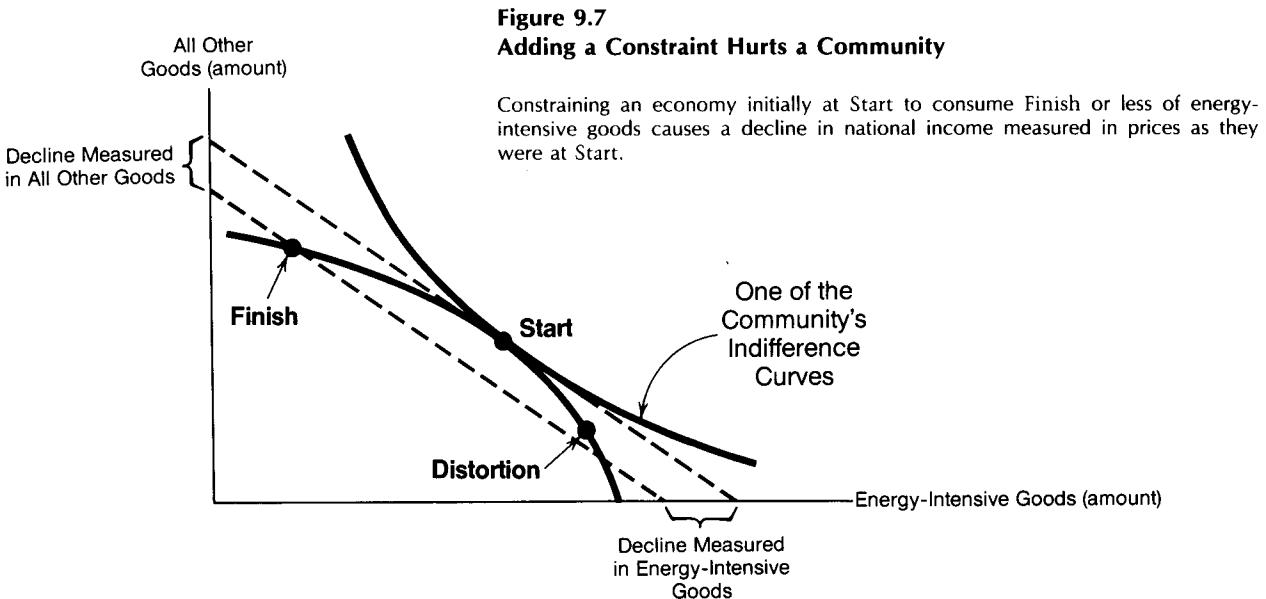
The idea of measuring the rise or fall of the budget line of the nation as though it were the budget line of a single person is a powerful one.

Q: Suppose that an economy produces and consumes two goods, an energy-intensive good and all others. The sages of the community determine that too much of the energy-intensive good is being produced and order its production cut. Illustrate how income declines in a diagram of the community's production possibility curve and indifference curves.

A: The community starts at Start (but see the comment following and refer to Figure 9.7). It finishes at Finish, the edict requiring that the amount of the energy-intensive good fall by this extent. Measured in the starting prices the finishing income is the lower dashed line. The decline in income must be expressed in one or the other good, but in either it declines. Saving energy is bad.

COMMENT The implied criticism of the policy of cutting the output of the energy-intensive good is valid if the start is at Start. But if the economy is indeed producing too much energy-intensive goods by the economy's criterion of a divergence between supply price and demand price—as at the point Distortion, perhaps because energy is underpriced (as a result, say, of another, earlier edict)—then an edict cutting the output of the energy-intensive good can be good for the economy.

The arithmetic of national income is the arithmetic of budget lines. The budget line is lower or higher depending on whether the sum of expenditures at some set of prices is lower or higher. In the United States over the past century, for example, the production per person of petroleum, paper, and rubber increased, while that of whale oil, hemp, and homemade tables decreased. To determine whether or not national income per person increased on balance, one must add up at the beginning of the century and at the end of the century the value of all goods *at some constant set of prices*. The outward movement of parallel (same-price) budget lines in a diagram corresponds to a higher value of consumption possibilities, or real income. The American sum expressed in 1975 prices



rose from about \$830 in 1840 (a real income per head roughly comparable with present-day Mexico) to about \$7000 in 1975. Although output of buggy whips and flintlocks declined, the average budget line as a whole moved out by a factor of about 8.4.

**An Application to
Comparisons
Between Two
Countries**

Q: In the first decade of the twentieth century, the average working-class family in England and in America consumed bread, beef, and pork in the amounts per week given in Table 9.1. Americans, in other words, consumed more meat but less bread than did Englishmen. Suppose (contrary to fact) that bread, beef, and pork were the only goods consumed. Prices in England were 1.25 pence per pound for bread, 6.75 pence per pound for beef, and 8 pence per pound for pork. Which bundle—American or English—was best? By how much?

A: The weekly “income” (so to speak) of Englishmen is the amount they spend on the three goods, that is, 1.25 pence per lb \times 22 lb + 6.75 pence per lb \times 4.5

lb + 8 pence per lb \times 0.5 lb = 27.5 + 30.38 + 4 pence = 61.9 pence. Similarly, the weekly income of Americans in *English prices* would be the American bundle evaluated at English prices; that is, (1.25) (8.25) + (6.75) (6.75) + (8) (2.25) = 73.9 pence. The American income is higher than the English income since 73.9 pence is higher than 61.9 pence: The American family consumed 12 pence worth more per week of bread, beef, and pork taken together. To put it another way, the relatively high American consumption of meat more than offsets (in view of the value that consumers put on meat relative to bread) the relatively high English consumption of bread. On balance, Americans ate $(73.9/61.9) - 1 = 19\%$ more.

Table 9.1
**Weekly Consumption of
Bread, Beef, and Pork by
an Average Working-Class
Family in England and
America c. 1900–1910**

	Pounds of:		
	Bakery Bread	Beef	Pork
England	22.0	4.5	0.5
America	8.25	6.75	2.25

Source: Great Britain Board of Trade, *Report of an Inquiry . . . into Working Class . . . Retail Prices*, Cd. 5609, British Parliamentary Papers 1911, vol. 88, pp. lxvi–lxvii.

Notice that simply adding up pounds of meat and bread would be foolish—the equivalent of adding one murder to one car theft to get a measure of total crime. If you tried to do this, you would find that by the pound “price” (one pound of sirloin equals in value one pound of bread), Englishmen are better off by $(22 + 4.5 + 0.5)/(8.25 + 6.75 + 2.25) = 27/17.24 = 1.57$, or 57%, and they could be made even better off by substituting still more bread for beef in hamburgers. But that’s nonsense.

To be sure, there are prices alternative to the market price that make some sense (if not cents), for example, calories. For some purposes of comparison, such as comparing body weights or the work a person can do, the caloric “prices” of 1250 per pound for bread, 1600 for beef, and 1800 for pork are relevant.

For comparing the ability to satisfy desires, however, the prevailing relative prices are the only relevant ones.

The Index Number Problem

D.
But it's
technical
problem
+
Not just over
time:
over space

A very serious problem in all of this, however, is that prevailing relative prices change from one year to the next. To return to Galenson consuming food and all other goods in 1983 and 1984, the question is: Which relative prices does one choose to weigh the goods, 1983 or 1984? There is no answer. Or to put it another way, one is free to choose either year's prices. And relative prices might change so much that using one year's prices would not give even the same *direction* for the change in real income as would using the other. If the price of food rises enough between 1983 and 1984, for example, valuing the two bundles at the two different relative prices gives two measures of real income, one rising and the other falling (see Figure 9.8).

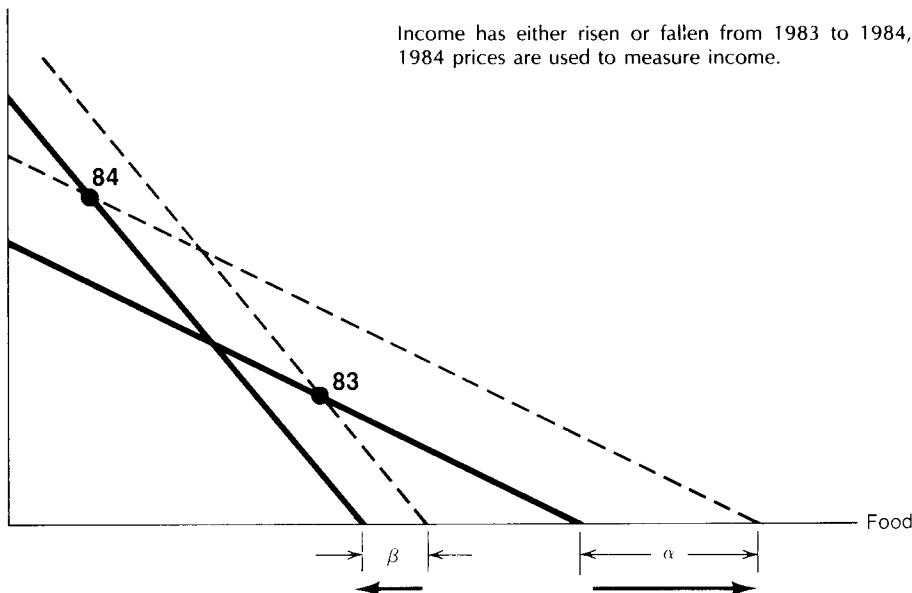
The α is the rise in income (expressed in real terms, in this case physical amounts of food) and β is the fall. Notice that the ambiguity comes out in revealed preference. At the prices actually observed in the years (the solid lines), neither the 1984 nor the 1983 bundle is revealed preferred to the other. The ambiguity is *the index number problem*.

The problem is that an assertion made from one point of view (from the point of view of 1984 prices real income fell between 1983 and 1984) may be false from another point of view (namely, the prices of 1983). It is most serious when prices change greatly, as when comparisons of real income are being made over long stretches of time or between two very different countries. The English-American comparison, for example, used English prices to evaluate the bundles of bread, beef, and pork. The American bundle was 19% larger by this standard.

All Other Goods (amount)

Figure 9.8
With Different Price Weights, Ambiguity Reinstated

Income has either risen or fallen from 1983 to 1984, depending on whether 1983 or 1984 prices are used to measure income.



But suppose that one had used American instead of English prices to evaluate the bundles in the two countries. Bakery bread, as it happens, was much more expensive relative to meat in America than it was in England, which would improve the relative position of the bread-rich English bundle. Indeed, it turns out that from the American point of view the English bundle was better, not worse (as the earlier point of view implied), than the American bundle. The calculations are, using the amounts in Table 9.1 and the American prices (2.79 pence per pound for bread, 7.00 pence per pound for beef, 6.50 pence per pound for pork):

$$\begin{aligned}\text{Value of English bundle} &= (2.79)(22 \text{ lb of bread}) \\ &\quad + (7.00)(4.5 \text{ lb of beef}) \\ &\quad + (6.50)(0.5 \text{ lb of pork}) \\ &= 96.13\end{aligned}$$

$$\begin{aligned}\text{Value of American bundle} &= (2.79)(8.25 \text{ lb of bread}) \\ &\quad + (7.00)(6.75 \text{ lb of beef}) \\ &\quad + (6.50)(2.25 \text{ lb of pork}) \\ &= 84.89 \text{ pence}\end{aligned}$$

By this reckoning the English bundle is 13% [that is, $(96.13/84.89) - 1$] *better* than the American, not 19% *worse*.

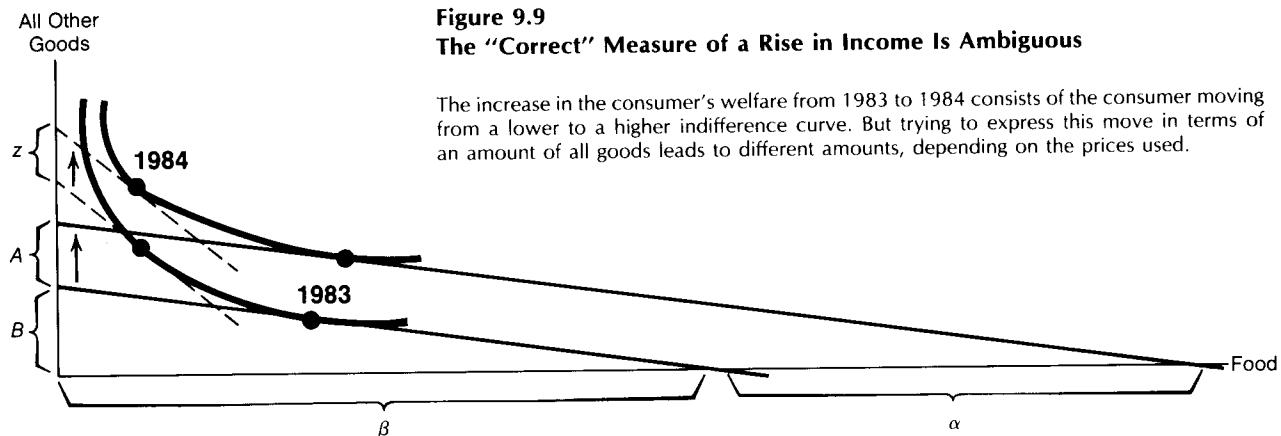
Taking a different point of view reverses the direction of superiority. And whether or not the superiority is reversed, it is in any case a different magnitude when different prices are used. Nothing could be more obvious yet more important. The index number problem is a numerical version of moral relativity. Different valuations produce different judgments of good and bad.

***Laspeyres Uses
Early, Paasche Late
Weights***

The index number problem can be given an expression in some jargon that will prove useful later. An index of Galenson's income is said to be a *Laspeyres index* if it weights the quantities of food and all other goods by *early* prices (that is, those of 1983 rather than 1984); it is said to be a *Paasche index* if, by contrast, it uses *late* prices (that is, 1984). The jargon of Laspeyres and Paasche commemorates two pioneers of index numbers. Laspeyres is pronounced as though it were spelled in English "la spairee" (the second word rhymes with "scarce"); Paasche rhymes with "squash"; you can keep in mind which is which by noticing that "L" comes before "P" in the alphabet and that the Laspeyres index uses weights that come before Paasche's.

The usefulness of the jargon is that, in making the simplest of all comparisons, between two bundles, one must choose either early prices or late prices, and it is nice to have distinct tags for each.⁶ Most indexes you will read about in the newspapers are in fact Laspeyres, early weighted indexes because it would be inconvenient to change the point of view as each year became the later year. For example, real national income (or "real GNP" or "real income" or "income in constant dollars") was for a long time measured in the United States in the prices of 1972. A comparison of real income per head in 1980 and in 1972 would be a Laspeyres comparison: It views products from the relative prices of 1972, not of 1980.

⁶ Mature reflection may suggest to you that in this case as in others the only merit of learning the jargon is that economists use it (rather than the simpler "early weighted" and "late weighted"). Such cynicism is a common accompaniment of maturity.



(E)
F L + P
indexes of
anything over
time:
early +
late

Laspeyres and Paasche indexes differ and neither can be said to be the correct index. Indeed, "correct" in this context has a Laspeyres and Paasche ambiguity of its very own. The notion of a correct index of, say, Galenson's income between 1983 and 1984 requires indifference curves running through the bundles consumed in the two years. The percentage increase in income is then the difference between the two indifference curves, divided by the level. Using the prices of 1983 in Figure 9.9, for example, the rise in real income is the vertical distance A —the distance between the 1983 budget line and a budget line that at the 1983 prices would just permit Galenson to achieve the happiness he actually achieved in 1984. The percentage change is A/B (notice that α/β measured horizontally is in the same ratio; you can measure the change in income along either the food or the all-other-goods axis).

But one might just as well use the prices prevailing in 1984 as the standard for constructing the tangency, using a Paasche instead of a Laspeyres measure of the distance between the two indifference curves. In this case the vertical distance Z , not A , would be the rise in real income. To put it another way, the "correct" indexes are measured from the budget lines of pure income effects, but there are as many definitions of the income effect as there are relative prices at which to compare two indifference curves.

Laspeyres and Paasche Are Bounds on the Truth

In any case, Figure 9.10 shows that that ordinary Laspeyres and Paasche indexes of income lie on either side of these "correct" measures. If you have calculated Laspeyres and Paasche indexes of the percentage change in Galenson's income between 1983 and 1984, you know at least that the "correct" measure is somewhere between the two. The Laspeyres difference in income is, you see, necessarily larger than the distance A ; likewise, the Paasche difference is necessarily smaller than distance Z . So the Laspeyres index is an upper bound on the true change in income and the Paasche is a lower bound.⁷

⁷ This bounding of truth, strictly speaking, depends on prior knowledge that the Laspeyres index of the growth of income is larger than the Paasche. Normally it is.

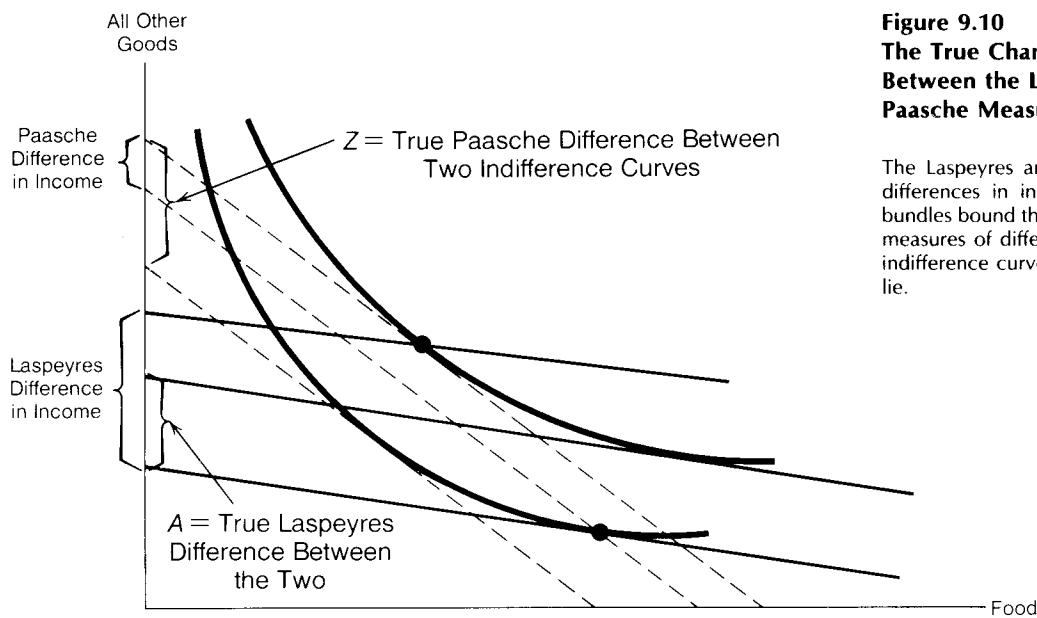


Figure 9.10
The True Change in Income Lies Between the Laspeyres and the Paasche Measures

The Laspeyres and Paasche measures of differences in income between the two bundles bound the Laspeyres and Paasche measures of differences between the two indifference curves on which the bundles lie.

Summary Talk of "America's well-being" involves treating America as one person. National income does. It is the sum of all expenditures by Americans on bread, rent, car insurance, movies, shows, schooling, roads, haircuts, and so forth. To compare two different years or two different places, the prices used to evaluate each thing must be chosen to be the same (that is, 1975 prices chosen to evaluate income in both 1975 and 1840, or British prices chosen to evaluate both British and American expenditures). Comparing two bundles, therefore, involves comparing two parallel (same-price) budget lines. The higher one is the larger sum. National income is simply the nation's budget.

But the choice of weights that must be made to add up quantities or prices into "real national income" or "the general price level" is arbitrary. It can be early or late weights (Laspeyres comes before Paasche in alphabetical order) or some entirely new set of weights. The choice, however, must be made: God or nature or Daddy does not make it. The index number problem, in other words, is insoluble. You must adopt a point of view to look at something. It is childish to suppose that the looking is uninfluenced by the point of view.

EXERCISES FOR SECTION 9.2

1. Rank these by increasing difficulty in measurement and explain: Galenson's change in happiness 1983–1984; Galenson's change in money (dollar) income 1983–1984; Galenson's change in real income measured in units of food 1983–1984.
2. Why in terms of income would the economy in Figure 9.7 be made better off by reducing the consumption of energy intensive goods if it started from Distortion?
3. In which of these comparisons of real income will there be an index number problem (that is, different results from different sets of prices used to weight the two bundles)?

- a. Real income in the United States in 1984 compared with real income in the Soviet Union in 1984.
- b. Russia in 1900 compared with the Soviet Union in 1939.
- c. Russia in 1900 compared with Russia in 1200.
- d. Moscow in 1984 compared with Kiev in 1984.
- e. *Apparatchiki* (officials of the Soviet Communist Party—they have access to special stores in which goods are sold at lower prices and many goods are available that are not elsewhere) compared with ordinary Soviet citizens.

PROBLEMS FOR SECTION 9.2

1. The production of grain (wheat, corn, rice) in the world in the early 1970s was about 1 billion metric tons selling at \$170 per metric ton; the production of all other goods was 3200 billion units selling at \$1 per unit. *True or false:* Since world income expressed in metric tons of grain was 19.8 billion metric tons of grain, the world could in fact have produced that much grain if it wanted to.

2. Comment on this headline in the *Chicago Tribune* for July 11, 1979: "Murder up 17%, rape 20%, but total crime off 4% here."

- 3. Bread, beef, and pork have natural units—pounds or loaves or whatever. But the method of comparing budget lines (that is, index numbers) applies also to whole classes of commodities, such as manufactures, that have no natural units. Between 1859 and 1874 in the United States, manufacturing output increased by 40% per capita, while agricultural output decreased by 8%.⁸ If the quantity of manufactured goods in 1859 is expressed as 100 manufacturing units selling at \$1 per unit and the quantity of agricultural goods as 168 agricultural units selling at \$1 per unit, how large was the percentage increase of all commodity output (manufacturers plus agriculture) from 1859 to 1874?
- 4. The median income of families in the United States was \$7000 in 1965 and \$13,700 in 1975. The consumer price index went from 94.5 in 1965 (the year 1967 is defined to be 100) to 161.2 in 1975. What was the real percentage increase in median income?

True or False

- 5. Opening trade increases national income.
- 6. If John Stuart's salary rises from 1985 to 1986 in the same proportion as does a Laspeyres index of his cost of living, then according to revealed preference he is better off after the rise.

9.3 The Economics of Politics

What to Read For Is there a general will in society? Is majority vote always a good thing? Why does the judgment that a project is good require unanimous agreement? Is bribery and buying of votes always bad? How do such activities get around the problem of unanimity? But what then is the problem of extortion? And would people tell the

⁸ These and the statistics following are taken from Robert Gallman, "Commodity Output 1839–1899," in National Bureau of Economic Research, *Trends in the American Economy in the Nineteenth Century* (Princeton, N.J.: Princeton University Press, 1960), pp. 46–48, 54, 56.

truth about valuations of public goods? Can one measure social gain created by a project without making an implicit judgment about the desirability of the income distribution?

Voting Is No Guide to the General Will

If everyone were a utilitarian and had precisely the same social utility function—say, “maximize national income” or “maximize the prospects of the least fortunate”—then choices of government policies would reduce to mere calculation. The Council of Economic Advisors or the Lord High Economist would calculate the economic impact of a quota on Japanese steel (as against no quota), a deregulation of airlines (as against continuing the Civil Aeronautics Board), closing of the Naval Supply Station at Norfolk, Virginia (as against keeping it open), or a 5-cent rise in cigarette taxes (as against no rise) and then insert the implied change in the distribution of happiness among Tom, Dick, and Harriet into the social utility function. The Council would then choose the alternative giving the greater social happiness, to the unanimous applause of the citizens.

But for better or for worse, no country has such unanimity. In other words, treating America as one person for purposes of measuring national income is obviously not acceptable. John and Laura may each have a social utility function, but they may not have the same function. Opinions about the social good differ. People disagree. Politics exist.

An enthusiast for social engineering might reply that this very politics will solve the problem of disagreement, because voting on the deregulation of airlines or the closing of the Norfolk Naval Supply Station reveals the general will. The reply sounds persuasive, but it is wrong. The arguments showing that it is wrong were discovered two centuries ago by a French philosopher, the Marquis de Condorcet. He pointed out that majority voting does not always lead to a decision, and when it leads to a decision does not always properly reflect the opinion of the voters.⁹

It is indeed not very surprising that a collection of people with varying tastes and interests cannot always agree. What is surprising is the gullibility with which people swallow the idea of the general will, identified as the simple majority or the vote of two-thirds of the Senate or the edicts of a man on a white horse. Cavaliers and roundheads, democrats and fascists cannot be expected to agree. John Adams estimated that fully a third of the people in the American colonies during the Revolution were loyal to the British crown.

**Market Solutions:
The Theory of Dollar Voting**

Economists are skeptical of the moral claims of majority rule. Someone will be outvoted, and there is no reason to give the minority’s dissatisfaction with the outcome less moral weight than the majority’s satisfaction. A decision to jail all communists would have pleased most Americans during the 1950s, yet would have displeased supporters of the Bill of Rights, not to mention the communists themselves. A decision to soak the rich is ever popular, but it ignores the rights of the rich to their riches. The decision during World War II to put 120,000 Japanese-Americans in concentration camps was acceptable to most voters (of non-Japanese descent), Congress, the Secretary of War, President Roose-

⁹ See Keith M. Baker, ed., *Condorcet: Selected Writings* (Indianapolis, Ind.: Bobbs-Merrill, 1976), pp. 52–53, and K. J. Arrow, “Values and Collective Decision-Making,” in P. Laslett and W. G. Runciman, eds., *Philosophy, Politics and Society*, vol. 3 (1967), reprinted in Phelps, ed., *Economic Justice*. The best elementary treatment of the subject and the others mentioned in this section is Dennis C. Mueller, *Public Choice* (Cambridge: Cambridge University Press, 1979).

velt, and even the Supreme Court, but such a weight of opinion against the Japanese-Americans is no guide to justice.

The economist's usual guide is unanimity. If literally everyone agrees that a new road should be built from the main highway to Great Durnford, then it should be built. If everyone has a veto over the taxing and spending by the village government to build the road, then the case is similar to a free trade between two people, because the people affected, like the two people trading, can enter or not enter the deal voluntarily. A new road or other project is desirable if it makes someone (or many) better off without making others worse off, that is, if the project is efficient.¹⁰ And in such a case, setting aside mere spite, everyone would agree to let the project go forward.

Unanimity is an absurdly strict test of a project's worth if no side payments, bribes, or vote trading are allowed. The new road could be a social bonanza, bringing a previously isolated village into contact with the wide world, enriching and enlightening everyone in the village—except the mule skinner, whose pack train was previously the only way out of the village and who is impoverished by the new road. He by himself can veto the road if unanimity is required, denying others the great benefits.

The solution, however, is simple. Allow the rest of the village to buy off the mule skinner. If the rest of the village gains from the road enough to share some of the gain with the mule skinner (enough to make him happy with the outcome) and yet still has some net gain left over, then the project is mutually advantageous and should go forward. Likewise, a rule of unanimity—or, for that matter, even a strict rule of majority—would paralyze Congress if its members were somehow forbidden to trade votes. But with vote trading allowed, Congress is able to pass hundreds of bills a year, for better or for worse. The practice has bad-sounding names: vote trading, graft, vote buying, bribery, venality, logrolling (you help roll my log, I'll help roll yours). But in fact the results can be good, selling a bill to the highest bidder and using the money collected to compensate losers in the bidding. Voting without vote trading enriches the winners at the expense of the losers; voting with vote trading compensates the losers at the expense of the winners.

Bargaining in good faith under a rule of unanimity, then, makes politics into a market. It solves the problem of discerning the general will: The general will is whatever money can buy. It suggests that, if democracy (the rule of the people) is to be fair to minorities, it must become plutocracy (the rule of the rich), with the losers in elections by dollar votes being compensated by the plutocrats. Obnoxious as the argument may seem when put plainly, it does at least in the style of all such market arguments ensure the largest happiness from a given distribution of money (that is, of power).

It has some merit, furthermore, as a description of real politics—or, rather, *realpolitik*. The payment to losers need not be coin of the realm. Farmers wanting the votes of a shipbuilding district for a farm subsidy may bribe the senator or representative involved by promising to vote for a Navy contract to build two submarines in American yards. Indeed, the payments need not be selfish. If the people to be bribed, the shipbuilders, are for some unselfish reason filled

¹⁰ Or in fancier language, *Pareto efficient* or *Pareto optimal*. The only point of the fanciness is to commemorate the role of the Italian sociologist and economist Vilfredo Pareto (1848–1923) in framing the idea clearly.

ASK:
How to decide
on amount
of aid to
countries.

The Vertical Addition of Marginal Valuation

with enthusiasm for the Salvation Army, the bribe may be a money contribution or spiritual support for the Salvation Army. Like ordinary exchange, with which it is now comparable in every detail, political exchange need not be direct to achieve efficiency.

The argument can be made more explicit by introducing the curve of marginal willingness to pay for roads on the part of the citizens of Durnford. An eight-lane superhighway is at one extreme of road quality, the blazing of a mere track through the forest at another, and the citizens have marginal valuations of road quality in between. For the mule Skinner the marginal valuation may be negative; for the rest of the citizens it is positive but diminishing as more is acquired. Suppose that the road is a commonly used consumer good in the sense that no one can be excluded from benefiting from it once it is built. National defense, public parks, police protection, and knowledge are similar goods. The society is unable or refuses to charge admission person by person to these services. Each person in the community gets the one level of service decided by the community as a whole.

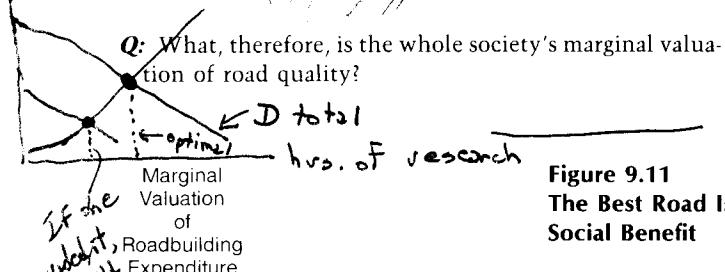
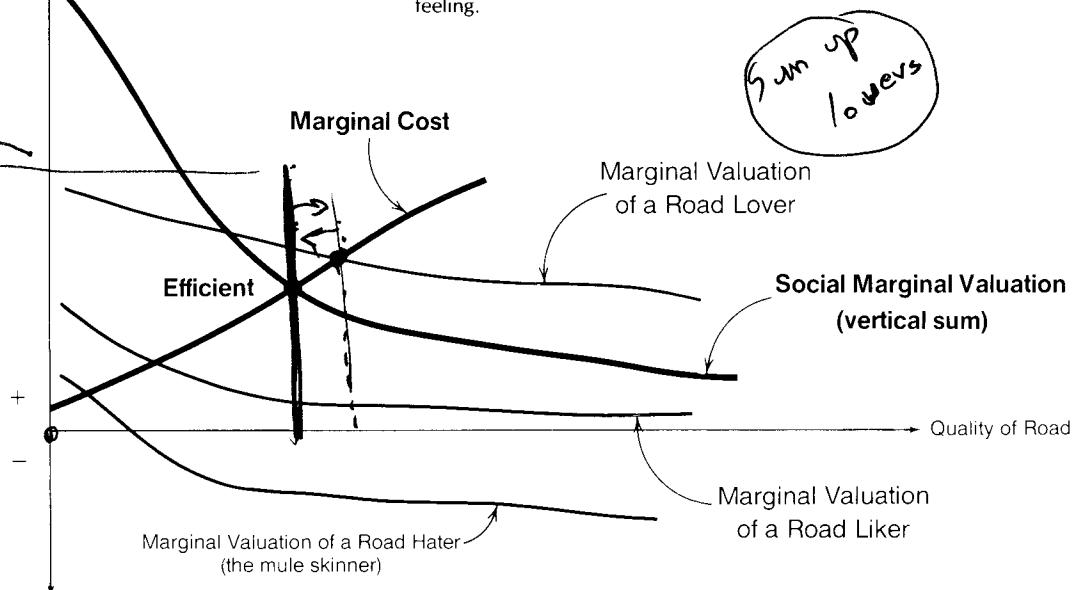


Figure 9.11
The Best Road Is the One Whose Marginal Cost Is Equal to Its Marginal Social Benefit

The social marginal valuation is represented by the *vertical sum* of the marginal valuations of three citizens: a road lover, a road liker, and a road hater. The point Efficient is the one the society should go to if it accepts marginal evaluation as a relevant measure of feeling.



sumed good is a good that *jointly supplies* services to each member of the community. Its demand curve is determined by vertical, not horizontal, addition.

Q: What is the efficient level of road quality?

A: It is the amount at which the marginal cost of road quality rises up to meet the diminishing marginal valuation of quality by the community as a whole, the point Efficient in Figure 9.11.

If citizens were able to cast dollar votes, in short, the right amount of road would get built. If politics is transformed into a perfect market, it works like a perfect market: efficiently.

Flaws in the Practice of Dollar Votes: Extortion

Very well, if citizens would give and accept honest graft under a constitution requiring the unanimous consent of the governed for every public act, then no one would be coerced by the majority or cheated by the minority. Supposing for a moment that the original distribution of dollar votes were thought to be good (or that it could be made good by redistribution), then the voluntary reshuffling of dollar votes in compensation for losses would also be good.

As a description of politics in democratic countries this picture is useful, within limits. Citizens with very strong feelings about the military draft or tuitions at public universities can and do band together into pressure groups to bid for the attention of Congress in perfectly honest but expensive ways, casting dollar votes as though political votes were literally for sale. The flaws in the picture as a description are political convictions and the costs of banding together. The costs may be low for a group of American television manufacturers seeking a tariff but high for the much larger group of American buyers of television sets hurt by the tariff. In the absence of such flaws one could affirm that whatever public policy is undertaken is in fact the one preferred, on balance, by the people casting *dollar* votes.

But the charming picture of winners compensating losers in the economic game has flaws even as a prescription of how society should look. The nature of the flaws is suggested by the rarity of its central element, unanimous consent. Criminal trials by jury, certain club memberships, and a few other social decisions are made unanimously. All other decisions require something less than everyone's consent, from a two-thirds majority in the Senate to ratify a treaty down to Idi Amin's solitary consent to execute his enemies. Why?

In Idi Amin's and like cases, the answer is brute force. He wishes it to be so. But even in free countries the usual procedure is to coerce by outvoting the few opponents of a treaty or a police action or a tax. This is odd. Surely, one would suppose, well-wishers of humankind would urge that unanimous consent be required before a superhighway is built through a city neighborhood or before income taxes are imposed to pay for national defense. But the reason the well-wishers do not so urge is clear from the examples. One cantankerous householder (and therefore all householders) could stop a highway, claiming that her love for her house in its present location was worth \$6 million to her. The point is familiar from the chapter on exchange. Since one's vote is essential, a rule of unanimity puts each person in the position of bargaining with the rest of the community. Unanimity encourages extortion.

Flaws When Commonly Consumed Goods Are Present: Lying

Furthermore, even if the procedure were not literal unanimity but merely dollar contributions for a commonly consumed good, there would be the problem of lying. A taxpayer paying voluntarily (supposedly in accord with his marginal valuation of the commonly consumed good) has an incentive like anyone in a

bargaining situation to lie about that marginal valuation, putting it too high when the pollster comes to the door and too low when the tax collector comes to the door. The bargainer in the market stall does not tell you that he would actually be willing to sell the antique picture for \$10; he tells you instead that if he sells it for anything less than \$50 his poor old mother will go without bread.

Competition between sellers to gain customers—that is, the prospect of losing your dollars if he does not sell for the low price—is what in the end forces him to admit his true willingness to accept the low payment and brings the society to the correct amount of paintings sold. Likewise, the householder overstates how much she would be willing to accept to allow her house to be demolished for a road and the taxpayer understates how much he would be willing to pay for national defense. In these cases, however, there is no competition driving the householder or taxpayer to tell the truth. On the contrary, if the householder overstates the value of her house, she gets the large amount; if the taxpayer understates the value of national defense, other taxpayers nonetheless pay and he gets a free ride on the backs of his fellows. There ought to be some system to induce people to reveal truthfully their valuation of commonly consumed goods in the way that ordinary markets induce them to reveal truthfully their valuation of private goods. There is, but it is complex. The existence of commonly consumed goods spoils the elegant theory of dollar votes as a prescription for social happiness.

The Fundamental Flaw: The Smell of the Income Distribution

The most fundamental objection to using willingness to pay as a criterion of social choice, however, goes well beyond these practical difficulties. The objection is that your willingness to pay or to accept payment depends on how rich you are. George Grantham's willingness to accept a bribe in exchange for allowing the city to build a road through his front yard depends on his original income. If he is poor he will presumably accept little, on the grounds that his marginal valuation of the amenity of a nice front lawn is small considering that he would rather spend most of his pitiful income on food. If he is rich he will only accept a large amount. The marginal valuations are dependent on the initial distribution of income.

One is driven back to judgments that unanimity was supposed to avoid, judgments about the ethical value to be put on George's income versus Scott's or Elyce's. A dollar is no longer a dollar, and total national income is no longer an uncontroversial way of testing society's happiness. Who gets what slice of the pie now matters, even if the whole pie becomes bigger. The introduction of the power loom into textiles in Britain in the early nineteenth century increased national income a great deal, but it ruined the livelihoods of the hand-loom weavers. True though it is that cheaper cloth made many people happier, there is no morally uncontroversial way of asserting that their happiness outweighed the misery of the weavers.

Economists sometimes try to skirt the issue by observing that the gainers from cheaper cloth could fully compensate the losers from less demand for hand weaving. Out of the proceeds of efficiency could come a fund for fairness, following the logic of dollar votes under a rule of unanimity. The important difference, however, is that it is hypothetical. *If* winners could bribe losers, then winning should go forward without restraint. In such a scheme the winners do not actually have to offer the compensation; they merely have it to offer

should they wish. But the world is not obviously a better place if the rich *could* out of their new riches keep the poor from getting poorer *but do not*. Without some explicit judgment on the morality of the distribution of income, in short, measures of its size are only partial guides to policy.

The counterargument is that higher average national income is after all good on average. Taking the good with the bad, a person contemplating entering a high-income society or a low-income society would prefer the high-income society. If Betsy Hoffman has no knowledge of her ultimate position—if she makes her choice behind “a veil of ignorance,” as this method of constitutional decision making is called—she would expect to do better on average as an American than as an Ethiopian.

The argument helps in framing rules of the game, not in determining the actual distribution of income between George and Elyce or weavers and clothiers. It asserts that in the long run a society with the rule “maximize income per head” will do better for the average person than will one with the rule, say, “keep incomes equal.” In a sense, then, there is a democratic justification for the economist’s fascination with income per head.

Only in a sense, however. The present distribution of income continues to haunt the argument. The present distribution of income in society determines the demand curves, which determine the prices with which one evaluates the changes in national income. The very prices one uses to weigh together goods in national income, in other words, smell of the distribution of the society’s income. If the rich love large automobiles and distant beaches while the poor love beans and corrugated iron, and if the rich are very rich, then large automobiles and distant beaches will have high prices. The “society” will value highly things that a more equal society would value less. The problem is the index number problem in another guise. To evaluate a rise in national income using today’s prices is to build in an opinion that today’s distribution of income is an appropriate point of view.

The Moral

The hard truth about counting society’s blessings, then, is that there is no perfectly satisfactory way to do so. It is nevertheless worth doing, if only to provide magnitudes against which to balance the uncountable. The staff of the Senate committee counts the cost of the proposed dam. That the senator decides that its high cost to the nation is small when compared with the uncountable enrichment of the fine construction workers and ranchers of his state does not make the calculation of the cost irrelevant, merely less than decisive. The distant nation of Santa Americana experiences a doubling of per capita income, as counted. That the doubling leaves millions living in squalid huts in the chief city’s garbage dump countably worse off than before and uncountably powerless besides does not make the calculation of per capita income irrelevant. It merely makes it less than complete.

Summary

Despite their modesty on these points, economists have a great deal to say about social morality. Most of what they have to say is that there are no infallible guides to the social good. Voting is certainly no such guide. And unless one accepts a utilitarian ethic that permits one person to be used as a means to another’s end, the only sure guide can be unanimity, this being the only assurance that no one is hurt. At first the theory of dollar votes seems a promising extension of the rule of unanimity. It is in fact the moral premise that underlies much

practical economic thinking, such as that exhibited in the next chapter. But it, too, is flawed. It depends on a premise that the original distribution of dollar votes is morally acceptable.

For this and other reasons you should be skeptical of national income as an all-purpose measure of national happiness. It ignores the distribution of income and it treats the status quo (with its distribution) as special. Your skepticism should not, however, paralyze all thought. The national income measure is not worthless merely because it is partial. The weather is not the only cause of a good or bad day, yet we all want to know the weather report. Average income is not the only cause of a good or bad society, yet we all want to know the nation's income.

EXERCISES FOR SECTION 9.3

1. Make a brief case in favor of the Chicago system of illegal bribes to building inspectors.
2. *True or false:* If the mule Skinner didn't hate roads, the social marginal valuation in Figure 9.11 would be entirely above the curve marked Marginal Valuation of a Road Lover.
3. *True or false:* The larger the number of road lovers or likers, the higher the social marginal valuation of the road.
4. Why would it not be a good idea, socially speaking, to have more than the Efficient quantity of a road for Durnford?

PROBLEMS FOR SECTION 9.3

- 1. Three furniture moving companies, Ripoff by Regulation, ICC Incognito, and Highwayman, have pooled their business to make profits as a monopoly. They now sit down to divide up the spoils. A majority of the three can determine how the spoils are divided. *True or false:* If there is no limit (set by an agenda, say) on the bargaining permitted, the majorities for one or another division will be cyclic; that is, the companies will not be able to agree on how to divide up the spoils.
- 2. In the United States, education through high school is supplied by local governments, with majority voting on the amount to be spent per pupil, collected through taxes.
 - a. Suppose that a certain community has three social classes, U (upper), M (middle), and L (lower), each of which acts as a single person having different tastes for expenditure per pupil (depending on the average number of children in school for each social class, the importance of education to the social class, and so on). Suppose that U values educational expenditures most, M next, and L least. Show the three demand curves for expenditure per pupil. How do you construct the marginal *social* willingness-to-pay (demand) curve for expenditure per pupil from these three separate demand curves, recognizing that all members of the community must consume the same amount of expenditure per pupil (all the children go to the same school)? Draw this social curve, along with the three separate class curves, all on the same graph.
 - b. If the marginal cost of additional expenditure per pupil is constant, what is the socially optimal output of expenditure per pupil? Explain why. Draw this point on the graph of (a).
 - c. If the marginal cost is shared *equally* in taxes among the three classes, at what points will each class begin to vote against more expenditure per pupil? In general will majority rule (2 out of 3 in this case, generally, 51%) produce the socially optimal expenditure? Explain.

- d. Show how manipulating the burden of the school tax among the three classes could, given majority voting, yield the socially optimal output.
3. Suppose that two people, Harold Hawk and Donald Dove, are willing to announce truthfully the money value each puts on number of bombs used for the defense of their nation (of which they are the only citizens).
- With 100 bombs in the arsenal, another bomb has a marginal valuation of \$10 to Hawk and \$4 to Dove. What is the whole nation's marginal valuation of a bomb?
 - With 200 bombs another bomb has a marginal valuation of \$5 to Hawk and zero to Dove. What now is the whole nation's marginal valuation?
 - If bombs for the nation's arsenal can be produced by sacrificing other commodities valued at a constant marginal cost of \$10 per bomb, will 100 be too few and 200 too many bombs? Why or why not? Draw a diagram with marginal cost and marginal valuation that illustrates the socially correct number of bombs.
 - Show that if bombs cost \$10 each at the margin, the only division of the marginal tax burden to pay for the last bomb that both Hawk and Dove will accept is a division according to their marginal valuations of the bomb.

True or False

- Giving bribes to traffic police is socially speaking more efficient than is taking the consequences of a traffic ticket (that is, having to go to court, paying a fine and court costs).
- According to Problem 4, therefore, we should encourage the police to accept and the public to offer bribes.

10

Consumers' Surplus

10.1 Consumers' Surplus: The Elements

What to Read For **What is value in exchange? Value in use? Consumers' surplus? How is consumers' surplus related to the area under a demand curve? How is the area under a demand curve related to what society should pay for a bridge? What price should be charged to cross a bridge? What is producers' surplus? What are some alternative names for it? Why does competitive equilibrium make the sum of producers' and consumers' surplus as large as possible? Does the equilibrium leave any mutually advantageous deals unexploited? Why does rent control leave advantageous deals unexploited?**

Value-in-Use Is Higher Than Value-in-Exchange

National income is a way, flawed though it may be, of measuring happiness in money. So is *consumers' surplus*. In fact, as will be shown in the next section, a correctly measured change in consumers' surplus is the *same thing* as a correctly measured change in national income.

The definition of "consumers' surplus," however, sounds very far from national income. Consumers' surplus is the money value of the willingness of consumers to pay in excess of what the market price requires them to pay. If you had to, you would pay more to get a pizza or an education than you do. As George Stigler put it, "When a reflective man buys a crowbar to open a treasure chest, he may well remark to himself that if necessary he would have been willing to pay tenfold the price. . . . Marshall gave the odd name of 'consumer's surplus' to these fugitive sentiments."¹ The crowbar costs \$5 but the man would have been willing to pay up to the entire value of the opened chest—\$50 or \$500 or \$50,000. The excess—\$45 or \$495 or \$49,995—is a measure of the man's *gain from trade*.

The difference is between the money amounts of value in use and value in exchange. In this terminology it has long been understood that *value in use*

¹ George Stigler, *The Theory of Price*, 3rd ed. (New York: Macmillan, 1966), p. 78.

minus value in exchange equals consumers' surplus. For voluntary exchange it is always positive. The value in use always exceeds the value in exchange, simply because the most you would pay must always be more than what you actually pay—or else you wouldn't pay. You buy a house for \$40,000. Unless you are just indifferent between buying and not buying, the value in use of the house is larger than \$40,000; if it were only, say \$10,000, you would not pay the \$40,000 value in exchange to get it.

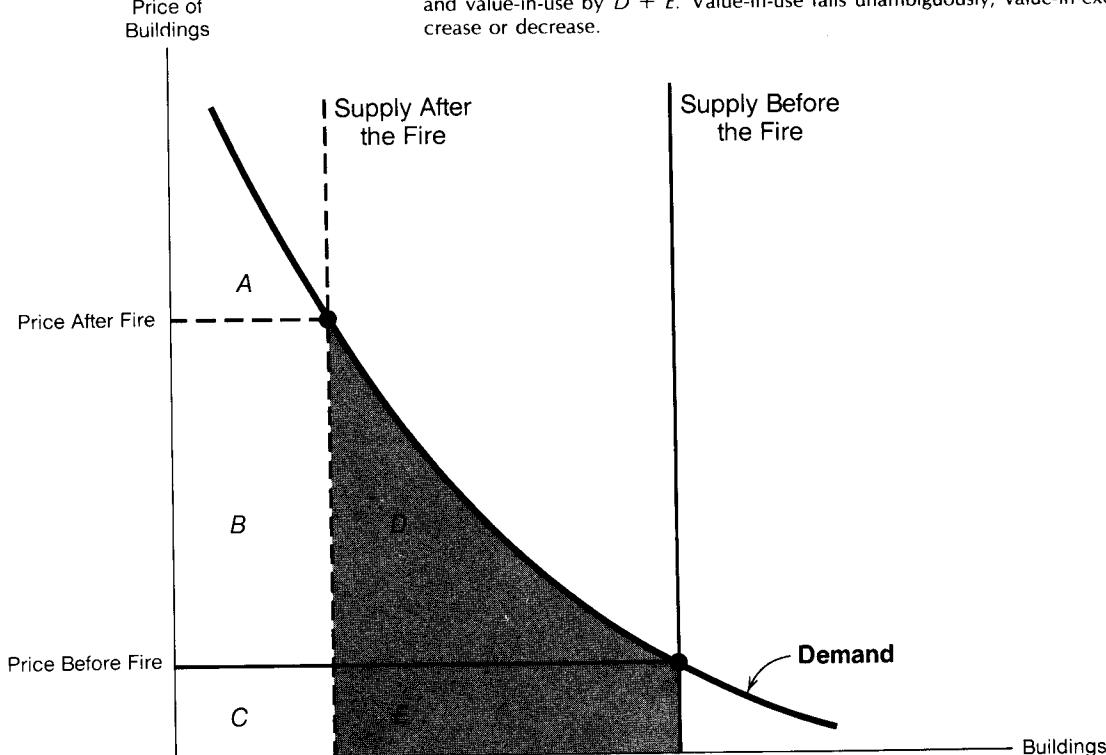
Q: The Chicago Fire of 1871 destroyed half the cubic contents of the city's buildings. The value of the 50 surviving buildings out of 100 *after* the fire probably exceeded the value of all 100 original buildings *before*. *True or false:* We see here the madness of economics and of economies, which witness a rise in value despite a great calamity.

A: What we see is the ambiguity of the word "value." The value in exchange, which is what is meant in the

second sentence, may go up, to be sure. But because the fire leaves fewer usable buildings, the value in use, which is what is meant in the third sentence, always goes down, in accord with good sense. The maximum amount that someone would pay for a house like the one at 1205 S. DeKoven Street—say, \$500—is unchanged by the fire. The market price might rise from \$100 to \$250, closer to the maximum amount. But the maximum itself is still \$500. Burning down \$500 worth of value in use cannot be a good thing. So, false.

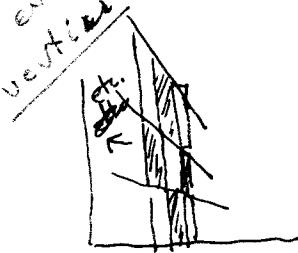
Figure 10.1
A Fall in Supply Reduces the Value-in-Use

The decrease in the supply of building services increases value-in-exchange by $B - E$ and value-in-use by $D + E$. Value-in-use falls unambiguously; value-in-exchange may increase or decrease.



→ Here's how think of projects: 198 → Doesn't matter who actually pays it: that leaves on road pay is perhaps just, but not efficient necessarily.

② Run program (1) private
③ Set laws (2) public
④ Correct way to evaluate public projects =
⑤ Evaluate 2 of d's.



This is how
to think

(B) What people actually pay is always less than what it's worth

The distinction between value in use and value in exchange explodes in usefulness when it is attached to a demand curve.² The problem on the Chicago Fire can illustrate the point. Interpret the demand curve for buildings as a curve of marginal valuation, that is, *marginal* willingness to pay. That is, its height is how much consumers will pay for each successive increment in buildings. The area under the demand curve, therefore, is the *total* willingness to pay, because adding up all the marginal willingness out to the total number of buildings will exhaust all the willingness. But the total willingness to pay is the same as the value in use. The total amount consumers will pay if all their willingness is extracted building by building is evidently the maximum they would be willing to pay in a lump sum to get the buildings. (Or so at least we can assume until the argument is made more rigorous later).

Now look at Figure 10.1. The value-in-exchange before the fire is the area $C + E$, the value in exchange after is $C + B$. Because the demand curve is inelastic it has risen. But the total value in use before is the whole area $A + B + C + D + E$; after, it is only $A + B + C$, a fall by the shaded area $D + E$. The diagram captures the verbal argument that the maximum people would pay does not rise and that the burning of half the buildings does reduce (by the shaded area) the "value" enjoyed in this sense. The diagram makes the argument much clearer. The triangular area $A + B + D$ is the consumers' surplus from the buildings before the fire; the area $C + E$ is the income of building owners. The fire causes consumers' surplus to fall to merely the triangle A , with some of the loss of consumers' surplus, the area B , reappearing as income to building owners and some, D , disappearing entirely.

Consumers' Surplus Is Useful for Reckoning the Benefit of Public Projects

Q: Suppose that the city fathers of Boston and Cambridge contemplate building a bridge, the Lars Anderson Bridge, to link their fair cities across the Charles River. Suppose, too, that the clientele for the bridge for its life divides into three parts: 500,000 Harvard students, who each would be willing to pay \$10 for a career of crossings (in view of the alternative routes); 1,000,000 townies, who would be willing to pay \$5 each; and all the rest of the people, who would be willing to pay nothing, and would require in fact some slight payment to go out of their way to cross it.

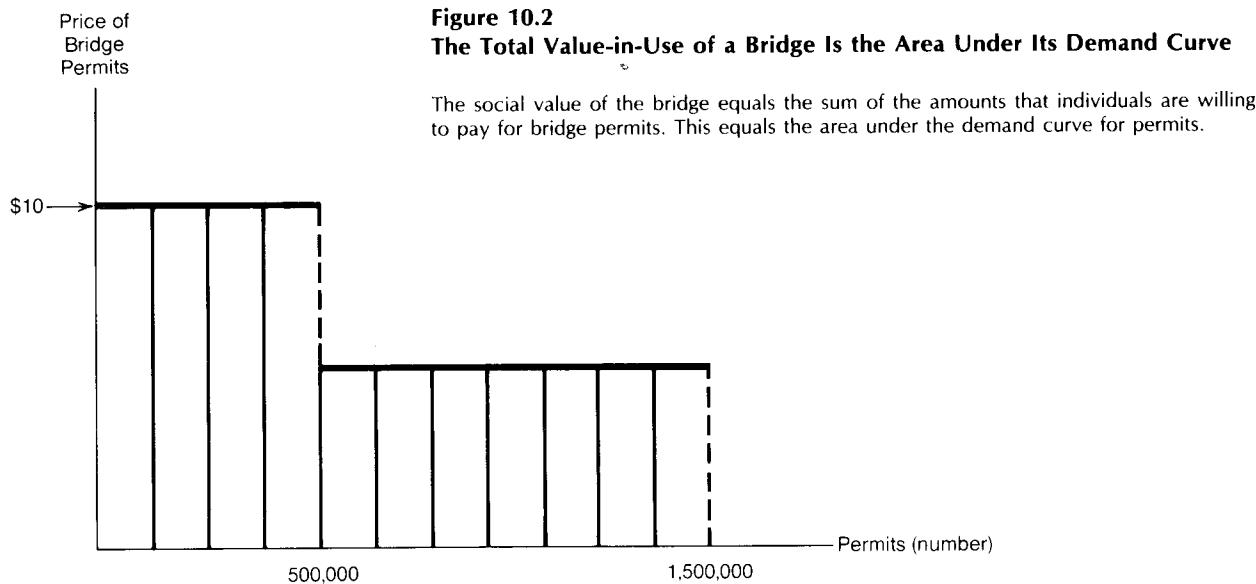
² The idea of doing so is commonly attributed to Alfred Marshall, but it is more accurately attributed to a French engineer, Jules Dupuit, "On the Measurement of the Utility of Public Works," first published in 1844, reprinted in K. J. Arrow and T. Scitovsky, eds., *Readings in Welfare Economics* (Homewood, Ill.: Irwin, 1969), reinvented for the English-speaking world by another engineer, Fleeming Jenkin, "On the Principles Which Regulate the Incidence of Taxes," first published in 1871–1872, reprinted in R. A. Musgrave and C. S. Shoup, eds., *Readings in the Economics of Taxation* (Homewood, Ill.: Irwin, 1959). That engineers invented it testifies to its use in evaluating engineering projects such as roads and dams.

The idea that an area under a demand curve is a money measure of total happiness is astonishingly fruitful. The first fruit is a rule for constructing toll-free bridges, dams, post offices, railway lines, highways, and other public edifices for which there is no charge. Since there is no charge, there is no revenue by which to judge the usefulness of the edifice. Unlike a flower stand or an auto factory, the market does not toss up a measure of usefulness.

1. What is the demand curve for lifetime permits to cross? (Use the assertions in the question.) If no price is charged for the permits, what is the number of people who will cross?

2. How much would they be willing to pay, if necessary, to do so? What, then, is the total value of the bridge? (*Hint:* Add up the areas under the demand curve.)

3. If the bridge costs in sacrificed opportunities elsewhere \$9,999,999 to build, should it be built? What if it costs \$10,000,001?



A: 1. The demand curve is a stairway. At 500,000 the Harvard stair is used up and the marginal valuation falls down to \$5 (see Figure 10.2). If the price is zero (no toll), exactly 1,500,000 Harvard students and townies cross.

2. Each of the 500,000 \$10-people would pay \$10, which yields \$5,000,000 from all of them; each of the 1,000,000 \$5-people would pay \$5, which yields another \$5,000,000, for a total of \$10,000,000 in willing-

ness to pay, that is, in value in use (the value in exchange is zero if the toll is zero).

3. If the bridge costs less than the money value of the satisfaction it brings (that is, willingness to pay), it should be built. Otherwise, as when it costs a dollar more than \$10,000,000, it should not. A society that indulged in such projects would fill in the Great Lakes for ski resorts or level the Rockies for farmland.

Consumers' Surplus Is Useful for Deciding How to Charge for the Project

Q: Suppose that another crosser at 10 P.M. adds nothing to the cost of the Golden Gate Bridge in either replacement costs of the bridge itself or in inconvenience to others by crowding. Suppose that the cost of the bridge divided by the number of crossings that will occur over the life of the bridge is 50 cents. *True or false:* Therefore, the toll at 10 P.M. should be 50 cents. (*Hint:* Forget about fairness; the point is to get the best use out of the existing bridge.)

A: It should be zero, which is the additional cost of the crossing. The social purpose of prices is not to punish people, or even to achieve the apparent fairness of mak-

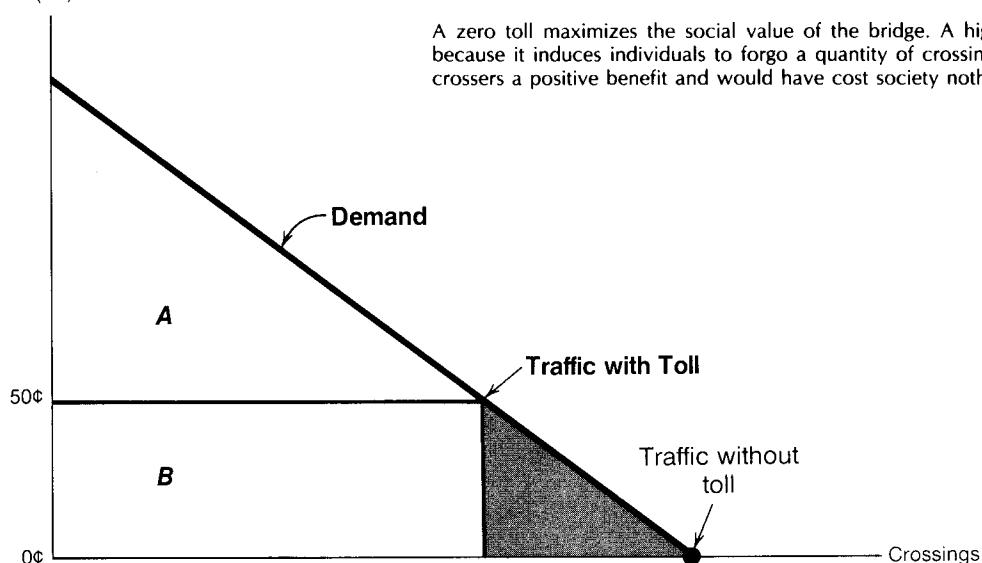
The method of value-in-use, then, is a guide to whether Mars landings, aqueducts, radar systems, and the like should occur at all. Of course, in many cases it would require the aid of a wizard to know the demand curve to evaluate the projects. But the method gives general guidance to how they should be paid for.

ing those who use the bridge pay for it. The purpose is efficiency. We want to induce people to use the bridge in a way that will maximize the value in use gotten from it and from the alternative uses of resources. For this purpose the relevant fact is that Molly McClelland as the extra crosser causes no sacrificed opportunities elsewhere. So McClelland should cross free. If her crossing wore off 8 cents worth of paint and bridge, it would be desirable to make her pay this cost in an 8-cent toll, to induce her to forgo the crossing unless it were worth in use at least the 8 cents of sacrificed opportunities it cost. But if not, then the past cost of the Golden Gate Bridge (incurred, say, to build it) is irrelevant to

Crossings
Price
(toll)

Figure 10.3
The Best Toll on a Bridge Is Zero

A zero toll maximizes the social value of the bridge. A higher toll causes a social loss, because it induces individuals to forgo a quantity of crossings that would have given the crossers a positive benefit and would have cost society nothing.



marginal cost—that is, it is irrelevant to the additional cost of one more crossing. By the assumption of the question the marginal cost is zero, and so should be the toll. The toll of 50 cents reduces society's happiness

by the shaded triangle of willingness to pay (see Figure 10.3). So, false; the average cost of constructing the bridge is beside the point.

With a 50-cent toll the consumers get area *A* in surplus (the excess of value in use, *A* + *B*, over value in exchange, *B*). The *B* area is the part of willingness to pay extracted from crossers and given to whoever owns the bridge. From the social point of view the crossings that still take place with a 50-cent toll cause happiness in the value of *A* + *B*. The distribution of income is changed by the toll, shifting *B* dollars of purchasing power from crossers to owners. The critical point is that *the shaded triangle is sheer waste*. Crossings not worth at least 50 cents do not take place, even though these crossings have some value and no opportunity cost. "The distribution of wealth among members of the community is affected by the mode of payment adopted for the bridge, but not the total wealth, except that it is diminished by tolls. . . . This is such plain common sense that toll bridges have now largely disappeared from civilized communities."³

³ Harold Hotelling, "The General Welfare in Relation to Problems of Taxation and of Railway and Utility Rates," first published in 1938, reprinted in Musgrave and Shoup, eds., p. 158. He added: "But New York City's bridge [sic] and tunnels across the Hudson are still operated on a toll basis, because of the pressure of real estate interests anxious to shift the tax burden to wayfarers, and the possibility of collecting considerable sums from persons who do not vote in the city." Today 6 out of 19 roads off the island of Manhattan charge tolls, and all the roads to New Jersey do. Yet New York does not fall below San Francisco in civilization; all four of the bridges across San Francisco Bay charge tolls. As we shall see in the next chapter, however, assigning degrees of civilization is not quite so easy when it is realized that the taxation to pay for the bridges (if tolls are to be zero) itself creates triangles of net social loss under the demand curves for the taxed goods. It is a choice between two evils.

**Producers' Surplus
Is Analogous, the
Producers' Gain
from Trade**

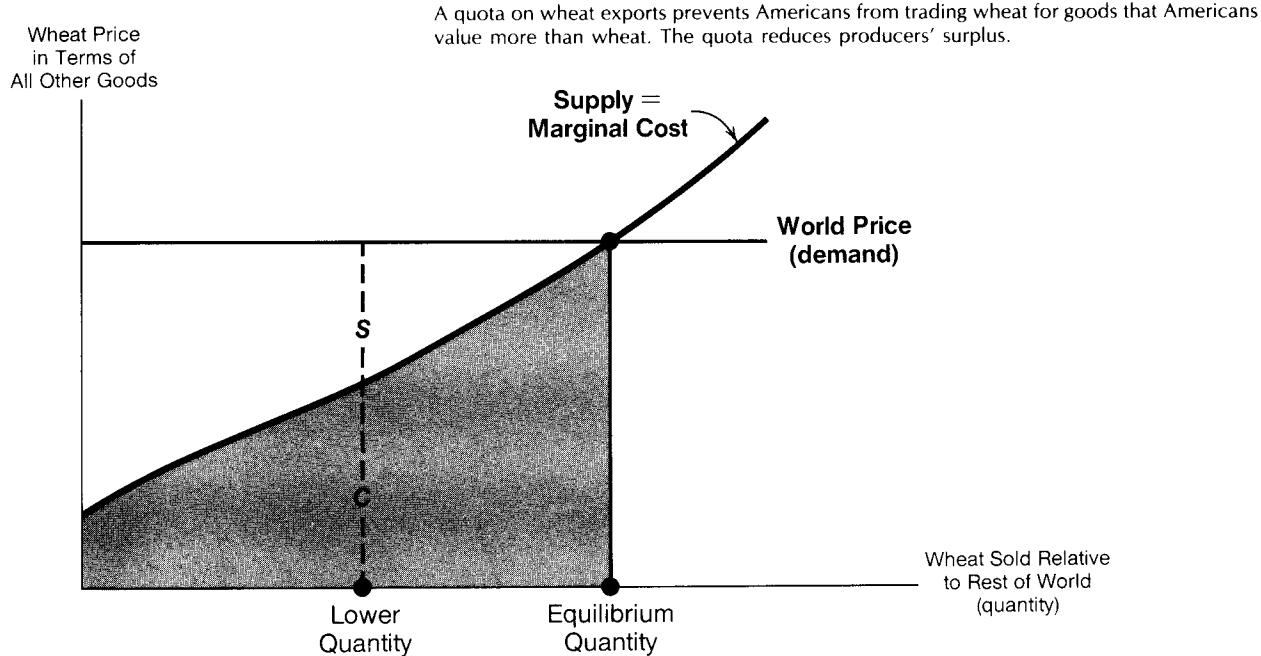
Consumers' surplus, then, is the consumers' net gain from trade, namely, a triangle equal to their gross gain (the trapezoid of value in use) minus their cost (the rectangle of value in exchange). Consumers imply producers. A *producers' surplus*, therefore, ought to exist, similar to consumers' surplus. In fact it does, and like consumers' surplus is the producers' net gain from trade. Consider, for example, American exports of wheat.

Q: When the crop is poor and wheat is expensive in America, it seems reasonable to restrict American exports. After all, our own citizens should be fed first, shouldn't they?

A: Our own citizens will be best off when the trade in grain is free. The supply curve of wheat to the rest of the world slopes upward in Figure 10.4 because more wheat is producible only at the expense of ever more costly resources in shipping, growing, and the like, and because more wheat from a given production is suppleable to the rest of the world only at the expense of ever more valuable uses in consumption at home. That is, the supply curve is the marginal (opportunity) cost.

The shaded area, C , under the marginal cost is therefore the whole cost of supplying the amount marked as the Equilibrium Quantity, for it is the summation of all the successive increments to cost. The rectangle $C + S$ is the revenue from selling the Equilibrium Quantity to the rest of the world at the World Price. The unshaded triangular area, S , then, is the *surplus of producers' revenues over costs*. In a word, it is America's *profit*. In another, it is America's *economic rent*. In still another, it is America's *producers' surplus*. It reaches a maximum when trade is free and the Equilibrium Quantity is sold. At a Lower Quantity the value of wheat is lower in America than in the rest of the world. Americans could do better in terms of their own valuation

Figure 10.4
Free Trade Maximizes Producers' Surplus



of wheat and all other goods by selling more wheat at the high World Price. The amount of all other goods

they get more than outweigh in value to them the wheat they give to the rest of the world. That is, false.

The Two Surpluses Together Show That Free Trade Is Good

A generalization of the argument makes the familiar point that free trade is best, and even shows how it might not be. Should there be, for instance, a free market for motorcycles in Columbus, Ohio? Well, suppose that competition would drive the market to Equilibrium in Figure 10.5. The sum of consumers' and producers' surplus is a measure of the entire gain from trade. The entire gain is the gain of student motorcyclists at Ohio State University plus the gain of Columbus motorcycle sellers. The position marked Equilibrium maximizes it. This fact is obvious for quantities Less Than Equilibrium, because the big, shaded triangular area (bounded by the three points, m , a , x) is chopped down to a smaller trapezoidal area (m , i , n , x).

It requires a little proof for quantities larger than Equilibrium. Consider Figure 10.6. Suppose, just to be definite, that the market deal takes place at a point on the demand curve (not along the supply curve or along neither). What needs to be proven is that, at More Than Equilibrium compared with Equilibrium, the gain by student motorcyclists of a large quantity and low price is more than offset by the loss by motorcycle sellers. The rise in Consumers' surplus in the move from Equilibrium to More Than Equilibrium is the horizontal trapezoid $A + B + C$. Is the fall in producers' surplus larger, as it should be if More Than Equilibrium is on balance bad for society? Yes. The area $A + B$ is the fall in revenue to producers on the old quantity of motorcycles when they are sold at the new, low price. The Equilibrium quantity still costs what it did before. But its revenue is less by $A + B$ since the market price is lower. The More Quantity out to More Than Equilibrium costs producers $C + L + R$, but gets them only R in revenue. So on this account, they lose $C + L$, to add to the loss of $A + B$. The whole fall in producers' surplus is $A + B + C + L$, which is L more than the rise in consumers' surplus. This would be true of any point

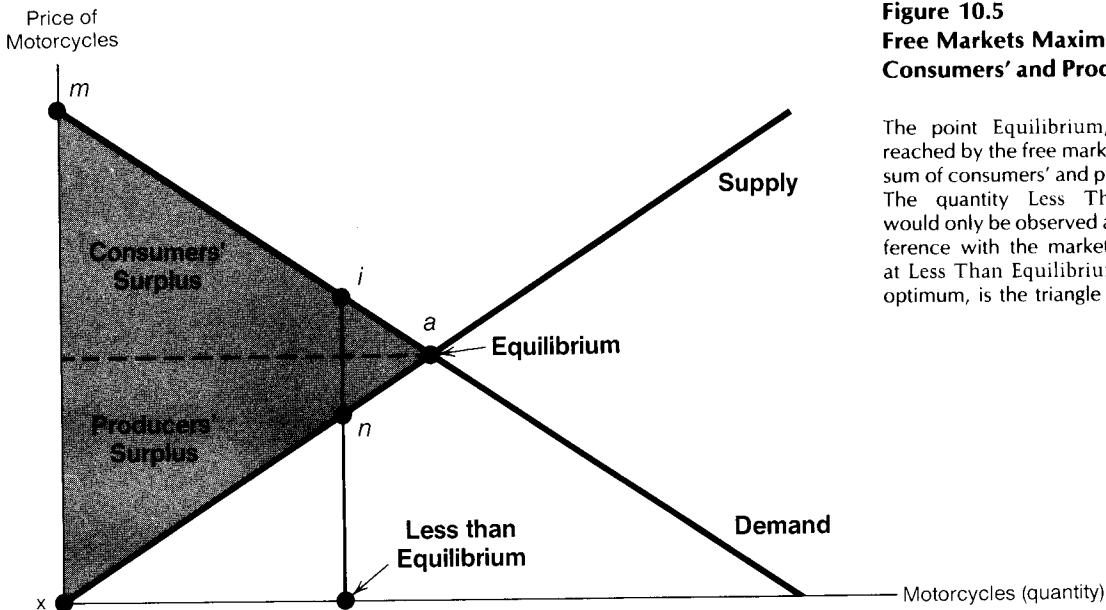


Figure 10.5
Free Markets Maximize the Sum of Consumers' and Producers' Surplus

The point Equilibrium, which will be reached by the free market, maximizes the sum of consumers' and producers' surplus. The quantity Less Than Equilibrium would only be observed as a result of interference with the market. The social loss at Less Than Equilibrium, relative to the optimum, is the triangle ina .

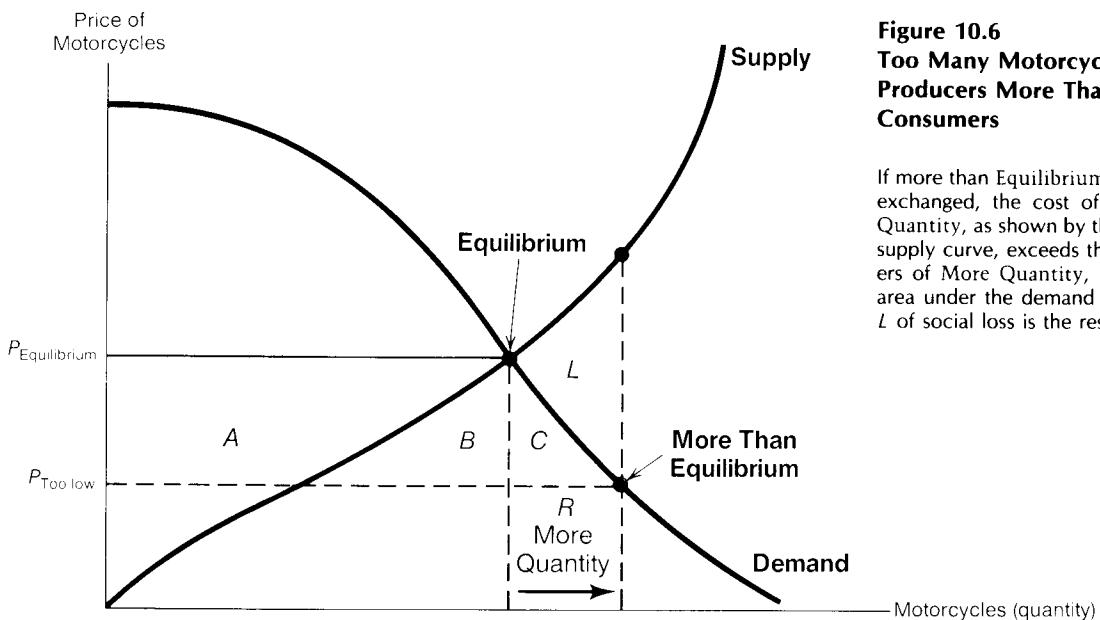


Figure 10.6
Too Many Motorcycles Hurts
Producers More Than It Helps
Consumers

If more than Equilibrium motorcycles are exchanged, the cost of supplying More Quantity, as shown by the area under the supply curve, exceeds the benefit to buyers of More Quantity, as shown by the area under the demand curve. A triangle *L* of social loss is the result.

out of Equilibrium. So only at Equilibrium is the sum of consumers' and producers' surplus maximized, which was to be demonstrated. Equilibrium, in other words, is efficient.

Notice that, in the way of efficiency, Equilibrium is not best for either consumers *by themselves* or producers *by themselves*. Consumers, for example, could do better if they could compel producers to give away the motorcycles; and short of such robbery they could do better, even if they had to stay on the producers' supply curve, by restricting their own purchases a little.⁴ Equilibrium is best for "society," that is, for consumers and producers together on balance.

The Feasible Area of Exchange Is the Triangle of Surplus

There is another way of exhibiting consumers' surplus and its uses.⁵ The landlords of New Orleans, say, have a supply curve of rental housing. It is as usual upward sloping, because more houses are expensive to build and because the landlords themselves can be enticed to live elsewhere at a high enough price. Look at Figure 10.7. But the supply curve is a schedule of the *minimum acceptable* price for any quantity, such as Small. Landlords would be pleased to get more than the price Barely Acceptable to Suppliers for the quantity Small. So it is with other quantities along the supply curve. The whole vertically shaded region *above* the supply curve, in other words, is the region of deals acceptable to landlords. Tenants likewise would be pleased to get the quantity Small at any price *less* than the Price Barely Acceptable to *Demanders*, and the whole horizontally shaded region *below* the demand curve in the diagram is the region of deals acceptable to tenants. Putting the two areas together yields the cross-hatched area of possible deals compared with no deals (zero housing rented) that make both landlords and tenants better off. The crosshatched area, in other

⁴ As will be shown in Chapter 17 on monopoly and in a question at the end of this section.

⁵ The idea is Milton Friedman's in his *Price Theory* (Chicago: Aldine, 1976), Chapter 2.

words, is the area of mutually advantageous deals, and these are exhausted at the corner marked Equilibrium. At Equilibrium the mutually advantageous deals have been exhausted. To have larger or smaller quantities rented by landlords to tenants would make one or both worse off; at Small, for instance, the deals have not been exhausted, and more renting would make both better off.

Interference in Free Trade

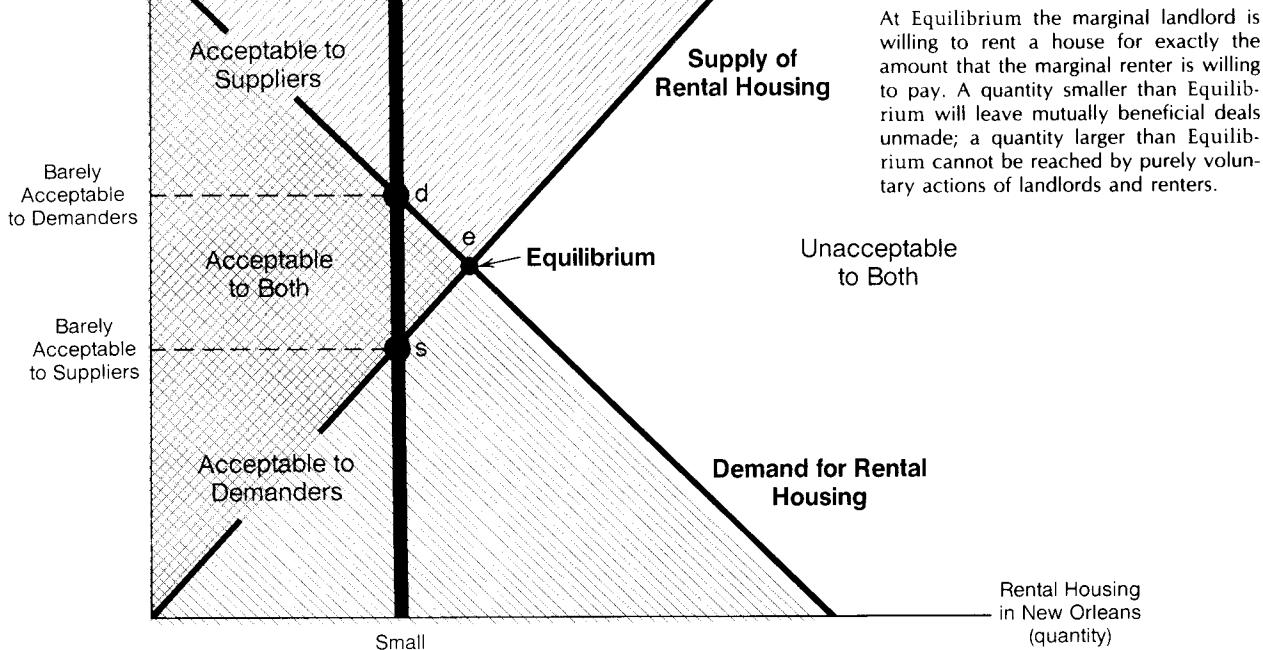
The connection between consumers' surplus and this way of looking at curves of supply and demand is that the areas of mutual benefit are the same money measures as consumers' surplus. The triangle *des* in Figure 10.7 is a money measure of the inefficiency of staying at Small.

Q: Suppose that the market for rental housing in New Orleans is in Equilibrium initially and that at the Old Price the government imposes rent control (that is, making it illegal to offer or accept a rental above the Old Price). Now suppose that the demand curve of tenants rises, perhaps because the population of New Orleans or the money income of tenants has risen. Describe the social loss from the rent control, assuming that it is enforced.

A: The old demand curve, which determined the Old Price, is the dashed line in Figure 10.8. Because the

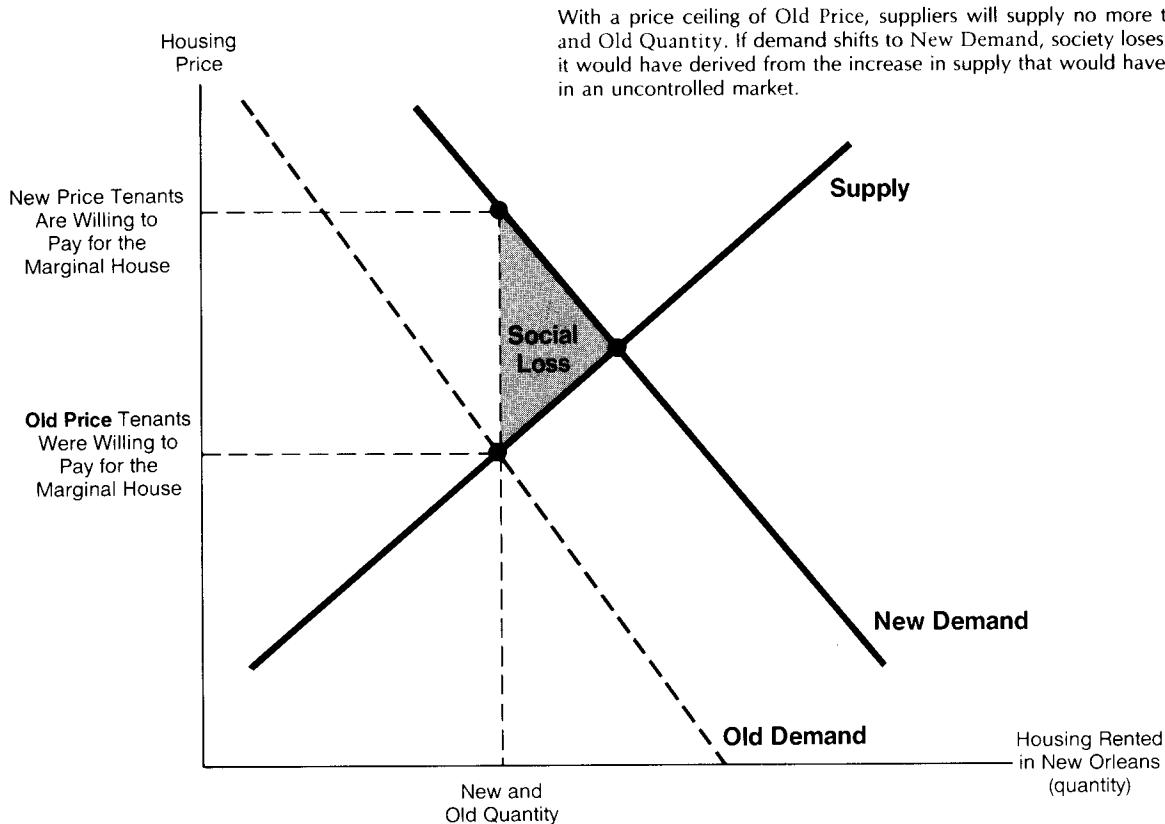
government forbids rentals above the Old Price, a landlord has no incentive to surrender more housing. Tenants are unable to express in money offers to the landlords their new and larger willingness to pay for housing. The shaded area is the money measure of inefficiency. It is the loss to everyone—tenants as well as landlords—from making new deals impossible. Or, to put it another way, the shaded area is the sum of consumers' and producers' surplus forgone.

Figure 10.7
Competitive Equilibrium Exhausts the Opportunities for Mutually Advantageous Exchange



At Equilibrium the marginal landlord is willing to rent a house for exactly the amount that the marginal renter is willing to pay. A quantity smaller than Equilibrium will leave mutually beneficial deals unmade; a quantity larger than Equilibrium cannot be reached by purely voluntary actions of landlords and renters.

Figure 10.8
Why Most Economists Disapprove of Rent Control



Summary Consumers' surplus is the excess of value in use (accruing to the consumers) over value in exchange; likewise, producers' surplus is the excess of value in exchange (accruing to producers) over value in use elsewhere (as for example in use by the suppliers themselves). The expression "consumers' surplus" is sometimes used to stand for both surpluses. The two are the gains from trade to all the parties, that is, money values of the happiness gained over not trading. The sum of the two surpluses is the total social gain from trade. It is maximized when the quantity traded is that of supply equal to demand. The idea is unremarkable. What is remarkable is its wide applicability to social questions, from free trade in housing and grain to the desirability of building bridges.

EXERCISES FOR SECTION 10.1

1. If you personally had the choice only between buying them from your usual supplier or *not being able to buy at all*, what would you pay at most for the following items? What do you actually pay? What's your consumer's surplus for each?

- a. Lunch today.
 b. Your shoes (all of them).
 c. Your running shoes.
 d. Your college education, anywhere.
 e. Your college education, at this college.
2. If you do not eat lunch today, what is the relation for the lunch between value in use (what you get) and value in exchange (what you pay)?
3. Fill in some quantities and prices in Figure 10.1 that would give a "paradox"—that is, value (in one sense) going up while value (in another sense) is going down. Then fill some in that would *not* give the paradox. What's the difference?
4. *True or false:* Users of a road should pay for what it cost originally to build the road, because that is efficient.

And this is wrong



PROBLEMS FOR SECTION 10.1

- 1. "Some very capable engineers wanted to know what was the utility of the French roads, and starting from the datum that the prices paid by society for their use amounted to 500 million (francs) per annum . . . they said that since society consents to pay 500 million for these transport facilities, their utility is 500 million."⁶ Are the engineers capable in economics? (*Hint:* Is the amount paid the value in exchange or the value in use?)
2. The cost of electricity from a dam is very largely the cost of constructing the dam. Once constructed, the costs of opening and closing the sluice gates, repairing the generators, checking the security of the dam, and so forth are essentially trivial. If the demand does not crowd the capacity of the dam, the opportunity cost of an additional kilowatt-hour of electricity produced is essentially zero. Suppose that the cost of a dam is \$30 million, that the demand curve for electricity over the life of the dam (expressed as a price dependent on the quantity sold, the quantity measured in 100,000 millions of kilowatt-hours) is dollars per kilowatt-hour = $\$0.40 - 4Q$, and the dam can supply with no crowding all the electricity demanded on it. Should the dam be built? (Measure Q in units of 100,000 millions. Price is measured in dollars per kilowatt-hour.)
- 3. Show the consumers' and producers' surplus that, for a point both above the demand curve and below the supply curve for motorcycles (that is, at a quantity other than Equilibrium), the gain for one group to equilibrium is more than offset by the loss to another. Assume that demand plus supply curves properly measure all social benefits and costs. (Notice that the version of the problem worked in the text assumed that the deals took place along the demand curve. The present version complicates the situation by considering a deal off the demand curve and the supply curve, in which *neither* consumers nor producers are pleased.)
 - 4. Is it true that consumers do better than Equilibrium if they can somewhat restrict their purchases from producers, driving down the purchase price enough to offset the loss of value in use from smaller purchases? (*Hint:* Try it out. Look carefully at the new areas of surplus, compared with the old.)
 - 5. A motorcycle costs society and its owner the sacrificed alternatives of the steel, labor, and so forth that have gone into making it. But it also costs the sacrificed alternative of peace and quiet in the neighborhood. Unless Nick Lash is exceptionally public spirited, he will not, when he buys a motorcycle, include the cost of noise in the cost he reckons when deciding whether to buy. Show with areas of consumers' and producers' surplus how many motorcycles people will buy, how this number differs from the social optimum,

⁶ From Dupuit (1844) in Arrow and Scitovsky, eds., *Readings in Welfare Economics*, p. 256.

and how much society as a whole (the aggregate of students, sellers, and the new, third party of people who sleep) is hurt by the difference. Think in terms of two marginal costs of various amounts of motorcycles, a *marginal private cost* (which the buyer faces) and a *marginal social cost* (which is higher at each amount by the value of sacrificed peace and quiet the neighborhood faces).

6. Show that if motorcyclists foolishly underestimate the probability of killing themselves on motorcycles, and would value motorcycles less from an informed perspective, then too many motorcycles will be produced. Identify the social loss.

True or False

7. It does not matter for the portrayal of the market equilibrium or of the loss relative to the optimum from receiving it whether the cost of noise pollution from motorcycles is viewed as an addition to society's supply curve or as a subtraction from society's demand curve.
- o 8. In the matter of loud motorcycles in Columbus, the hurt to sleepers exceeds the gains to producers and consumers from moving from the optimum point (viewed from supply, say) to the market point.
 - 9. As long as the socially correct quantity of housing is somehow demanded and supplied (perhaps by compulsion), it does not matter at all for the maximization of the sum of consumers' and producers' surplus how the supply is priced.

10.2 Further Uses of Consumers' Surplus: Middlemen and Other Exchanges

What to Read For

What is arbitrage? Why is storage and resale of coffee like transporting coffee from a low-price town to a high-price town? Is the transporting a good thing? When does the transporting stop? Are middlemen, speculators, and so forth good things? Are they good if they are mistaken about the future price? Do consumers like middlemen? Are they justified in their dislike? Should society as a whole like middlemen? What is the use of horizontal as distinct from vertical consumers' surplus? Is it a good idea to restrict exports of corn when the American crop is bad?

Middlemen Are Merely Transporters of Goods Across Time

Consumers' surplus is a fine tool for examining the goodness or badness of markets. Middlemen in commodity markets, for example, have always attracted suspicion. It is said that someone who makes a living buying wheat or copper or hog bellies at low prices in one month and selling them at high prices in the next month is not making anything or doing anything. In the language of medieval regulation the person is said to be a "regrater"; in the language of ersatz economics the person is said to be a "profiteer." At one time or another, tulip bulbs, British government bonds, canals, sugar, coffee, cotton, export goods, building sites, land confiscated from nobles during the French Revolution, public lands, copper, foreign money, gold, buildings, commodity futures, stock options, and ownership certificates in English country banks, the British East India Company, foreign mines, railroads, joint-stock banks, discount houses, new corpora-

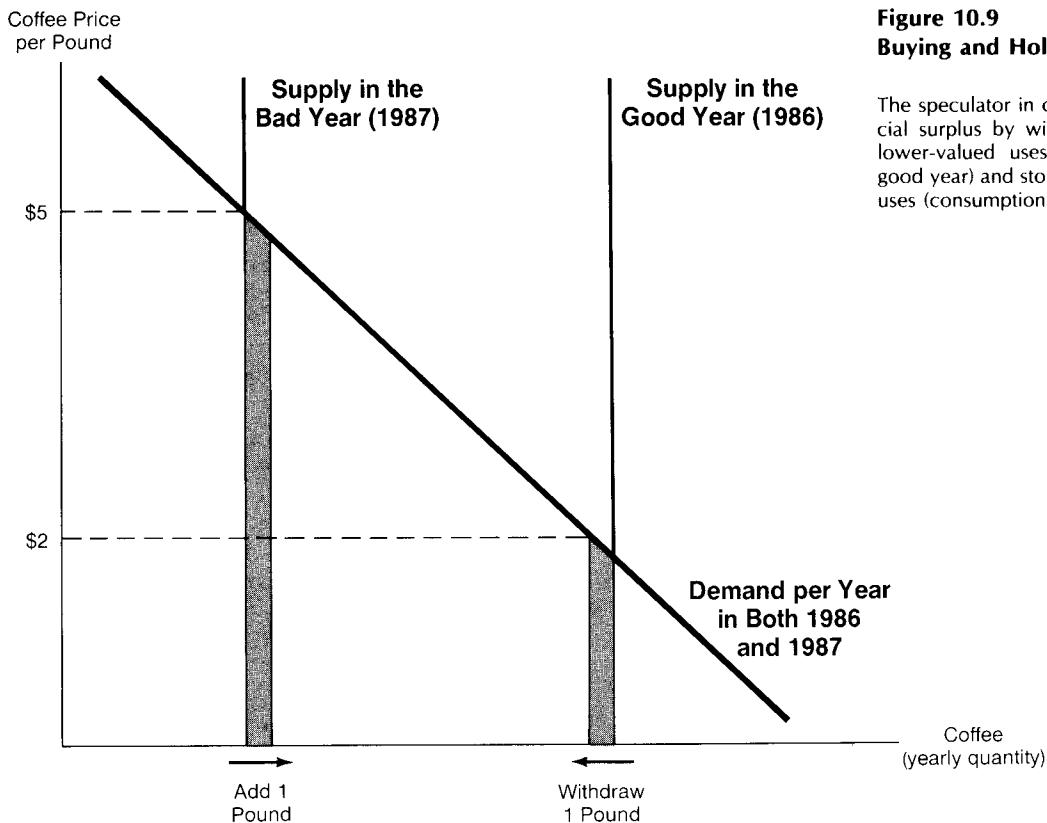


Figure 10.9
Buying and Holding for Profit

The speculator in coffee increases net social surplus by withdrawing coffee from lower-valued uses (consumption in the good year) and storing it for higher-valued uses (consumption in the bad year).

tions, and old corporations have been objects of buying low and selling high.⁷ An air of evil hangs over these markets. But it is not justified.

T or F: If the crop of coffee will certainly fail in 1987, the middleperson's purchase of 1986 coffee to sell at an obscene profit in 1987 is socially desirable.

A: The middleperson is a *mere transporter of coffee*. Coffee in 1987 will be more expensive than in 1986 if the middleperson, Teresa Baker, does not transport coffee. She can therefore make money doing so. In other words, true. In trying to make money out of buying low and selling high, to put it another way, the speculator plays the same role as someone trying to make money out of a difference in automobile prices between

California and Illinois or a difference in wheat prices between Canada and China. For the same reason that ordinary trade across space is desirable, so too is trade across time. The pound of coffee withdrawn from consumption to be stored in 1986 is worth only, say, \$2 a pound, emerging from the storehouse in the year of crop failure (1987) worth, say, \$5 a pound. Since the demand curve in both years is the same (the consumers are roughly the same people), values such as these can be read off one curve for both years (see Figure 10.9). Baker induces coffee drinkers in the good year to give up a little coffee that they themselves will value higher

⁷ The list is from Charles P. Kindleberger's *Manias, Panics, and Crashes: A History of Financial Crises* (New York: Basic Books, 1978), pp. 39–40. The book is a good one and a good antidote to the views on arbitrage, speculation, and the like expressed in the pages that follow.

in the bad year. She induces them to do *what they would want to do for themselves* if they had the good information and good storage facilities she has. Value

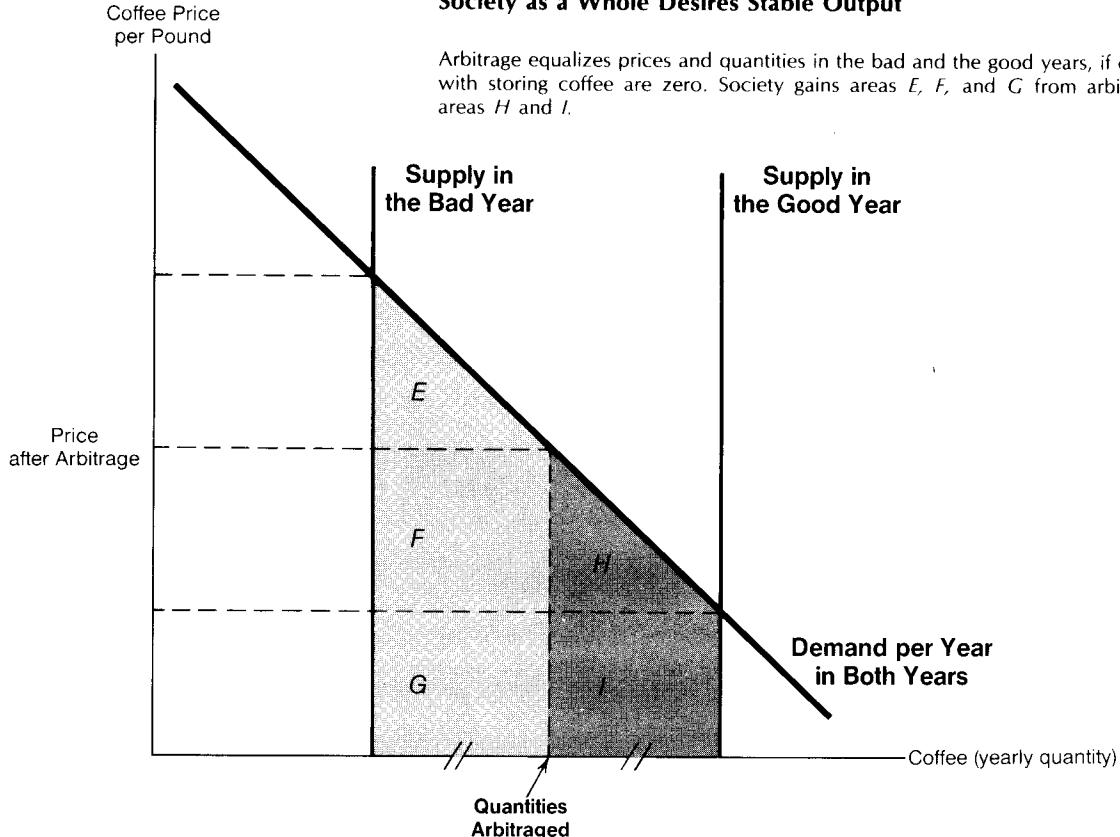
has been created by mere waiting, the difference between the tall and short columns of willingness to pay being the net gain in value.

As Baker and her competitors continue to *arbitrage* the present and future price (that is, to buy in the low-price market in order to sell in the high-price market), the prices converge. If arbitrage and storage have no costs, and if the future is certain, the process stops only when prices in the two years are equal, that is, only when quantities are equal, as at the point Quantities Arbitraged in Figure 10.10. The larger shaded area ($E + F + G$) is the gain to society from having more coffee in the bad year, and the smaller area ($H + I$) is the loss to society from having less coffee in the good year. The larger area is indeed larger. So if the demand curve slopes downward there is always a net gain from the speculators. For the straight-line demand curve the gain is area F . In general it is $E + F - H$ (since G and I are equal and cancel).

Another way of putting the point is to say that society as a whole gains from the *stabilization* of output. Who exactly gets the gain is irrelevant to the question of whether society as a whole gains. A monopoly of middlemen,

Figure 10.10
Society as a Whole Desires Stable Output

Arbitrage equalizes prices and quantities in the bad and the good years, if costs associated with storing coffee are zero. Society gains areas E , F , and G from arbitrage and loses areas H and I .



for example, might be able to extract for itself all the net gain, by taking every dime of excess willingness to pay. *But middlemen are people, too.* They are part of the "society" that gains: Who gains is a matter of distribution, not efficiency; of the ownership of national income, not its size. Usually the middlemen do not get all the gain, for they compete with each other in offering cheap good-year coffee to consumers in the bad year. In the extreme they get none of the gain, being paid only enough to bring them out of other occupations and into being middlemen. But whoever gets the gain, it exists. National income rises. Trade is productive, whether over space or over time, and the middleman who carries out the trade deserves two cheers and our heartfelt gratitude.

Such an analysis of arbitrage, speculation, and the like requires two comments. The comments lead to an intricate analysis in the remainder of this section that you might as well skip on first reading.

1. The future is uncertain. Our gratitude would sour if the speculators were wrong. If the harvest were in fact very good instead of very bad, the stored coffee released in the good year would merely reduce the price further, causing a social *loss* relative to no storage rather than a social gain. The storage and resale would amount to transport from a region of scarcity and high prices to a region of abundance and low prices, the opposite of the desirable flow. Middlemen cannot always be right. Sometimes they will expect next year's crop to be bad when in fact it will be good, leaving them (and society) with still more coffee in a year of abundance. Wrong arbitrage is bad.

That middlemen are sometimes wrong, however, is not an argument for outlawing them. Unless someone else—God, the Department of Agriculture, or whoever—can do better, and supposing that the predictions are not usually wrong, a case can be made for leaving the middlemen alone to put their money where their mouths are. Unlike their critics, after all, they back their predictions with their money. If they make money, their predictions were on average right, and society is on average better off. If they lose money, their predictions were on average wrong, and society is on average worse off. And the arbitrators who lose money do not stay around to make further bad predictions. Exactly contrary to the teachings of ersatz economics, you see, the more profitable the arbitrage the greater the good the arbitrators do for society. Profitability on average is the guarantee of goodness on average.⁸

2. Another comment is that consumers alone would prefer to stop the middlemen. The comment depends on the distribution of the gains. The details of the distribution of the gains are veiled when the analysis uses only "vertical" areas of willingness to pay, as it has so far. There is another way of looking at the question, using *horizontal* areas, that lifts the veil. It comes to the same result if done correctly. If done incorrectly, as it frequently is, however, it leads to strange results. First, a routine point:

T or F: For a straight-line demand curve, the larger the amount of coffee already being consumed, the larger

the gain to getting it even cheaper and the larger, therefore, the increased consumers' surplus.

⁸ The diagram used is, in truth, ill suited to speaking of goodness or rightness or anything else "on average." It contains no uncertainty about the future crop. The middlemen in such a case are perfectly certain in a good year that next year will be bad. They are in fact mere storers and transporters, not "speculators," for what speculation is there in a certainty? A small amendment to the diagram given as a problem, however, brings in uncertainty. The amendment is due to Bart Taub.

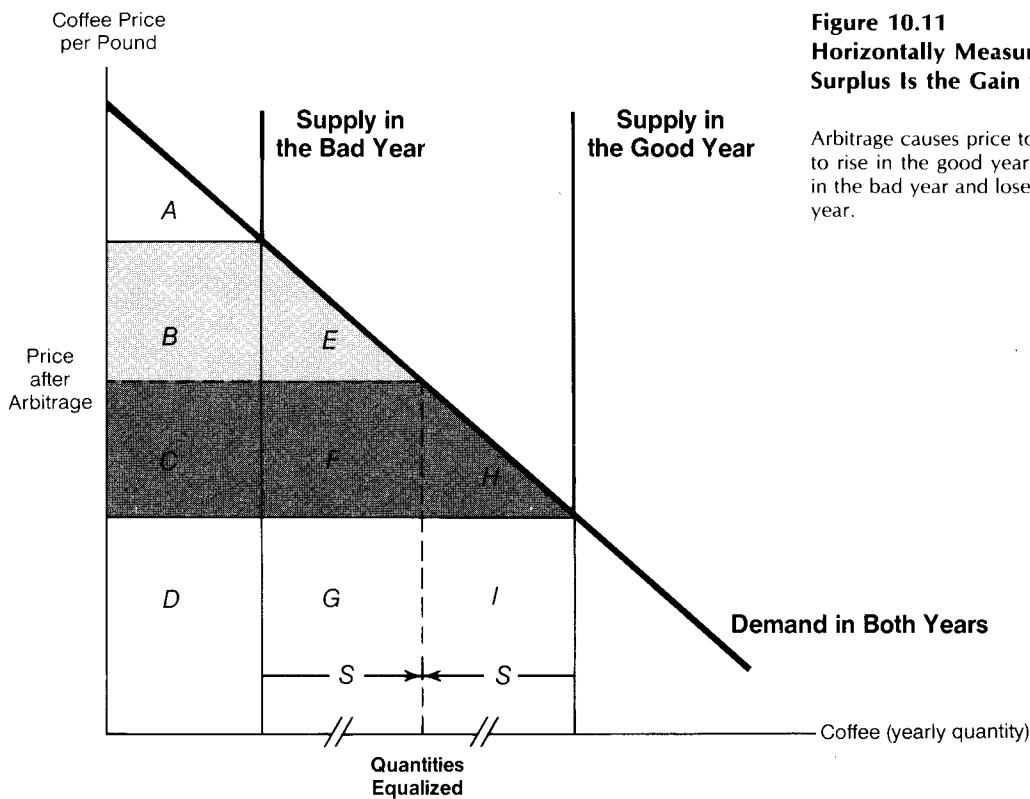


Figure 10.11
Horizontally Measured Consumers'
Surplus Is the Gain to Consumers Alone

Arbitrage causes price to fall in the bad year and to rise in the good year. Consumers gain $B + E$ in the bad year and lose $C + F + H$ in the good year.

A: For example, consumers of the Supply in the Bad Year in Figure 10.11 get only area $B + E$ in increased consumers' surplus if they are allowed to buy from the good year at the lower price after storage and sale. Work it out step by step. Before arbitrage they paid $B + C + D$ for coffee, giving them $B + C + D + A$ in satisfaction. This left A in consumers' surplus. After arbitrage they pay $C + F + D + G$ for coffee, giving them $F + C + D + G + A + B + E$ in satisfaction. This leaves $A +$

$B + E$. The difference between consumers' surplus before arbitrage and after arbitrage—the net gain—is $B + E$. By the same reasoning a move from paying for Quantities Equalized at the high price to paying for Supply in the Good Year at the lower price gives the consumers the much larger area $C + F + H$ in increased consumers' surplus. That is, the more they consume, the more they like it to be even cheaper. So, true.

The strange implication of this argument is as follows.

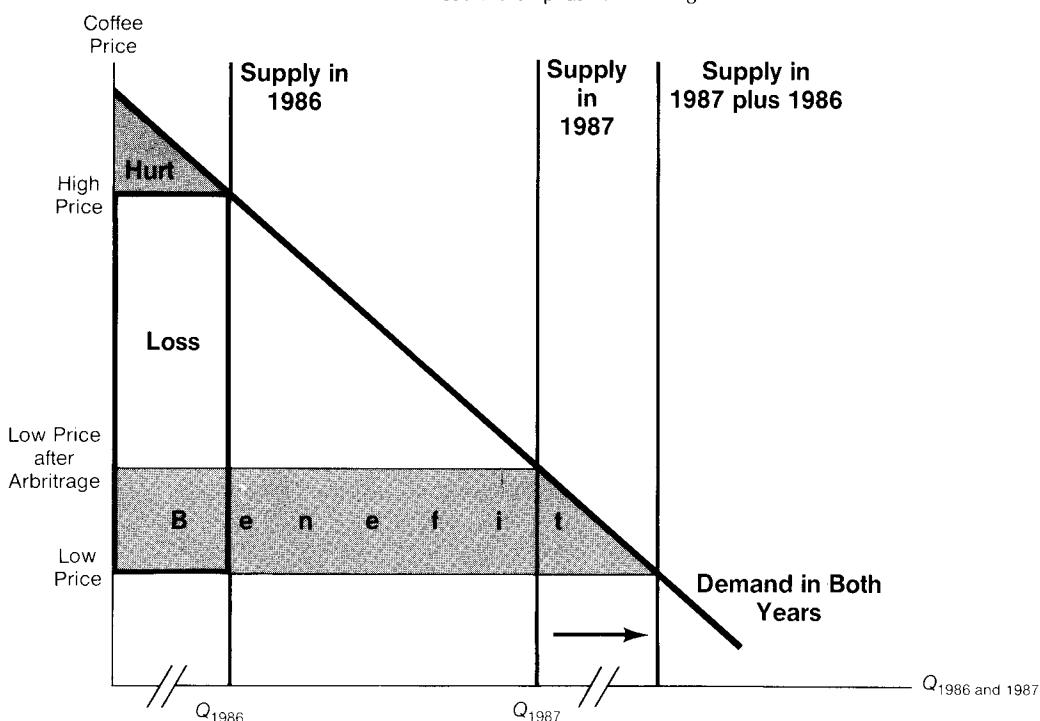
T or F: Far from wanting a smooth path of consumption, consumers will sometimes prefer to buy all the output in one year and none in the other.

A: With identical straight-line demand curves in two years, for example, the consumers get the most consumers' surplus when all the supply is shifted to one year, driving down the price in the year. The supply of 1986

shifted to 1987 causes the area Hurt, but is more than offset by the area Benefit (see Figure 10.12). The consumers by themselves would like to see coffee moved from a year of scarcity to one of abundance. So, true. This may explain the hostility of consumers to middlemen, arbitrators, speculators, and other movers of goods from low to high prices.

Figure 10.12
Consumers Can Prefer Unstable Prices

If all the coffee were sold in 1987 at Low Price, consumers would gain area Benefit and lose area Hurt. Consumers would gain if arbitrage did not occur. But arbitragers would lose the emphasized rectangle Loss.



A Solution to the Apparent Paradox

But would anyone else like it to be so moved? No, they would not. Buying the amount indicated at High Price and selling it at Low Price gives the emphasized area of Loss. No one would engage in such a silly deal voluntarily. In other words, consumers will not be able to find suckers with whom to make the deal of cheapening cheap coffee still more (by buying expensive coffee and dumping it on the cheap market). As Paul Samuelson has put it, the experiment would require an "outside Santa Claus."⁹

The key to the correct result is to include all members of society—a Christmas with mutual gift giving but without a white-bearded visitor on a sleigh. Someone owns the coffee. The gains or losses to people as *consumers* of coffee are matched by gains or losses to people as *owners of stocks* of coffee. The gains and losses include not only those on coffee stored and resold but also those on coffee owned. If arbitrage causes the price of coffee to fall below what it would have

⁹ Paul Samuelson, "The Consumer Does Benefit from Feasible Price Stability," *Quarterly Journal of Economics* 86 (August 1972): 476–493. His article is a correction of an influential article by F. W. Waugh in 1944 that made the error of leaving out the suppliers. See also B. F. Massell, "Price Stabilization and Welfare," *Quarterly Journal of Economics* 83 (May 1969): 284–298.

been in 1987, the owners of a given amount of 1987 coffee are made worse off by having less command over resources than they would have had.

The earlier Figure 10.11 can be seen to contain such areas of gain and loss if looked at closely (see Figure 10.13, which reproduces it). Consumers in the good and bad years gain and lose the "horizontal" areas $B + E$ and $C + F + H$. To collapse the rest of society into one group, suppose that the original owners of coffee are the middlemen. Owners of coffee in the bad year lose the area B through the storage and resale of coffee in the good year. That is, the owners of coffee in the bad year lose because of the reduced value in the good year of the existing coffee. Likewise, owners of coffee in the good year gain the area $C + F$, that is, the increased value of the coffee owned and consumed in the good year. They also gain the profit on storing the quantity stored and selling it at the price after storage and sale. The area is the two triangles $H + J$ or equivalently (since what is stored in the good years equals what is sold in the bad) the rectangle F . The complete analysis, then, adds up in both years all the gains (+) and losses (-) for both consumers and owners, as shown in Table 10.1.

The important point is that the net gain, $E + F - H$, is exactly the same as

Figure 10.13
The Gains and Losses to Owners

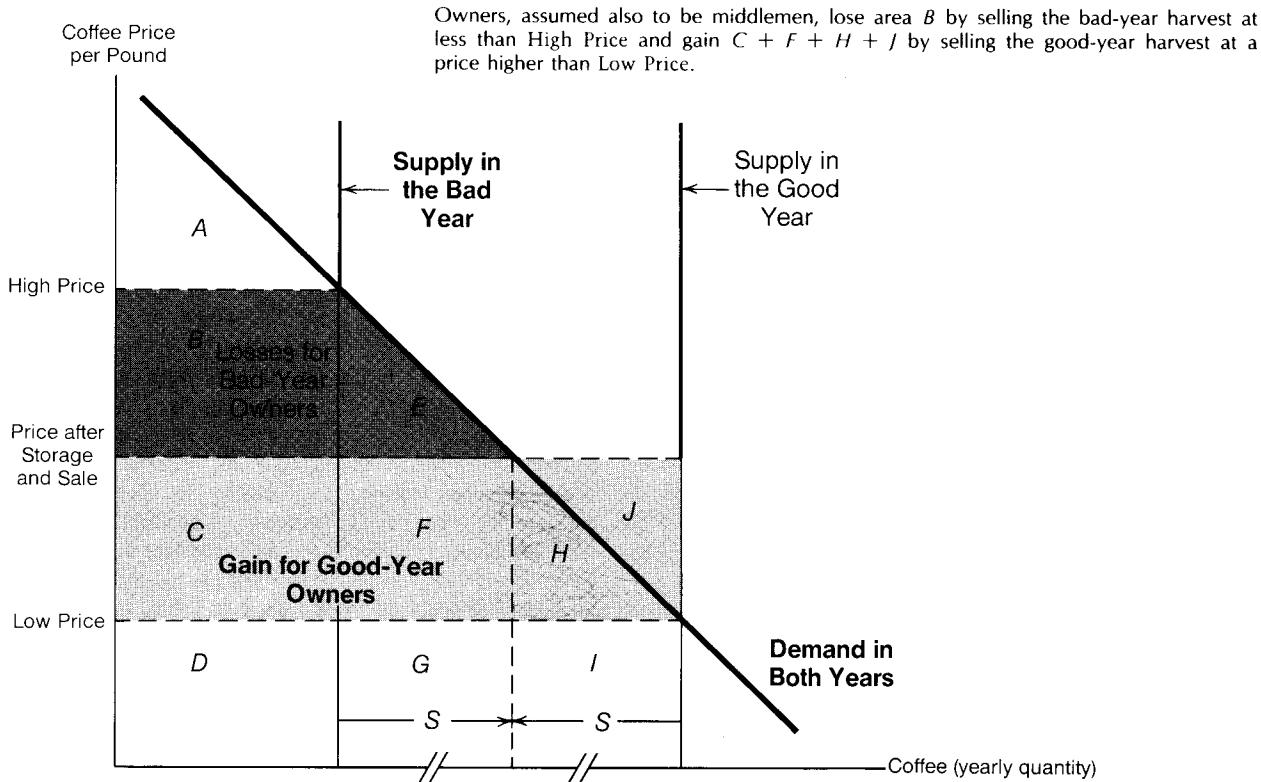


Table 10.1
The Balance of Gains and Losses¹

To	Areas	Description
Bad-year consumers	$+B + E$	Consumers' surplus gained
Good-year consumers	$-C - F - H$	Consumers' surplus lost
Bad-year owners	$-B$	Inventory value lost
Good-year owners	$+C + F$	Inventory value gained
Plus	$+H + J = +F$	Profit on stored and resold coffee
Society as a whole ²	$+E + F - H$	Net social gain from storage and resale

¹ Gains (+) and losses (-) from storage and resale.

² Add up all of the above.

the net gain calculated earlier by looking at the vertical areas of willingness to pay. The horizontal consumers' surplus amounts to the same thing as the vertical consumers' surplus. The vertical analysis is appropriate when the question is one of income as a whole, the horizontal when the question is one of the distribution of income among various participants in the market. The method of vertical areas ignores the distribution of the gain but is simpler and more sure; the method of horizontal areas exhibits who benefits but is more complicated and easier to get wrong. Take your choice.

Other Sorts of Trade

The examples so far have involved trade across time. All the principles of using consumers' surplus vertically and horizontally can be illustrated for trade across space.

Q: A popular policy in times of unusual scarcity is to forbid exports. Thus, if wheat prices rise after a bad world harvest and American harvest, American consumers will clamor for restrictions on exports of American wheat to China or the Soviet Union. Suppose that the world price is fixed. Using both the horizontal and vertical approach to consumers' surplus, describe diagrammatically the effects of forbidding exports:

1. The loss to Americans as a whole.
2. The gain to American consumers.
3. The loss to the owners of the American wheat crop (after the harvest: ignore costs of production).
4. Compare (1), (2), and (3).

A: Figure 10.14 portrays a World Price high enough to cause Free Exports if they were permitted. The perfectly inelastic Supply curve expresses the costlessness of the wheat crop after it has been produced. Part 1 is best answered with vertical areas, for it treats the nation as one enterprise. When trade (Free Exports) is permitted, the nation as a whole earns the value (in all

other goods) of the rectangle $D + E + F$. But the nation would be willing to pay only the trapezoid $E + F$ to use that quantity for its own consumption rather than export it. The net social loss from forbidding exports, forcing all the supply onto the American consumers, is therefore the shaded area D . Part 2 is a question about the gain to one segment of society, and is best answered with horizontal areas. At the Autarky (no trade) Price, consumers get $B + E$ more consumers' surplus. No wonder they clamor for restrictions. If the world price were lower, presumably the clamor would be lower as well. Part 3 is simply a matter of rectangles of revenue: At the World Price the owners of the fixed supply get $C + F + (B + E + D)$; at the autarky price they get only $C + F$. They lose $B + E + D$ on balance. But according to (2), the $B + E$ portion of the loss is a gain to consumers. So—to answer 4—the loss to American owners over and above what American consumers gain is the shaded area D . The vertical and horizontal methods yield the same result.

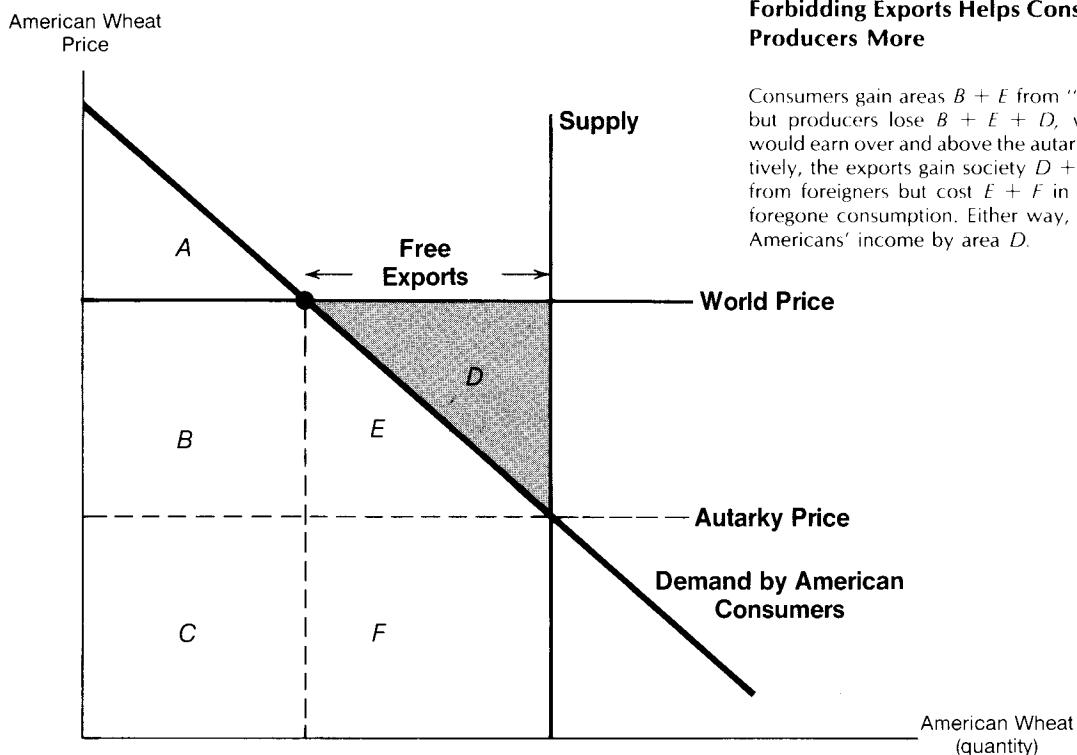


Figure 10.14
Forbidding Exports Helps Consumers but Hurts Producers More

Consumers gain areas **B** + **E** from "autarky" (no trade), but producers lose **B** + **E** + **D**, which is what they would earn over and above the autarky revenue. Alternatively, the exports gain society **D** + **E** + **F** in payments from foreigners but cost **E** + **F** in opportunity cost of foregone consumption. Either way, free trade increases Americans' income by area **D**.

Summary You are now equipped with a powerful tool called consumers' surplus. It can be used two ways, vertically and horizontally, which yield the same answer if the question is the same. The horizontal way emphasizes the distribution of gains and losses; the vertical emphasizes their size. The key to using the tool correctly is to include all the people in the analysis. An analysis of middlemen that ignores owners of the coffee or copper or wheat, for example, will arrive at the result that middlemen are bad. But the correct analysis shows that middlemen—like other traders between times and latitudes—are on the whole good. Consumers feel that middlemen are bad for them, and they are perfectly justified in the feeling. But consumers are not the whole of society; or, to put it another way, people play many roles, only one of which is the role of consumer. The whole of society includes people in their roles as owners, too. If you shift a stone from your left hand to your right, your left hand is relieved, but you as a whole are not. Staring fixedly at one hand or at one party to a transaction is a mistake if the question is about the whole body. A question about wholes requires an analysis of wholes. Consumers' surplus is the economist's way of adding up whole groups of consumers, owners, producers, and people in general.

EXERCISES FOR SECTION 10.2

1. Robinson Crusoe lives alone on a desert island, eating coconuts. The coconuts are available on the trees only half the year. Describe what he does to keep himself from

starving when the trees have no coconuts, and compare it to Teresa Baker's activities in the coffee market.

2. Why in Figure 10.10 does the area *H* (of loss) cancel exactly against the area *E* (of gain)? (*Hint:* Would it cancel if the demand curve were not straight?)

3. How do the following unsavory types make society better off?

- a. The scalper of basketball tickets.
- b. The blackmarket seller of automobile tires in a country where tires are strictly rationed.
- c. The profiteering oil company, raising prices now in anticipation of an oil shortage.

4. *True or false:* Consumers are happier when prices fluctuate than when they are smoothed out by speculators because the consumers can buy a lot when the price falls low, and refrain from consuming when it rises high.

PROBLEMS FOR SECTION 10.2

- **1.** True speculation would involve uncertainty about the crop. The following problem shows how to allow for uncertainty. Suppose that there are only two sorts of years for the coffee harvest, good and bad, the one twice as large as the other. Suppose now that you are in a good year and are deciding whether or not to store 10% of the crop in the hope that next year will be bad. Suppose that coffee can be stored only one year, at which point it must be sold. Suppose further that there is a 50–50 chance that next year will be bad. And suppose finally that the demand for coffee is the same straight line in all years.
 - a. You buy and store and sell next year 10% of this (good) year's crop, paying and getting money along the demand curve. Illustrate in a diagram how much you as a middleman make if next year is bad. Use Figure 10.9 as a model.
 - b. What do you make if contrary to your hope next year is good? Note that the price will be *lower* because you'll add to the quantity in a *good* year.
 - c. Answer (a) and (b) for the case of buying the *next* 10% of this (good) year's crop.
 - d. If you and your many competitors are neutral toward risk, at what point do you all stop buying up this year's crop to store and sell next year? That is, at what point do you expect, on average, to make no profit? Compare the result with that of perfect certainty about the future (that is, the result of the earlier exposition).
- 2.** The chance of next year's crop being bad was 50% in Problem 1. If the chance were 25%, what would happen to the optimal amount stored?

True or False

- **3.** If the demand curve for coffee is approximately a straight line, the social gain from costless storing of coffee in anticipation of a perfectly foreseen bad crop is approximately half the difference between the unarbitraged quantities multiplied by half the difference between the unarbitraged prices.
- **4.** If an anticipated fall of 10% in the wheat crop would cause a 20% increase in price along a straight-line demand curve without storage, allowing (free) storage to occur will increase world income by $(10\%)(20\%) = 2\%$ of the value of the wheat crop.
- 5.** Consumers by themselves prefer instability in prices, even though society as a whole prefers stability in supplies (and therefore in prices).
- 6.** If middlemen buy and sell all the coffee stored in accurate anticipation of a harvest failure at the price after arbitrage, their profits are zero and the social gain is also zero.
- 7.** If demand curves for a storable commodity are not straight lines, consumers might benefit (even strictly as consumers) from shifting output from a good year to a bad year.

- 8. For nonstraight, nonidentical demand curves in the good and bad years, middlemen will shift output from good to bad until prices are equalized.

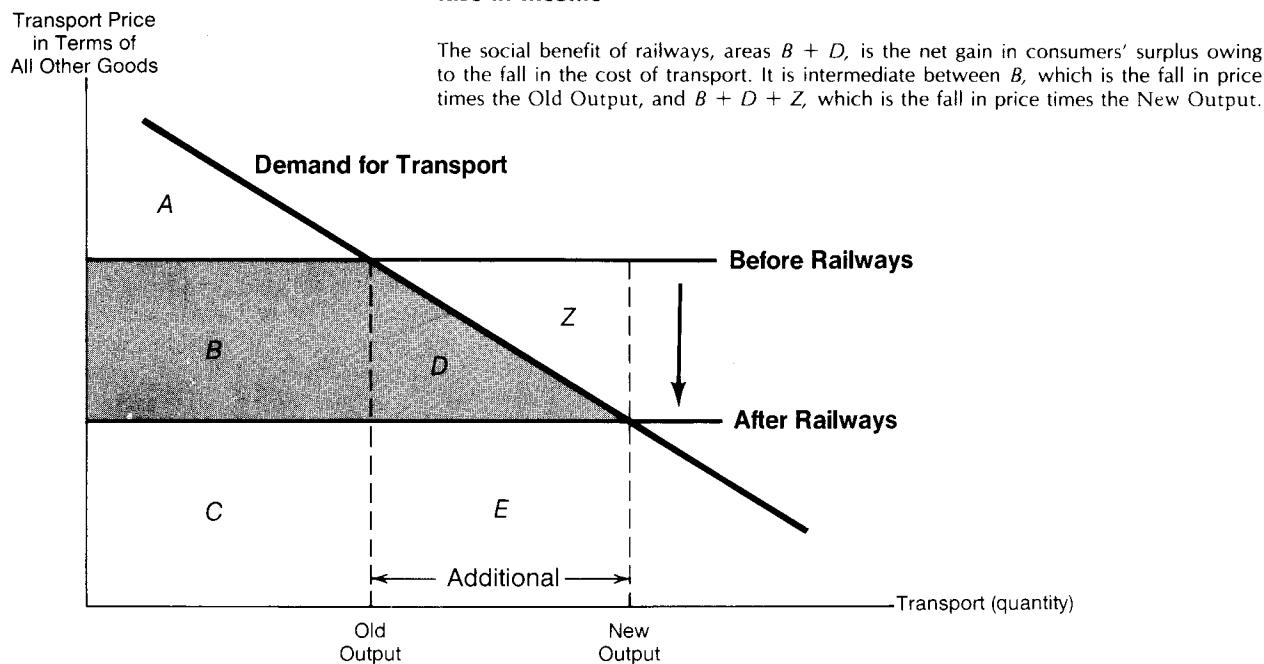
10.3 A Change in Consumers' and Producers' Surplus Is the Same as a Change in National Income

What to Read For What is the benefit from cheaper railways, expressed in terms of consumers' surplus? What is the relation between this benefit and the Laspeyres measure of the change of national income? The Paasche measure? Which is an easily observable upper bound on the "true" benefit from cheaper railways? Is the change in consumer surplus any more difficult to measure than the change in national income?

Surplus as a Tool in Cost-Benefit Analysis

Consumers' surplus measures the benefit side of costly projects. The railways of the nineteenth century, for example, were so beneficial and costly that many observers have believed they were the main cause of economic growth. Railways were spectacular machines, built by giants and run by heroes, and it is therefore not astonishing that they have been thought to be epoch making. Stripped of their glamor, however, they were merely a cheaper way of moving some sorts

Figure 10.15
Railways Led to a Fall in the Cost of Transport, and the Fall in Cost Is a Rise in Income



of freight and passengers than the canals, rivers, and highways with which they competed. As a result of the coming of railways, in other words, the opportunity cost of transport per ton-mile or passenger-mile fell.¹⁰ One can think of the society, then, moving down the demand curve for transport, as in Figure 10.15.

The benefit from the coming of the railways is simply the horizontal area $B + D$, that is, the increase in consumers' surplus. Recall that areas in such a diagram are measured in units of all other goods. Before railways, getting $A + B + C$ in willingness to pay required the sacrifice of the rectangle $B + C$ of all other goods in opportunity cost. After railways, getting the larger area $A + B + C + D + E$ required the shorter but broader rectangle $C + E$ of all other goods in cost. The net gain in consumers' surplus over opportunity cost is the shaded trapezoid $B + D$. To put the matter another way, consumers of transport got the old quantity at a cheaper price in opportunity cost by the amount B , and furthermore they got additional transport at the low bargain price After Railways (for benefit of D). The total benefit to society was the saving in opportunity cost on the old output (namely, B) plus the advantage of buying additional transport at a price below its value in use (namely, D).

The Change in Surplus Is the Change in Income

The argument so far has just restated the idea that the sum of consumers' and producers' surplus measures the benefit from a new device, such as the railway. Another way of measuring benefit is national income. The similarity of vocabulary between "surpluses" and "national income" suggests a similarity in substance. As a matter of fact, *national income goes up after the railway by exactly the same amount as does the sum of consumers' and producers' surplus*.

That national income goes up by some amount is no surprise, since railways are a better way of transporting and transport is part of the bundle of things called "national income." But that it goes up by the amount of the shaded area of consumers' and producers' surplus in a diagram of demand and supply is a great surprise and requires proof, as follows.

Return to Figure 10.15. The rectangle B is the portion of the shaded benefit coming from cost saving. Since the relative price is expressed in amounts of all other goods per amount of transport, the area is the additional amount of wheat, houses, iron, education, and all other goods the society can have if it can provide the amount marked Old Output at the low, After Railways price. In other words, B is one measure of the increment to national income. It is merely one measure of income because it chooses one of many solutions to the index number problem. The opposite solution is to look at the cost saving from the point of view of the larger New Output. The rectangle area $D + Z$ added to the rectangle B is the cost saving on the New (as opposed to Old) Output. If the New Output had to be produced by Before Railway techniques, it would cost in forgone amounts of all other goods the whole rectangle $B + D + Z$. The two areas of cost saving (that is, B by itself and B plus D plus Z) are alternative measures of the increase in national income. But note well that these two alternative measures of increased income bracket the area of

¹⁰ This thought and its elaboration is due to Robert W. Fogel, *Railroads and American Economic Growth* (Baltimore: Johns Hopkins University Press, 1964); Albert Fishlow, *American Railroads and the Transformation of the Ante-Bellum Economy* (Cambridge, Mass.: Harvard University Press, 1965); and dozens of similar studies in the economic history of transportation published since then.

net consumers' surplus (the horizontal trapezoid $B + D$). Therefore, to perform the final twist, the area of net consumers' surplus is a *compromise* measure of the increase in national income, which falls between the two alternative measures. In other words, to the accuracy allowed by the index number problem, the rise in consumers' (and producers') surplus *is* the increase in national income. This is the surprising truth.

The two areas of cost savings, in fact, can be shown to be precisely the Laspeyres (early prices weighting output) and Paasche (late prices) measures of the change in national income resulting from a fall in transport costs. Take the Paasche as an example.

Q: Using the diagram of production possibilities between transport and all other goods (Figure 10.16, bottom panel), show that the area B (in earlier diagrams) is the Paasche measure expressed in all other goods of the rise in national income.

A: To establish such correspondences requires a two-diagram picture familiar from earlier chapters. The steps are numbered in the diagrams in Figure 10.16. Put the supply and demand diagram on top of the diagram of production possibilities (note that the lines are straight, to capture the constancy of opportunity cost).

The dashed line (marked 10) in the bottom panel is drawn parallel to the new, farther-out production possibilities curve that corresponds to the New Price. The dashed line has the same slope as the new production possibilities curve.

The vertical distance marked C in the bottom panel is the same as the area C in the top panel. To see this, consider the following. The New Price after railways is the relative price of transport expressed in terms of all other goods. The New Price is also the opportunity cost of transport in terms of all other goods. The slope of the production possibilities curve, which is the same

as the slope of the dashed line marked 10, is equal to the opportunity cost of transport in terms of all other goods. The slope of the dashed line is therefore equal to the New Price. But the slope of the dashed line is also equal to the vertical distance marked C divided by the Old Output (in other words, the rise over the run). The slope of the dashed line multiplied by the Old Output is equal to the vertical distance C . But the New Price multiplied by the Old Output is also equal to the area C in the top panel. The vertical distance C is therefore the same as the area C in the top panel.

By identical reasoning (the Old Output multiplied by the Old Price), the vertical distance marked $C + B$ is exactly the same as the area $C + B$ in the top panel. It follows that the differences between the two are equal. The emphasized vertical line B in the bottom panel is equivalent to the shaded area B in the top. B is B . But B in the bottom is the Paasche measure (expressed in all other goods) of the rise in national income, since it is the distance measured in new prices between the value of the Before and the value of the After bundles of transport and all other goods. The area B in the top panel, then, is the Paasche rise in national income.

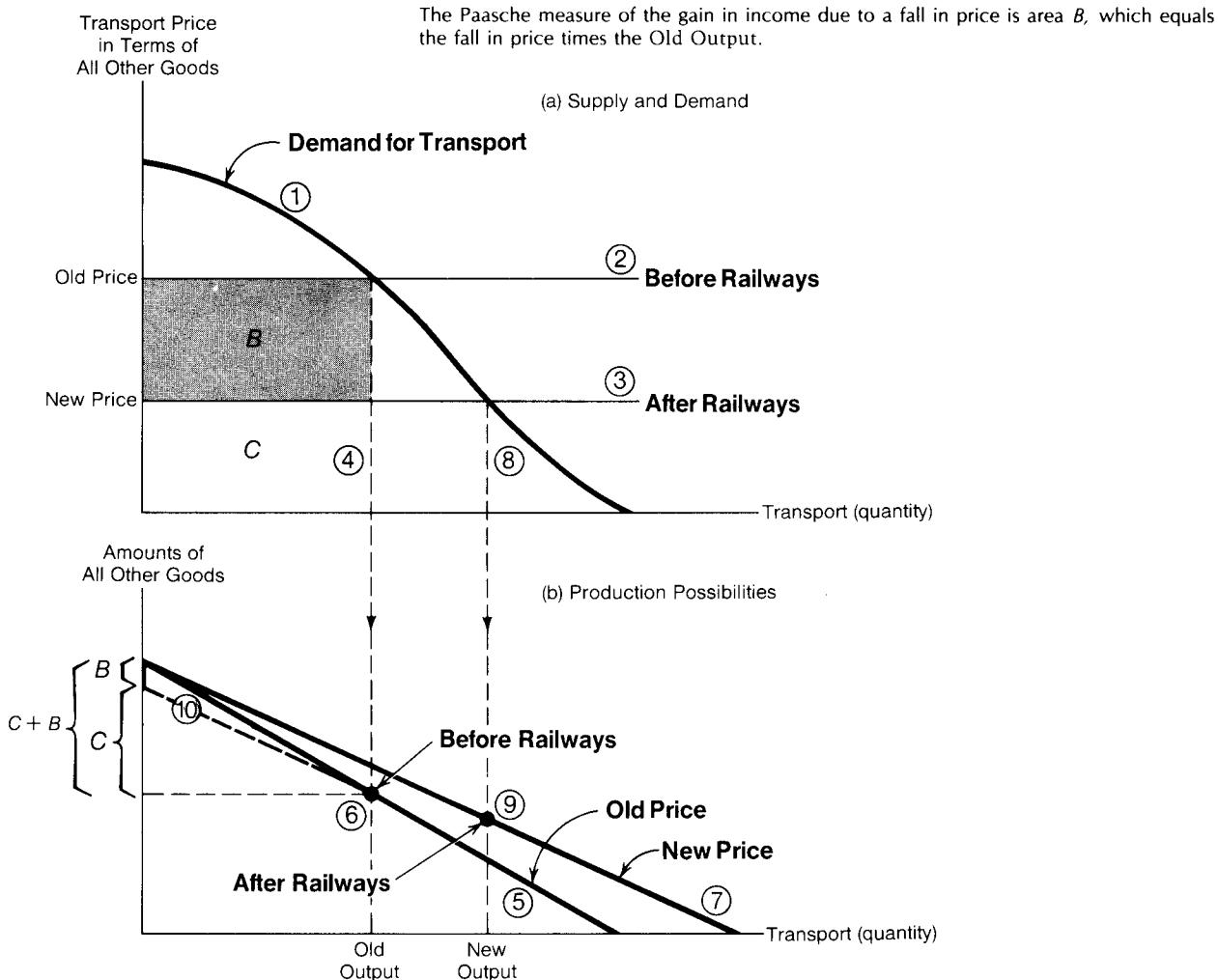
A similar proof applies to the Laspeyres rise. The benefit from railways measured as consumers' surplus lies between these two measures of national income.

T or F: If American transport of crops was 1 billion units in 1890, if the price of carrying it without the railways would have been \$0.50 more per unit than it actually was, and if American national income was \$12 billion in 1890, the increase in national income from the cheaper carrying of crops was 4.2% at most.

A: The statistics give the dimensions of the rectangle of cost savings on the after-railway quantity of transport (that is, the large rectangle $B + D + Z$ in earlier diagrams). The area of the recentangle is 1 billion units

of transport times \$0.50 more cost per unit of transport, which equals \$0.50 billion in more cost of transport without the railway, or $\$0.50 \text{ billion} \div \$12 \text{ billion} = 4.2\%$ of 1890 national income. So, true. The increase is "at most" 4.2% because the rectangle is the larger, upper bound (the rectangle using the Before Railway's quantity is the smaller, lower bound). With such methods, Robert Fogel was able to assert that in American economic growth during the nineteenth century railways were far from indispensable.

Figure 10.16
Laspeyres and Paasche Measures of National Income Are Upper and Lower Bounds on the Benefit in Consumers' Surplus



What makes the calculation simple is that you do not need to know the shape of the demand curve; to get an upper bound, all you need to know is the one point on it actually achieved in 1890. By assuming, so to speak, that the demand curve for transport was perfectly inelastic, you get an upper bound on the true rise in consumers' surplus. The question arises, "If you know the demand curve, can you refine the estimate of the benefit?" Yes, you can. But the refined measures are ambiguous, and it is doubtful that the refinement gains much in the end. The bounding of the surplus by two measures of national income is the crudest but most important truth.

Surplus Has the Same Difficulties as Income The change in consumers' surplus, then, is the same as the change in income. It is not a more precise method demanding more stringent assumptions for its validity. One will hear it said that consumers' surplus requires the "constancy of marginal utility of income" or some other more or less improbable-sounding condition to be true. This is false. Consumers' surplus is a true measure of social happiness to the exact extent national income is.¹¹

National income, of course, is a doubtful measure of social happiness. At the technical level, both national income and consumers' surplus must face the index number problem. At the moral level, unless a change in national income has been generated by a just society with just rules for making alterations in its affairs, a rise in national income—whether it is called a rise in income or a rise in consumers' surplus—is not morally decisive. But if the criticisms of national income apply also to consumers' surplus, so too do the replies: that a rise in the sum of producers' and consumers' surplus (to use the more comprehensive term) is morally defensible as the sort of change one might want to encourage from behind a prenatal veil of ignorance of where exactly in society one would end up when born; and that morally defensible or not, the rise and fall in the surplus is a statistic one would in any case like to know, the better to balance it against other considerations.

Summary

The change in consumers' and producers' surplus is a measure of the benefit from some innovation, such as the railway. The measure turns out to be, in fact, a compromise estimate of the change in national income caused by the innovation, lying somewhere between the Paasche and the Laspeyres measures of the change in national income. In other words, the change in the surplus is the change in income. It is not a more elaborate or more questionable measure. It is questionable to the same degree as national income is questionable. And this is scarcely odd, because the two are the same.

EXERCISES FOR SECTION 10.3

1. You consume now 100 hamburgers a year at \$2 each. The price falls to \$1 each, at which price you consume 150 a year. What is the cost saving on the old consumption (100 hamburgers)? What is the cost saving on the new consumption? If your demand curve for hamburgers was a straight line, what would be your increase in consumers' surplus arising from the new, low price?
2. Why doesn't the society gain the area $C + E$ in Figure 10.15? (Hint: What does it go for?)

PROBLEMS FOR SECTION 10.3

- 1. The total social benefit from a railway can be divided up among separate demanders and suppliers. It is clear that each of, say, two demand curves, one for the West and the other for the East, could be faced separately with the same fall in transport costs. Can they be faced with it together, in one diagram of Western and Eastern demands added together? (Hint: Do it both ways and compare.)

¹¹ Compare A. C. Harberger, "Three Basic Postulates for Applied Welfare Economics," *Journal of Economic Literature* 9 (September 1971): 785–797.

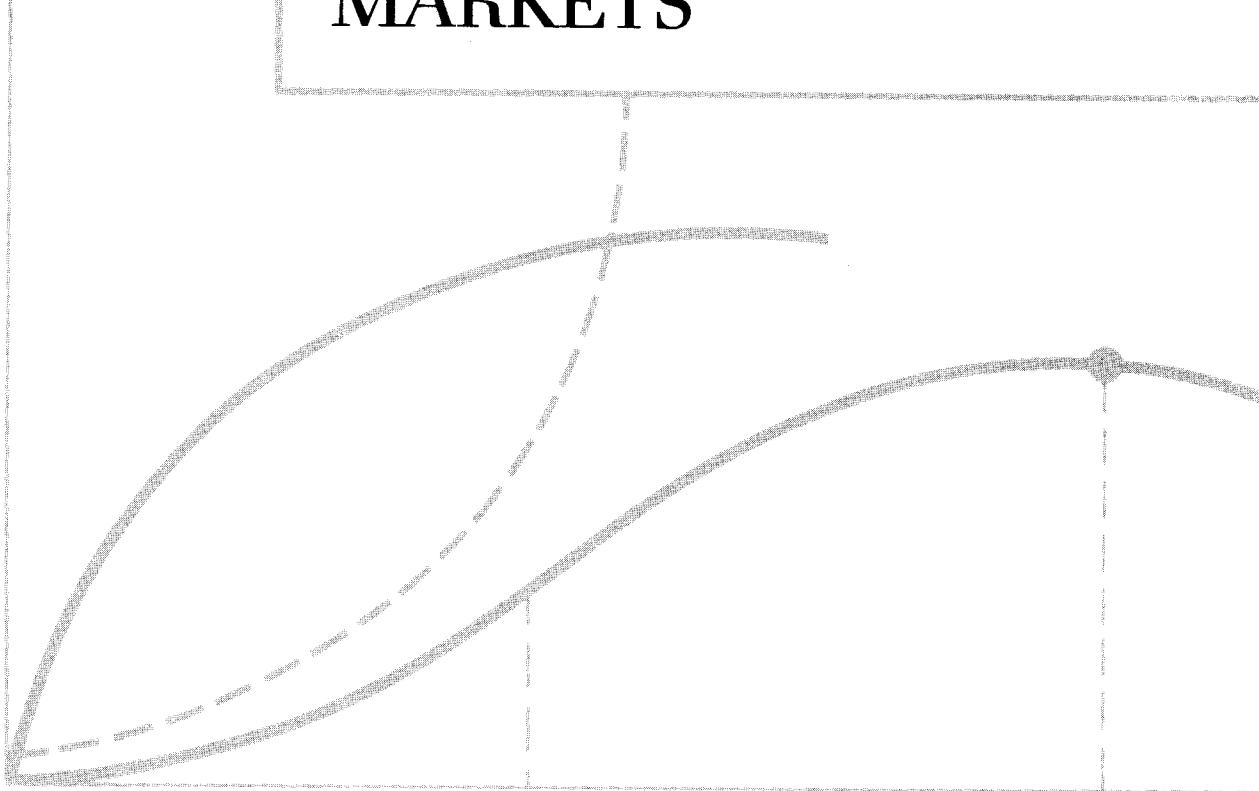
2. Prove that the Laspeyres measure of the rise in national income from the invention of railways is the cost saving on the After Railways amount of transport.

True or False

- 3. Even if the supply curves of transport were upward sloping before and after the railways, the area of social benefit would have the same definition as with flat supply curves, namely, the area bounded by the axes, the demand curves, and the lower and higher supply curves.
- 4. If the demand curve is a straight line, the benefit from railways measured in consumers' surplus is exactly halfway between the Laspeyres and Paasche measures of the rise in national income caused by the railways.

IV

PRODUCTION AND MARKETS



The Firm

$\pi^* = \text{good enough}$

Firms = economist's theory of cooperation.

Theory of firms as anti-social/polymerization
blind economic language.

Value of firm is true maximum. $S = LR = SR$

Q is just one of many control variables.

11.1 Whether and Why Firms Exist

What to Read For

Whom do you buy from when you buy an Oldsmobile? What is the wheel of wealth? Does it mean that firms are without use? In what sense are firms "islands of conscious power in an ocean of unconscious cooperation"? What alternatives to firms exist? When does society adopt the firm rather than self-help or the market?

The Firm Is an Agent for Other People

(2) Complexity of
Specialization
Wooden Pencil

A dark place in the picture of the economy drawn so far is that between single people and the whole economy. "The economy" or "the society" is said to choose between guns and butter; "a consumer" is said to buy butter from "a farmer." But in countries without central planning there is no general will choosing guns or butter. And in most countries there is little direct exchange between eaters and farmers. You do not trade, really, with the Iowa farmer Richard Zecher. Millions of people are involved directly or indirectly to supply you with this or that grain of corn.

Between the economy and the individual stands the business enterprise or the government agency or the charitable institution: in a word—the economist's usual word—the "firm." The United States has one profit-pursuing business (usually a tiny proprietorship) for every 15 people. Federal, state, and local governments employ in countless agencies a fifth of the nonfarm labor force. Each of about 7000 hospitals employs on average 400 workers. Only 7% of workers work for themselves; the rest receive a check as employees of some firm. The present chapter casts light on this hitherto dark firm.

But do not let the light blind you to the irrelevance for many purposes of the farmers, manufacturers, wholesalers, retailers, regulators, carriers, bankers, educators, and so forth between you and Zecher. The firm is a mere intermediary: in the vernacular, a middleman. Daniel decides to save some money for a rainy day, putting it in (loaning it to) the Hyde Park Bank and Trust. The bank loans the money to Margaret to buy equipment for her new law offices. Underneath it all, Daniel is loaning money to Margaret, even though we speak of

"the bank" as making the loan. Likewise, Jacob Metzer grows an orange, sells it to a buyer in Haifa, who ships it to a wholesaler in New Zealand, who sends it by truck to a supermarket chain, where it is bought by Lynn Jones. Underneath it all, Metzer is selling to Jones, even though we speak of "the kibbutz" or "the supermarket" making the sale. Again, Larry Westphal works nights at the plant that makes the engine for the Oldsmobile sold to railway executive Steven Weiss. Underneath it all, Westphal is working for Weiss and Weiss for Westphal, trading hours of assembly work for hours of railway execution, even though we speak of Westphal working "for Oldsmobile" and Weiss "for the Milwaukee Road railway."

Looking at a firm without realizing that its people are agents of other people is a common source of error:

(1)
We are
it:
"taxes"
"corporations"
"them,"
not
people"

T or F: In making rental apartment buildings into condominiums (that is, apartments to be bought outright rather than rented), money-hungry bankers, realtors, and developers benefit no one but themselves.

A: These people, so detestably money hungry (unlike the former renters), do benefit themselves. Sure. But they also benefit the present owners of the buildings and the future buyers of the apartments. If they did

not do so they would not make money and their money hunger would go unsatisfied. They buy the buildings from the present owners without coercion. They resell them (improved or not) to the future owners without coercion. Both must feel that the deal is doing them a favor. So, false. The present and future owners may for some reason be less worthy of our concern than the present renters, but they are not "no one."

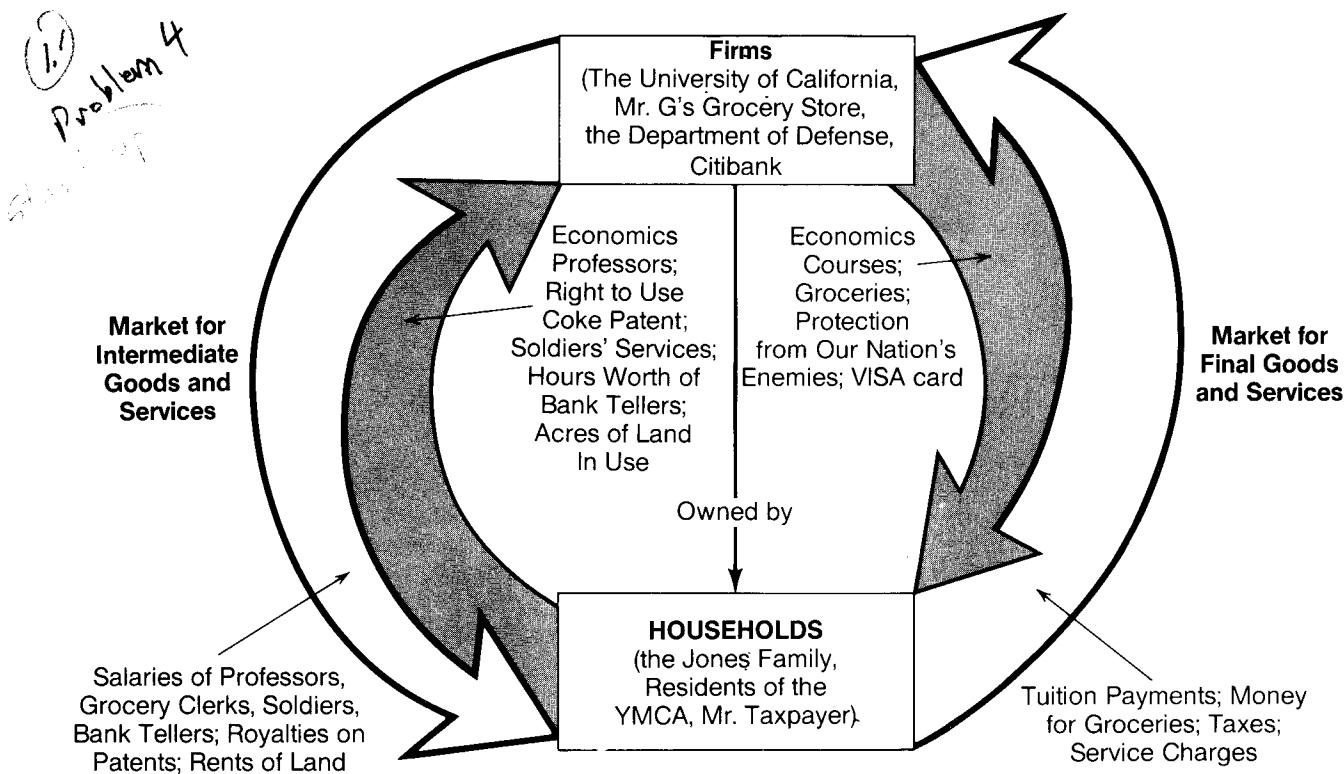
The Wheel of Wealth Deals take place at one point on a *wheel of wealth*. But an economist must look around it. On the bottom of the wheel are households, to which everyone belongs; on the top are firms of various sorts (universities, grocery stores, government agencies, banks), which someone owns. Money payments go around the wheel in one direction, goods and services in another, as shown in Figure 11.1. Only the households really exist; the firms are merely legal fictions. The households own the firms, work "for" them, buy "from" them. Underneath it all, households are dealing with each other. To repeat an earlier formulation, "we has met the enemy and he is us."

Another use of the wheel of wealth depends on its balance. Since firms are merely accounting devices to organize transactions among households, what firms pay to workers, landowners, and capitalists must equal what firms get from customers. The customers *are* the workers, landowners, and capitalists. A seller's income is the same as the buyer's outgo. The seller in this case is the buyer. The resulting balance of the wheel of wealth, with \$6000 per head of expenditures by American consumers equal exactly to their earnings of \$6000 per head, is merely a restatement that income equals outgo.

The consequences are important. For example, contrary to a widely held opinion, inflation does *not* hurt everybody. The inflation in the prices of meat at the grocery store, automobiles in the showroom, and shelter from the landlord is met penny for penny by an inflation of incomes of butchers, farmers, grocers, auto factory workers, and landlords, among others. The inflation of outgoings that we face is offset by an inflation of incomes. We is the enemy, simply because we are buying from and selling to ourselves. In moral terms the hurt of inflation to the old woman on a fixed income in money is not balanced by the land speculator's benefit, but in money terms the two cancel. And wage earners, on average, have no legitimate complaints.

Figure 11.1
The Wheel of Wealth

Households earn income by supplying services to firms, which demand the services of households to produce goods that households demand.



One must follow around the wheel the consequences of a shock. For example, a rise in meat prices that is a *demand pull* (consumers demand more) may look like a *cost push* if the rise is not traced to its source. Suppose that meat consumers demand more, perhaps because they have become richer. The grocer does not at first raise prices to match. Meat supplies on the grocer's shelves run out unusually fast; the butcher will order unusually large amounts from the supplier, who will order in turn unusually large amounts from the packer and, finally, from the farmer. The farmer will only supply more meat at a higher price. He charges it. The higher price will now be passed back around the wheel of wealth through packer, wholesaler, and grocer. Each will believe, quite reasonably, that the rise in price came from "costs." No one, least of all the consumers and their outraged representatives, will realize that the rise came from the demands of the consumers themselves. Firms hide the underlying transactions. The economist must look around the wheel of wealth to the households of farmers and meat consumers transacting with each other.

**Why Firms Exist:
They Are a Cheaper
Way to Do
Something**

Economists habitually look below the surface of an institution to the relations among people that they insist are "underlying." This is good. But if the habit is carried too far it leads to the wrong conviction that institutions are pointless, mere layers of middlemen clogging the market economy. The man in the street's version of the fallacy is as follows.

T or F: Because it can buy directly from the manufacturer, eliminating middlemen's profits, a large discount store can sell exactly the same air conditioner at a lower price than can a small neighborhood store.

A: The middleman performs services (shipping, stocking, guaranteeing) that the discount store must either per-

form itself or not perform at all. If it performs them itself, it earns the usual reward for the services in the price. If it does not perform them, it is offering less of an air conditioner. An air conditioner without a guarantee, without service, without delivery is not the same as one with them. In spirit, then, false.

The existence of a firm suggests that it does what it does cheaper than some other way of doing it. The discount store probably sells "no-frills" (that is, no-service) air conditioners cheaper than does the local store; and both stores probably sell all sorts of air conditioners cheaper than the average consumer would be able to get them herself, directly from the manufacturer or, still more directly, from the makers of the steel and copper and plastic that go into the air conditioner.

One alternative to a firm, then, is self-help; but the more important alternative is a market. A market, as we have seen repeatedly, induces people to specialize and cooperate *unconsciously*. A firm also induces them to specialize and cooperate, but *consciously*. Firms in the midst of markets are "islands of conscious power in the ocean of unconscious co-operation like lumps of butter coagulating in a pail of buttermilk."¹ In economies in which workers can move to another firm at will, the conscious power of the manager of the firm over the workers is not complete. But within the limits of the contract to which the worker and manager have agreed, the manager is a little dictator, the boss.

The alternative would be a market in which, say, checkers and baggers at a grocery store contracted with the manager to do the checking and bagging for a fee, with the manager required to keep her nose out of their business. The butchers, likewise, could be independent businessmen, as could be the stockmen and bookkeepers. The grocery store could become literally a marketplace, like the central market in many European towns. That the alternative is not adopted widely suggests that in this case "the operation of a market costs something and by forming an organization and allowing some authority (an 'entrepreneur') to direct the resources, certain marketing costs are saved."² In other words, the costs of locating people to trade with on the spot ("Is there a butcher in the house?") or of agreeing on a price for the particular service ("What am I bid for an hour's worth of hamburger grinding today?") may make it worthwhile to hire employees by the month and order them about inside a grocery firm.

¹ Dennis H. Robertson, *Control of Industry* (New York: Harcourt, 1923), p. 85, quoted in Ronald H. Coase, "The Nature of the Firm," *Economica* n.s. 4 (November 1937): 386-405, reprinted in G. J. Stigler and K. E. Boulding, eds., for the American Economic Association, *Readings in Price Theory* (Homewood, Ill.: Irwin, 1952). The view on the nature of the firm expressed here are Coase's.

² Coase, "The Nature of the Firm," p. 338.

(2)
Case of
university.
Apprentice-
ship
in law.
Old
downtown
Specialized
shops →
mall →
dept.
store.
It can
go back,
go to
some
degree
has.
Watch
repair:
jewelries.
or metal
out.

Summary The firm is a clump of consciousness standing between households in the unconscious order of a market economy. Although usually confined to profit-seeking enterprises, the word "firm" and the economic reasoning that accompanies it is in fact also applicable to hospitals, charities, cooperatives, post offices, the Navy, Ohio State University, and the Memphis Fire Department. A firm does not exist independently of the people who own, work for, sell to, and buy from it. Households dealing with a firm are implicitly dealing with other households. The wheel of wealth shows the payments that households make for things from firms and the payments firms make in turn for services from households. One must look around the wheel to understand why taxing "business" is a tax on households or why a "bank" loan is really a loan from one household to another.

But to say that the existence of the U.S. government or the local grocery store "does not matter" because "we own it" is an exaggeration. Of course it matters. The grocery store organized consciously does (presumably) a cheaper job of selling groceries than would a more traditional market organized by exchange. By contrast, a retailing industry organized by exchange into separate grocery, clothing, furniture, appliance, record, book, and magazines stores does (presumably) a cheaper job of selling consumer goods than would an all-purpose local store organized by command. The presumption—by no means always true, but always worth considering—is that the firms that exist do the job best.

EXERCISES FOR SECTION 11.1

1. List all the people you trade with when you buy a:
 - a. Number 2 lead pencil, wooden.
 - b. Hamburger at Roehl's Restaurant.
 - c. Education in economics.
2. You buy the pencil, hamburger, and education from a firm (the stationery store, Roehl's Restaurant, and the college). The firms bring together the ultimate suppliers to make each item available. Describe how the items could be supplied through self-help or through the market operating without the firms.

PROBLEMS FOR SECTION 11.1

- 1. Draw wheels of wealth for societies with:
 - a. Households buying food from grocers, wholesalers, packers, farmers.
 - b. Households' savings deposited in banks loaned to business.
 - c. Households and firms buying importables from and selling exportables to the rest of the world.
 - d. Households paying taxes to the government and receiving government services.
- 2. What is wrong with the bumper sticker "Tax Corporations, Not People"? (*Hint: Who gets taxed when a corporation is taxed?*)
- 3. Only bookstores, not individuals, can advertise for books in *The Clique*, the magazine of the secondhand book trade in Britain. One Mr. S. Murray-Smith complained in a letter to the *Times Literary Supplement* (June 18, 1971) that

What this in effect means is that the customer has to buy his book through two bookshops, rather than one: [the retail shop] . . . and the shop that sells to the first shop. He presumably has to pay two profit margins.

What is wrong with Mr. Murray-Smith's argument? (*Hint:* Is there presumably some reason for restricting *The Clique* to booksellers?)

True or False

- o 4. The nation's product can be measured not only by adding up the value of the goods and services produced for households but also by adding up the incomes of households as workers, landlords, and capitalists.

3

11.2 The Profit Motive

What to Read For

What is the best guess that economics has about what motivates the owner of a firm? Can you always believe people when they claim not to be motivated by money? Will a hired manager always care to do what is best for the owner? How can the owner put a stop to the manager's lack of care? Can he completely stop it? How can the simple money-making theory be used as a standard against which to measure the value of nonpecuniary rewards?

What Does the Firm Maximize?

The theory of the firm asks why firms choose a certain scale of activity. Why, for example, does Fred Carstensen, a farmer in Illinois, produce 30,000 rather than 60,000 or 10,000 bushels of corn? The usual answer is that 30,000 bushels is the most Fred can produce, the notion being that firms produce as much as they can. But the notion is surely false. He could produce more by doing more—applying more fertilizer, working longer hours, hiring extra help—or produce less by doing less.

Likewise, as a mere matter of engineering, General Motors *could* produce more automobiles; the Army Corps of Engineers *could* build more dams than it does; the college of the University of Chicago *could* have 25,000 rather than 2500 students. Firms choose not to produce all they could. A farmer could run a farm in the intensely cultivated fashion of a city garden, getting 200 bushels of corn an acre; he chooses in fact to run it like a farm, getting 100 bushels an acre. Why?

The answer obviously must depend on Fred's motivations, that is, on his utility function, that is, on what he likes. He could like having lots of land, as the lord of a large domain. He could like the soil and its riches, cultivating deep. He could like to follow in his father's footsteps. He could like leisure. He could like independence. And last but not least, he could like having the things that can be bought with money. The making of money for what it can buy is the most powerful description of why people engage in producing for sale things or services. Other descriptions are possible and could form the basis of alternative theories of the business firm. They do in social sciences other than economics. Most governmental firms, for example, appear to be run for the nonmoney benefit of certain groups in the community or of bureaucrats and politicians, not for making money—although the newspapers are filled with scandals consisting of cases in which they *were*, directly. (And indirectly the power exercised in government bestows goods and services.) The motivations of governmental firms is the leading subject of political science. The motivation of other firms (by the economist's widest definition)—churches, armies, neigh-

10 12

borhood clubs—is the leading subject of sociology. Economists, however, believe that the money motive is dominant for business firms and well worth mentioning, at the least, for other firms. In short, Fred produces 30,000 bushels because at 30,000 he makes the most money.

Moneymaking Is a Means to Other Ends

There is nothing very surprising about the assertion that an Illinois farmer farms for money. Nonetheless, the widespread feeling that moneymaking is nasty makes many people unhappy with the assertion. Surely it demeans the nobility and dignity of humankind, they say, to attribute to it these base motives. "The age of chivalry is gone. That of sophisters, economists, and calculators, has succeeded, and the glory of Europe is extinguished forever."

The distaste for moneymaking, indeed, is characteristically (and paradoxically) European.³ No member of European civilization or its offshoots much likes the idea that his dear old Uncle McClelland farming by Cayuga Lake in upstate New York resembles the purest definition of "economic man." Still, it is some comfort to note that although "economic man is incapable of sympathy, benevolence or love . . . he is also incapable of envy, malevolence or hatred. In short, he is splendidly neutral to others." Even the money-grubbing subspecies of economic man has this quality of neutrality.⁴

In any case it is a mistake, though common, to confuse the moral issue of the goodness or badness of moneymaking with the descriptive issue of how common it is. In view of the confusion, the testimony of the money-makers often misleads on the matter of how common it is. You will hear a truck driver declaring that he does not work for money but for his family; or a lawyer that she does not work for money but for the good opinion of her colleagues; or an owner of a shoe factory that he does not work for money but for his own independence and for the well-being of his workers. The declarations can be honest and even true, yet can all be consistent with the pursuit of money. Money buys these things. Even when it is not the plain motive, then, moneymaking pops up.

Q: The hired manager of a large company, it is said, has little incentive to do what is best for the owners of the company. He wants power, prestige, a high salary, and other things for himself, motives quite different from maximizing the owners' profits. Therefore, the "separation of ownership from control" so common in the modern world is said to leave moneymaking (for

the owners) a poor description of the working economy.⁵ *True or false:* If the manager gets power, prestige, high salary, and whatever else he likes by pleasing the owners, the reasoning is false.

A: The business page of the newspaper is filled with reports of takeover bids and of the comings and goings

³ Eastern civilizations are candidly materialistic, contrary to a widely held belief among Westerners that they are holier than we are. Contrast, for instance, the modest request in the Lord's Prayer for "our daily bread" to the prayer to Durga, the Mother Goddess of Hinduism: "Give me longevity, fame, good Fortune, O Goddess, give me sons, wealth, and all things desirable" [Nirad C. Chaudhuri, *A Passage to England* (London: Macmillan, 1959), p. 178; see also his Chapter V, "Money and the Englishman"].

⁴ David Collard, *Altruism and Economy: A Study of Non-selfish Economics* (New York: Oxford University Press, 1978), p. 6.

⁵ This is the usual line of argument against the first generation of economic thinking on ownership versus control, a first generation represented by Thorstein Veblen, and by Adolf Berle, and Gardiner Means in *The Modern Corporation and Private Property* (New York: Commerce Clearing House, 1932) and, of late, J. K. Galbraith in *The New Industrial State* (Boston: Houghton Mifflin, 1967).

of managers. Such events are examples of a market incentive to please owners. Even the supposedly passive owner of a few shares of stock in a large company worries about its value, and when he thinks the management of the company is bad he sells out (probably too late). When many passive owners do so, the fall in stock values makes it likely that there will be a takeover

or merger by another company. In effect the owners hire and fire managers. However remote from the owners a manager seems in the organization of the company, the manager is forced by the market to be close to them in goals—making money for the owners. In short, true: It is not the case that the separation of ownership from control means that the owners' profit will be ignored.

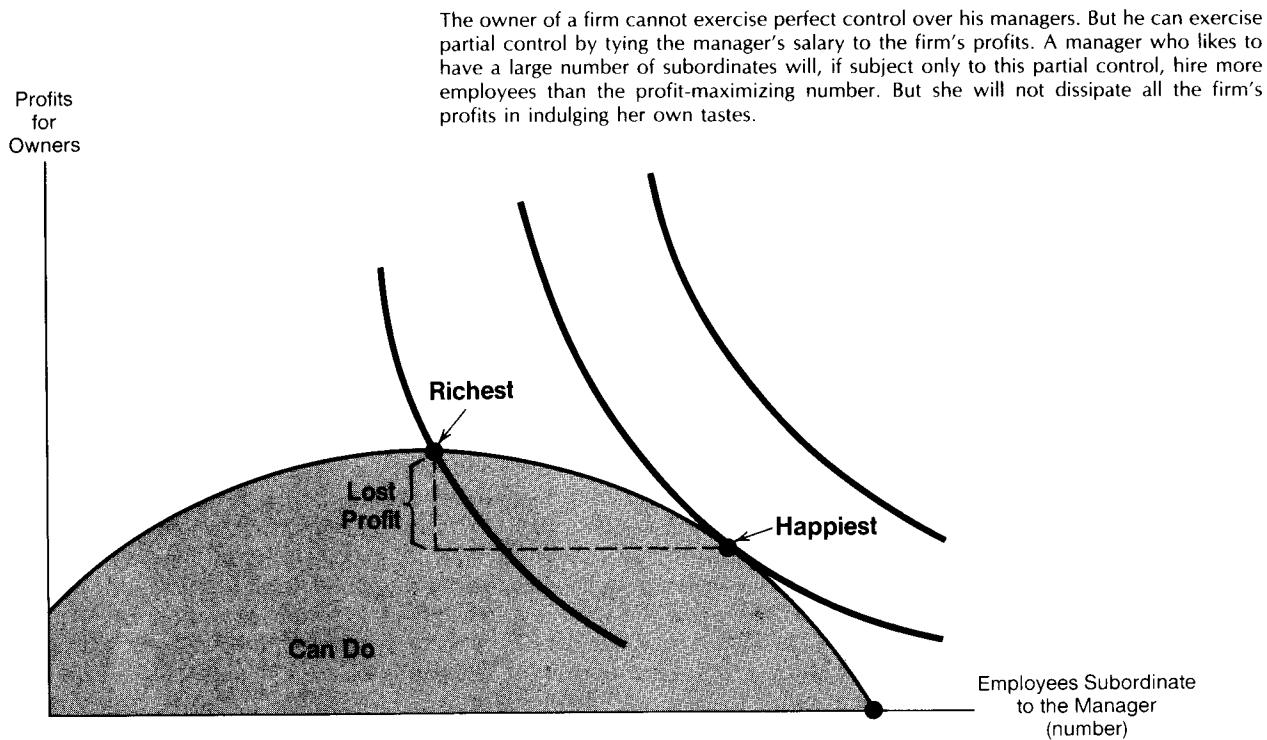
The Agency Problem Modifies the Profit Motive

You should not take the argument just given as a conclusive proof that a manager of, say, Imperial Chemical Industries never works for personal aggrandizement and always works in the interests of its stockholders. Separation of ownership from control might be a serious flaw in modern capitalism. The question is answerable only by fact, not by mere logic. But the logic guides the search. One could ask how frequently in fact bad sales managers lost their jobs; or whether in fact a vice president's lust for power proved to be inconsistent with enriching the owners. Figure 11.2 embodies the logic. Suppose that a manager, Teresa Baker, values the profits of the owners to some degree (or else she will get fired, replaced in a takeover, or passed over for promotion) but also values having large numbers of employees subordinate to her. She has, then, indifference contours in the profit-employees plane. When the number of employees is below the profit-maximizing level of Richest, the manager's goal of moving up her hill of utility is consistent with the owner's goal of high profits. But when the number is larger the goals are inconsistent. The possibilities are represented by the Can Do region in Figure 11.2. The manager is happiest by moving to Happiest, which gives her more of the pleasures of a large staff at the expense of lost profits for the owners.⁶ If the owners could at no cost detect and punish the manager for failing in this way to be a good agent, they could force her to run the firm at Richest for them. The "agency problem" in economic life is precisely that the detection and punishment is costly. It is costly to stop managers from pleasing themselves instead of the owners, just as it is costly to stop government bureaucrats from pleasing themselves instead of taxpayers or to stop professors from pleasing themselves instead of students and other patrons of learning.⁷

⁶ The analysis is that of Oliver E. Williamson, *The Economics of Discretionary Behavior: Managerial Objectives in a Theory of the Firm* (Englewood Cliffs, N.J.: Prentice-Hall, 1964), as summarized in Armen Alchian, "The Basis of Some Recent Advances in the Theory of Management of the Firm," *Journal of Industrial Economics* 14 (November 1965): 30–41, reprinted in William Breit and Harold Hochman, eds., *Readings in Microeconomics*, 2nd ed. (New York: Holt, Rinehart and Winston, 1971). Alchian's article is a good introduction to the second generation of economic thinking on ownership versus control.

⁷ Although such reasoning is widely accepted by economists, it is incomplete, as will become clear in Chapter 25 on labor supply. It speaks as though paying a manager partly in numbers of subordinates were an accident of the technology of detecting and punishing the manager's failures to maximize profits. In doing so it overlooks the supply of managers: If the manager can get employment in other industries in which total pay (both money and many subordinates, plush carpets, fancy bathrooms, and other perquisites of the executive suite) gives the same satisfaction as does Happiest in the diagram, then the owners in the industry cannot do better than to let the manager stay at Happiest. Otherwise the manager will leave. To put it another way, the cheapest way to entice a manager to do good for owners may be to pay her partly in nonmoney coin, and it may be the cheapest way even when it is easy to detect and punish movements away from Richest.

Figure 11.2
How to Succeed in Business Without Really Trying



The Uses of the Hypothesis of Profit Maximization

Moneymaking, then, is not the only motive even in market economies. Especially in parts of the economy sheltered from the test of profit, the managers or workers can indulge tastes for a quiet life, comfortable offices, socially good behavior, socially bad behavior, leisurely schedules, and people whom they like on the job. An amusing example from the sheltered parts is the former Department of Health, Education and Welfare, which was found to discriminate against women (indulging a taste for working with men?), at any rate according to the peculiar standards the department itself applied to universities accused of such discrimination.⁸

The matter of discrimination in the labor market will come up again. For the present it is enough to acknowledge the possibility of other motives in the activities of firms and then pass firmly (so to speak) to the fact, namely, that as a good approximation firms maximize money profits alone.

T or F: The garment industry is fiercely competitive, and since prejudices are expensive to indulge, the garment industry is less likely than a sleepy monopoly to discriminate in hiring against Jews and blacks.

A: True, to the extent that the "fierce" competition forces the bulk of the firms in the industry to minimize money costs. One does not minimize costs by turning away every Jewish stockboy or black cutter who seeks

⁸ George J. Borjas, "Discrimination in HEW: Is the Doctor Sick or Are the Patients Healthy?" *Journal of Law and Economics* 21 (April 1978): 97-110.

employment on Seventh Avenue when they are skilled and inexpensive relative to the preferred group. The money motive is here the friend of the worker. And

in fact the ethnic composition of the garment industry has undergone several swift transformations over the last century.

The question has this answer, in truth, by mere assumption. The assumption is that the firm's utility is simply money alone, not a function of money and the quiet life and the pleasures of command and of prejudice and so forth. But the assumption appears to be factually powerful. People do not rush into obviously unprofitable businesses; a rise in the price of wheat does cause farmers to grow more; money invested in various different industries does earn roughly similar returns in each. The facts do not always speak clearly, but they suggest that the simple utility function $U = \$$ is promising.

Many of the disagreements economists have with other observers of society turn out to be disagreements about whether or not things other than dollars significantly affect the utility of firms. The test of whether or not they do is simple. Could the firm make more money by behaving differently? In terms of Figure 11.2, the question is whether the firm is at Happiest or at Richest. Consider the following example:

Q: Historians have long believed that British managers in the late nineteenth century valued leisure and tradition too much, neglecting innovations that would have been profitable in money. A leading example is alleged to be the neglect of cheap iron ore in Lincolnshire for making iron. The Lincolnshire ore in 1900 was half a shilling per ton cheaper than were the traditional ores. On the other hand, coke (not a cola, but a treated coal used in ironmaking) was 2 shillings more expensive per ton in Lincolnshire. If making a ton of iron required 3.3 tons of ore and 1.5 tons of coke, could British ironmasters have made more money by using the Lincolnshire ore? Do leisure and tradition appear to have been important reasons for the neglect of Lincolnshire ore?

A: The difference in the cost of iron between Lincolnshire and the traditional locations is the lower cost of ore weighted by the ore required plus the higher costs of coke weighted by the coke required: $(-0.5 \text{ shilling}) (3.5 \text{ tons of ore per ton of iron}) + (2.0 \text{ shillings}) (1.5 \text{ tons of coke per ton of iron})$ or 1.35 shilling per ton of iron in total. That is, the Lincolnshire location was more expensive, not cheaper, than the traditional locations (albeit only slightly so, for iron sold at about 40 shillings a ton). There is no need to call on any motive beyond moneymaking to explain the "neglect" of Lincolnshire in 1900: It was a worse place to make iron.⁹

What is important here is the simplicity of the method for uncovering motives other than moneymaking: Find out whether the location or technique actually chosen was best for making money. The second-guessing of decisions to see whether they were moneymaking is a common activity of economists, especially when they face historians or government officials or others doubtful of single-minded rationality. The discovery reported earlier that American slaves were held for business not prestige is one example. The repeated finding of cold-blooded profit making in the decisions of allegedly irrational peasants in poor countries is another.¹⁰ Moneymaking is more common than most people would like to believe. The economists must face the facts, of course, as they do in the case of the typical large landowner in eighteenth-century England. That

⁹ See D. N. McCloskey, *Economic Maturity and Entrepreneurial Decline: British Iron and Steel 1870–1913* (Cambridge, Mass.: Harvard University Press, 1973), especially Chapter 4.

¹⁰ See Samuel Popkin, *The Rational Peasant* (Berkeley: University of California Press, 1979), especially Chapter 1.

his investment in a large estate earned only 2% in money when equally secure investments elsewhere were earning 5% implies that the landlord earned the value of 3% in nonmonetary prestige and power from his estate. Yet even in such cases, you see, the simple logic of moneymaking proves useful. It allows one to measure in money terms the value of nonmonetary motives. Nonmonetary motives leave measurable footprints in the snow of making money.

Summary

The question of why the owner of a firm does what he does is a question of what his utility function is and what constraints he faces. The constraints are the subject of the next section. The simplest and most powerful choice for the utility function of most business firms is money profit. Only a miser or a coin collector values the coin for itself. But the coin can buy power, prestige, comfort, safety—things most people value. Managers of firms may choose to pursue power and prestige directly, at the sacrifice of maximum money profits, just as they may indulge a taste for workers of a certain color or indulge a taste for an easy life. If they do so the simplicity of the theory is spoiled, although the simplest theory still provides a standard against which to measure how much they value the nonpecuniary rewards.

EXERCISES FOR SECTION 11.2

1. Explain how moneymaking can achieve the nonmoney goals of the Illinois farmer: "Having lots of land . . . like the soil . . . to follow in his father's footsteps," and so forth.
2. What, if any, are the incentives for an administrator in the Defense Department to get an inexpensive tank for the Army? Is this a factual or a logical question?
3. In the problem in the text the cost of coke was 2 shillings more expensive than it was in the other locations. How much would the cost difference of coke have to be to make Lincolnshire a good place to make iron?

11.3 Marginal Cost: Why the Firm Produces What It Produces

What to Read For

What does the familiar *Rule of Rational Life* have to do with the firm's decision on how much to produce? What is the *marginal cost* of output? *Marginal revenue?* *Fixed cost?* What areas in the diagram represent the losses from the wrong outputs? Do "profit margins" have anything to do with the decision on how much to produce? Do "short-run profits"? How closely does the theory of marginal cost and marginal revenue need to fit the world?

A Firm Follows the Rule of Rational Life

Moneymaking, then, is the economist's first, best explanation of buying, hiring, manufacturing, transporting, stocking, advertising, and selling by the business firm. This is part of the answer to the question posed by the theory of the firm: A firm does these things because it wants to make money. The rest of the answer tells why it does them in the amounts it does, why exactly Fred

Carstensen produces 30,000 rather than 60,000 bushels. The answer is an application of the same principle applied earlier to consumers, the Rule of Rational Life. The rule is, you recall, to pursue an activity until its marginal benefit is equal to its marginal cost. The marginal benefit of Fred growing another bushel of grain is the price he gets for the bushel; the marginal cost is by definition the change in the farmer's total costs caused by growing the additional bushel. Fred produces 30,000 bushels, then, because at 30,000 the marginal cost is equal to the marginal benefit. Diagrammatically, at 30,000 bushels his marginal cost and marginal benefit curves cross (see Figure 11.3).

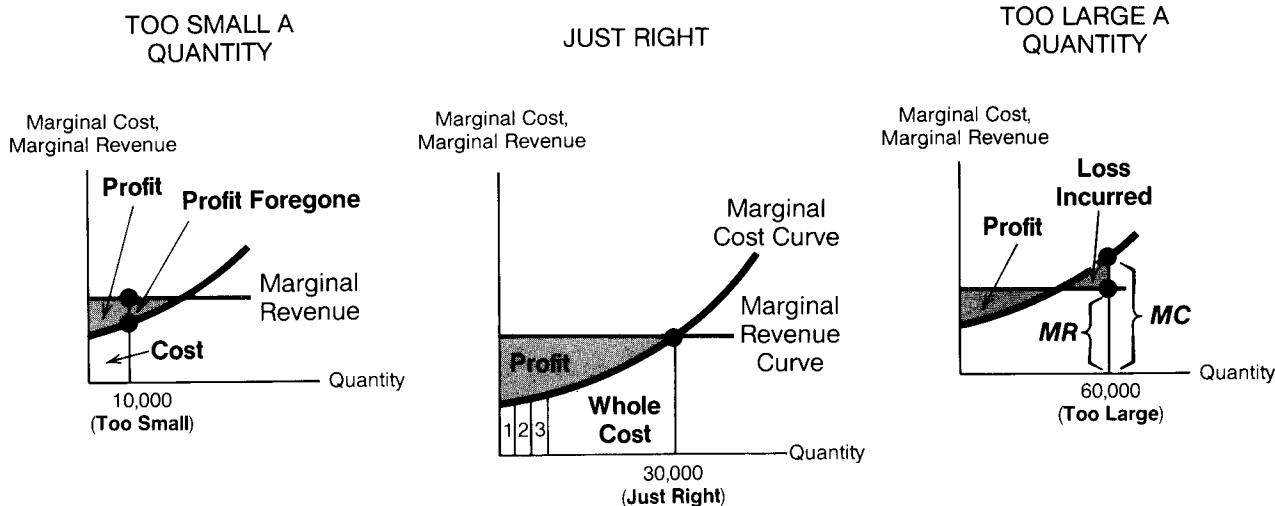
The diagrammatic reasoning is as follows. A familiar truth about marginal curves is that the cost of producing the first unit (given that the farm is in existence) is the area labeled 1 in the middle panel of Figure 11.3; the cost of the second is the area labeled 2, of the third 3, and so forth, the Whole Cost being therefore the sum of all these areas. That is, the area under the Marginal Cost Curve out to the quantity produced in the diagram is the area marked as the Whole Cost of that output (except for land rents, license fees, repayment of loans, and other *fixed* costs that do not vary with output). The rectangular area under the Marginal Revenue Curve is the whole revenue. It is price times quantity produced. Therefore, the shaded area Profit is just that (except for fixed costs): It is the whole revenue minus the Whole Cost. Producing the output Too Small leaves a positive area of Profit Foregone, as in the left panel. There is still some profit, but not the most that could be had. Producing Too Large

(1)

JL

Figure 11.3
Marginal Revenue Equals Marginal Cost and Makes the Most Profit

The firm that produces Too Small an output sacrifices potential profit because the marginal revenue of unit 10,000 is below its marginal cost. At Too Large, profit is sacrificed because marginal revenue is below marginal cost. Only at Just Right are profits maximized.



leaves a negative area of Loss Incurred (offsetting to some degree the area of Profit).

The simplest way to see that Just Right is just right is to note that anywhere else an increase or a decrease in quantity produced can make more money. At Too Small the marginal revenue from some more output is more than the marginal cost: A larger quantity will make money. At Too Large the marginal revenue is less than the marginal cost: A smaller quantity will save money, costs falling by more than revenues do. The area of profit, in other words, is at a maximum—with no profits foregone or losses incurred—at Just Right.¹¹

**Average Cost Is
Irrelevant to the
Rule**

An important point to understand here is that marginal cost (the change in cost from the production of another bushel) is not the same as average cost (all cost divided by the total number of bushels). Whatever its attractions to lawyers, regulators, and journalists, to an economist the idea of average cost is pretty much useless. Its only use, as one of several ways of expressing the desirability of going into business, will be described later. But it is no guide to how much to produce. The fallacy in using it as such a guide is the fallacy of “profit margins,” a piece of ersatz economics that is in fact antimarginal.

Q: In the year 1884 the Montsou Mining Company raised 250,000 tons of coal, sold it all at 3,000,000 francs, and paid only 2,000,000 francs for miners, steam engines, pit props, and so forth. *True or false:* With such a profit margin the owner would be foolish not to expand production still further, because he makes 33 centimes on each franc's worth of coal he brings out of the mines.

A: To be sure, the price of coal was 12 francs per ton ($3,000,000/250,000$) and the average cost of a ton only 8 francs ($2,000,000/250,000$). But this average cost over all output is irrelevant to the question of the effect of a small increment or decrement. That is, the company

does not earn 12 francs revenue minus 8 francs cost on an *additional* ton, only on all earlier tons. The cost of getting an additional ton to add to the 250,000 could well have been exactly 12 francs, in which case to try to get more would be unprofitable. That is, false: That there is profit on average is not a signal to expand. Marginal units at lower outputs evidently had cost less, but squeezing the last ton out would mean pushing against the capacity of the machinery, hiring another and less energetic person, digging deeper into a thinner seam of coal. The *marginal*, not the average cost measures the cost relevant to a comparison with marginal revenue. Rational life is a matter of margins, but not “profit margins.”

The logic of profit margins would in fact lead business managers into earning none. If any enterprise making a profit margin (or even a “healthy” profit margin) were expanded, it would arrive after its expansion at an output that would offset in Losses Incurred the profits available from stopping at Just Right, because only at such a point would *all* the profit margin and the accompanying “incentive” to expand be gone.

¹¹ The first-term calculus of all this is simple. Suppose that cost is a function, $C(q)$, of the quantity produced, and revenue another function, $R(q)$. Then the profit function is, obviously, $R(q) - C(q)$. It reaches a maximum (the top of the hill) when its derivative with respect to q (the slope of the hill) is zero, when $dR(q)/dq - dC(q)/dq = 0$, that is, when $dR(q)/dq = dC(q)/dq$. But the left-hand side of this, $dR(q)/dq$, is merely the mathematical expression for marginal revenue (the change in revenue per change in q), which is in the simple case discussed in the text simply the price; and the right-hand side, $dC(q)/dq$, is simply the marginal cost. At Just Right, then, marginal revenue equals marginal cost, and profit is maximized.

Why Firms Do Not Maximize the Average Profit

A related fallacy is that firms maximize average profit; that is, it is said that they seek the highest profit per unit of sale as against higher total profit at some larger output. Cautious firms are sometimes said to act this way. But the point of maximum average (per unit) profit is not, of course, the point of maximum total profit. The simplest way to convince yourself that it is not is to reflect that maximizing anything other than total profit will not in general achieve maximum profit. There is no good reason why a firm would maximize average instead of total profit, and at least one good reason why it would maximize the latter: Maximizing total profit maximizes total profit.

Why Firms Do Not Maximize Short-Run Profits

Another and more important fallacy is that of short- versus long-run profits. You will often hear it said that "the market (as opposed to the government or the Sierra Club) maximizes short-run rather than long-run profits." The phrase occurs often in discussions about conservation. The lumber barons of Michigan and Wisconsin in the late nineteenth century, for example, are said to have cut trees at a yearly rate too fast for long-run profits, though just right for short-run profits. The distinction between the two sorts of profits, however, is at best unhelpful. A dollar is a dollar. Aside from a discount to compensate people for waiting so long for it (discussed in Chapter 26), a dollar of long-run profits in 1899 is the same as a dollar of short-run profits in 1882. Both dollars have the same effect of raising by a dollar the owner's wealth, and it is wealth that a rational owner tries to maximize.

Another way of saying the same thing is that maximizing total profit summed over all future years will maximize the price that people will be willing to pay to buy out the firm. The price of the right of ownership in the firm is simply the summed stream of future profits, because if it were less or more, people would rush to buy up the firm or rush to sell out the firm. Unless the lumber barons did not own rights to future profits, as would be the case if the forest was not their private property, the rate of cutting would maximize long-run profit. That is, it would maximize the present value of the woods. A lumber baron who cuts so fast that the land is eroded, say, and grows no more trees, will find that the value of his forest is lowered, since no one wants to pay very much for the rights to a moonscape of eroded land. If he holds the forest for a long time he obviously has an incentive to exploit it properly. And even if he plans to sell it in six months he has the incentive, for buyers will not pay as much for an overexploited forest.

Profit Maximization Is Not Attained Exactly

The ersatz theory of the firm, then, must give way to the genuine theory, which asserts that a firm maximizes profit (or better, as we have just seen, the discounted stream of profits: wealth) and does so by choosing the output that brings marginal cost up to price. The maximization of profit or the equalization of marginal cost and price need not be perfect to 11 digits of accuracy for the theory to be useful. The ratio of the diameter to the circumference of a circle is not actually 3, or even 3.14; it is actually 3.1415926536 . . . ad infinitum. But the use of 3 or 3.14 rather than 3.1415926536 is for most uses an unimportant mistake. A map of Baltimore on a scale of 1 inch for every 2 miles is not actually a full representation of the city, for it is insufficiently detailed to name even South Baltimore, much less 1524 Hollins Street or H. L. Mencken's writing

Figure 11.4
The Reward to Better Managers Can Be the Additional Profits They Earn

A manager who orders production increased to Large when high price prevails earns area More Profit for the firm. A rational firm will pay such a manager up to, but no more than, the additional profit earned for the firm.

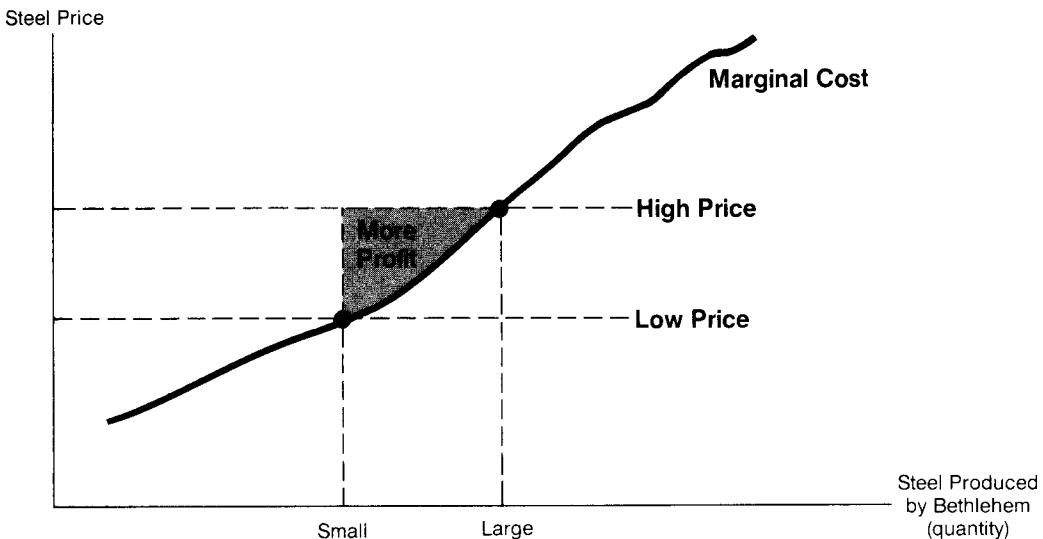


table. Any map less than the whole of Baltimore itself leaves something out and is "inaccurate" or "unrealistic." But the whole of Baltimore itself is a little hard to unfold in the car while attempting to bypass Baltimore. For this and many other purposes, the approximation works fine.

The consequence of a failure of price to equal marginal cost exactly is, of course, that profit is not at a maximum exactly. By adjusting output to yield a more exact fit of marginal cost to price the firm could make more money, just as by adjusting one's consumption of housing and other goods to every slight variation in the prices and income one faces one could be happier. But such adjustments of output or consumption are themselves costly. Indeed, the profit gained by doing a better job of profit maximizing is available to pay the manager who does the better job.

Suppose, for example, that the price facing Bethlehem Steel fluctuates unpredictably from High Price to Low Price. A stupid or cautious manager might produce always at Small output, sacrificing when the price was in fact High, the shaded area of More Profit (see Figure 11.4). That area—which could be millions of dollars for a large company like Bethlehem—is available to pay a brighter or bolder manager. If half the time when the price was High, for example, the bright manager realized it and reacted to it, then the company would be richer by half the area.

T or F: One would expect big companies in volatile industries to pay their top executives more than little companies in routine industries.

A: True, because bad decisions are more costly and good decisions more valuable for a big company in a volatile industry (for example, Bethlehem Steel in the steel in-

dustry as against Dr. Moggridge in the psychoanalysis industry). Notice that the prestige of the company or

the dignity of rank need have nothing to do with it.¹²

The Economics of Rules of Thumb

The alternatives to the usual theory of the firm were described as non-profit maximizing. The present reasoning suggests another description: The maximization of sales or profit margins or short-run profits might be merely rules of thumb that conserve the scarce resource of managerial intelligence. The rules are simpler to implement than "equalize over the long-run marginal cost to price." If they diverge too much from the strictly correct rule, they are abandoned. Tom Thumb would miss much if he used his thumb's breadth as an inch rule in carpentry and would change his rule of thumb. On the other hand, a company that aims at high sales (not profits) over the next three years (not the long run) may be acting rationally enough if high sales approximate maximum profit and if a dollar in the fourth and later years is worth little in view of the rate of interest and the uncertainties of life.

The crucial point, though, is that a firm in a market has an incentive to pull itself together if such rules of thumb are very poor approximations to maximizing profit. A firm outside a market has less incentive. The leading case in point is the behavior of firms in centrally planned economies such as that of the Soviet Union. With the best intentions but without a profit motive, a manager told to maximize, say, tonnage of paper clips or glass will tend to produce large paper clips and thick glass. It is said that managers in the Soviet Union produce 5-pound paper clips and 6-inch-thick plate glass.¹³ The closeness of the approximation to maximizing profits depends on the incentives to make it close.

Summary

Why does a farmer produce and sell 30,000 bushels of corn? Because that amount of production and sale makes the farmer as rich as possible—or so says the simplest theory of the firm. The alternative theories can be arranged in order of increasing production, from less to more than the output of a profit maximizer. Table 11.1 exhibits the order, and the sort of manager who would be associated with each output. The maximization of profit has great merit among these alternatives. It is a true description of the motives of many firms. It is defensible as an approximation forced on the firm by the market. And it is readily testable against the facts. For these reasons the economist's theory of the firm puts great emphasis on the equalization of marginal cost and price, using it as a standard of comparison even when it is false.

EXERCISES FOR SECTION 11.3

1. Suppose that Fred Carstensen, Paul Gregory, and Martin Spechler are three corn farmers in Illinois with three different schedules of total cost as follows:

¹² Compare Thomas Mayer, "The Distribution of Ability and Earnings," *Review of Economics and Statistics* 42 (May 1960): 189–195.

¹³ See Alec Nove, *The Soviet Economic System* (London: Allen & Unwin, 1977), pp. 94–96. The Soviet humor magazine *Krokodil* once published "a cartoon showing an enormous nail hanging in a large workshop. 'The month's plan fulfilled,' said the director, pointing to the nail" (p. 94).

Total Costs at Various Outputs	Farmer		
	Bushels per Year	Carstensen	Gregory
9,999	\$ 18,760.50	\$ 17,499	\$ 32,497
10,000	18,762	17,500	32,500
29,999	53,748	47,498	82,498
30,000	53,750	47,500	82,500
59,999	124,997.25	129,996.50	119,999.50
60,000	125,000	130,000	120,000

Table 11.1
What Various Theories of the Motivation of the Firm Imply About the Scale of Production

What Is Maximized	Output	Who Might Choose to Maximize in This Way
Marginal profit	Zero	A lunatic
Average profit	Low output	The French family firm in cases of high fixed costs (see Problem 2)
Wealth (total profit over the long run)	Profit-maximizing output	A selfless manager with full information or an owner-manager
Utility of manager	High output	A selfish manager who values a large staff
Output or revenue subject to some limit on losses	High output	A manager without full information following a rule of thumb; a manager in a command economy (Soviet Union, General Motors)
Output	Infinite	Another lunatic

(For the mathematically inclined, the total cost functions are $5000 + 1.25q + 0.0000125q^2$ for Carstensen; $10,000 + 0.5q + 0.000025q^2$ for Gregory; and $3.5q - 0.000025q^2$ for Spechler.) Calculate the profit (total revenue minus total cost) that Carstensen, Gregory, and Spechler make at 10,000, 30,000, and 60,000 bushels, when corn sells for \$2 a bushel. (Losses are negative profits).

2. Which of the three outputs is best for each farmer in Exercise 1? What do you suspect will happen to Spechler's profits beyond 60,000 bushels?

3. What is the dollar marginal cost—the change in cost caused by one bushel more or less—at 30,000 bushels for Carstensen? Fill in the table:

At:	Marginal Cost for:		
	Carstensen	Gregory	Spechler
10,000 bushels			
30,000 bushels			
60,000 bushels			

4. If corn sells for \$2.00 a bushel and each of the farmers takes the price as given, what is the Just Right output for Carstensen and Gregory (set Spechler aside for a moment)? Why? Compare the answer with the results of Exercise 2.
5. Does Spechler do as well as he can at 10,000 bushels? At 30,000? At 60,000? At any finite number of bushels? Explain.
6. What are the Profits Foregone by Carstensen and Gregory if they produce at 10,000 bushels? (*Hint:* Look back at Figure 11.3. What is the profit at the Just Right point for Carstensen and Gregory? What is the difference between this profit and what they get if they sit at 10,000 bushels?)
7. What are the Losses Incurred by Carstensen and Gregory from moving out to 60,000 bushels. (*Warning:* What is the distance between \$6250 and -\$5000? It isn't \$1250; nor is it \$6250.)
8. What is the change in profit for Spechler of moving from 30,000 bushels to 60,000? Compare to Exercise 2.

PROBLEMS FOR SECTION 11.3

- Show that 30,000 bushels, where price equals marginal cost, maximizes Spechler's losses, not his profits.
- French managers in the eighteenth century, it is said, were less rationally motivated than were their British counterparts. In France, and indeed elsewhere on the Continent, the owner of a little family bakery or jewelry shop had a "preference for the greatest possible profit per unit of sale, as against higher total profit at some larger output."¹⁴ True or false: If the marginal and average cost of producing bread or necklaces were rising, then by any definition of "profit per unit of sale" (average or marginal), a firm maximizing it would produce only one loaf of bread or one necklace per year.

11.4 Equimarginality: How the Firm Produces What It Produces at Minimum Cost

What to Read For

How does a firm decide how to allocate internally? How much does it take from each plant? What is the Principle of Equimarginality? Is the average cost of each plant relevant? What is the don't-do-it-yourself principle?

¹⁴ David Landes, *The Unbound Prometheus: Technological Change and Industrial Development in Western Europe from 1750 to the Present* (Cambridge: Cambridge University Press, 1969), p. 132.

The idea that profit is at a maximum when marginal benefit is equal to marginal cost is applicable, then, to the foreign policy of the firm, that is, to the policy of how much to produce and sell. It is also applicable to the domestic policy of the firm, that is, the policy of how to produce a given amount cheapest. The pursuit of minimum costs leads the firm to balance the marginal benefit against the marginal cost of reorganizing its activities.

The Two-Plant Problem

You allocate your time among activities to equalize the marginal value of an hour in each activity. So too with, say, the Brooks Shoe Company, allocating shoe output among its various separate plants. In the best allocation the company has shuffled output until nothing is to be gained by reshuffling. The analogy is the horizontal addition of demand curves for private goods such as housing. Just as the optimal allocation of housing among Higgs, DeCanio, Reid, and Wright leads to equal values of marginal benefits to each, so too the optimal (lowest-cost) allocation of tasks to different plants leads to equal marginal costs.

The simplest case is the firm with two plants, although the case is generalizable to any number of plants. How much should each plant produce to minimize the overall cost of producing a given output? The University of Maryland, for instance, can within limits assign students to its College Park campus or to its Baltimore County campus. Each student subtracted takes away costs. Clearly

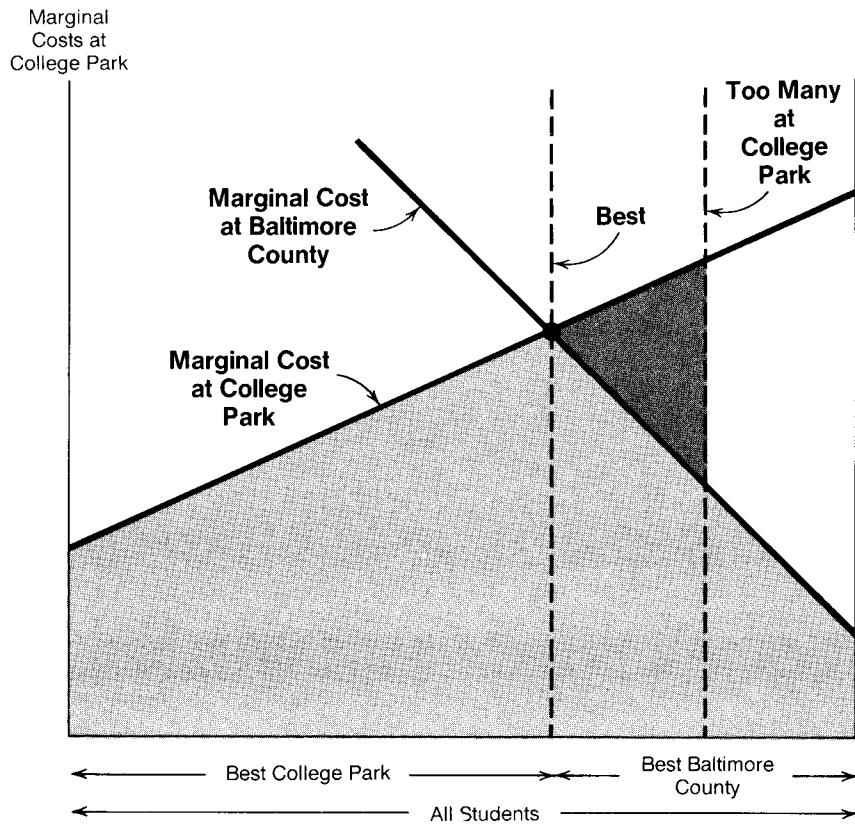


Figure 11.5
Equality of Marginal Costs
Achieves the Minimum Total Cost

A given number of students is allocated to two campuses at minimum cost if the marginal cost of a student is the same at both campuses. The lightly shaded area, equal to the difference between the cost actually incurred and the least-cost method, shows the costs of a misallocation.

the lowest cost to the state is achieved when the numbers of students on each campus are such that no reshuffling brings lower cost, that is, when moving one student from College Park to Baltimore County raises costs at Baltimore County by the same amount that it lowers costs at College Park. Dashed vertical lines in Figure 11.5 represent various alternative allocations of the total number of students to the two campuses. Notice that in the familiar style of such allocation diagrams, the enrollment at Baltimore County is measured *backward*, from right to left, and that its cost curve therefore rises backward as well.

For the allocation represented by the right line, Too Many at College Park, the marginal costs are *not* equal, which implies that costs can be made lower by reshuffling, in this case from College Park to Baltimore County. This reshuffling will result in a lowering of the Marginal Cost at College Park and a rise in the Marginal Cost at Baltimore County. The allocation Best achieves the lowest total cost, namely, the lowest heavily shaded area under the two marginal cost curves. Any other allocation will result in a lightly shaded triangle of additional cost such as that for Too Many at College Park. This is the *Principle of Equimarginality*: Set marginal costs equal in different activities.

The Irrelevance of Average Cost to the Two-Plant Problem

T or F: A given output of ammunition is produced in the best way if the output is divided up among ammunition factories so that the average cost per bullet is the same in each.

A: Only by accident would the average costs (below marginal costs if marginal costs are rising) be equal at the truly best way (that is, the way that equalizes marginal cost). What matters is whether a little reallocation can

Notice that it is marginal, not average, cost that is to be equalized.

reduce total cost, which depends on the little rises and falls in total cost caused by little reallocations. But these little rises and falls *are* marginal costs. If an extra bullet from the Enfield Arsenal costs \$1 and from the Watertown Arsenal only \$0.50, it is clear that the last bullet ought to be produced in Watertown, not Enfield, even if Enfield's cost per bullet for the sum of *previous* bullets is much less than Watertown's. Therefore, false.

The cost per bullet averaged over all previous bullets is of course much easier for a firm to measure than is marginal cost. You will sometimes find people justifying using a comparison of average rather than marginal costs for this reason. It is true that marginal cost will be measured with more error. If the true marginal cost of a bullet were \$1 when total costs were \$100,000, even a tiny percentage error in the cost estimate after the rise in output would cause a large mismeasurement of marginal cost. The firm would have to perform a larger experiment than raising output by just one bullet to get a reliable estimate of marginal cost. Or it would have to engage in an extremely subtle measurement of all the myriad ways in which rising output causes total cost to rise. By contrast, average cost is easy to measure accurately: Divide current costs (no experiments here) by current output. That it is easier to look for a lost wallet under the lamp post, however, is beside the point if the wallet was lost in the dark 50 feet away. Average cost is simply irrelevant. However crudely measured, marginal, not average cost guides allocation.

The Don't-Do-It-All-Yourself Principle

Another application of the same idea is to the question of the desirability of self-sufficiency, as when American Motors wonders whether or not it should make its own crankshafts.

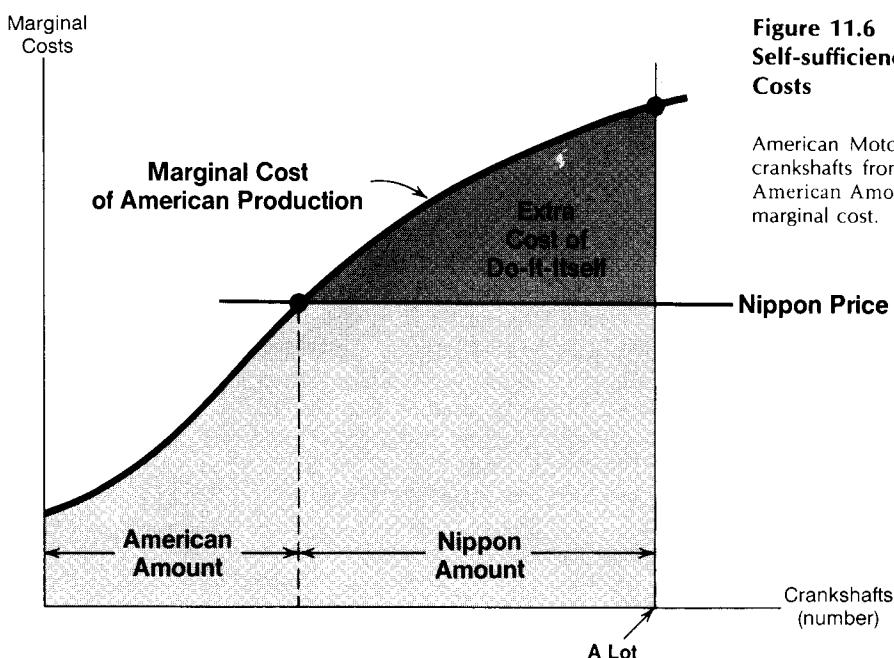


Figure 11.6
Self-sufficiency Will Not Equalize Marginal Costs

American Motors minimizes total costs by purchasing crankshafts from Nippon, because at quantities above American Amount Nippon's price is below American's marginal cost.

T or F: If it wants to maximize profits, American Motors should have a policy of making all its own crankshafts (a rising marginal cost per shaft) rather than buying some of them from Nippon Crankshaft (at a constant price).

A: "Having a policy" would mean that American Motors would buy American regardless of how cheaply Nippon sold the shafts. But to put it generally, it can never

be better to add such a constraint to one's behavior. To put it precisely, American Motors minimizes the cost of any given output by setting the amounts of its two sources of supply (itself and Nippon) so that the marginal cost of crankshafts is equal to both. And since Nippon's offer is a constant price for any amount bought, the Nippon price fixes the marginal cost (see Figure 11.6). Therefore, false.

The similarity between this allocation diagram and the last one reflects the similarity in the economic arguments. If American wanted A Lot of crankshafts, it would make the American Amount itself and buy the rest from Nippon, such that marginal costs of crankshafts (like marginal costs of students at the University of Maryland) were equal. Self-sufficiency, doing-it-itself, would cost it the extra shaded area. The lightly shaded area of total cost is minimized, given that A Lot of crankshafts are to be produced, by letting Nippon do some of it. The policy of self-sufficiency among companies, as among nations, is no way to maximize profits. Therefore, false; American should not do it all itself.

Nippon would do all of it if its price were everywhere below American's marginal cost. The Nippon price would be this low if there were *economies of scale* in the production of crankshafts (that is, marginal cost falling) that Nippon could exploit because of the large extent of the market for crankshafts.¹⁵

¹⁵ The definition of economies of scale is to some degree arbitrary. Here it is falling marginal cost. More usually it is falling average cost.

Even if American Motors also had access to the technology with economies of scale, it might be sensible for American (and Ford and General Motors and Toyota and others) to abandon its own, low-volume production of crankshafts in favor of adding its business to the already extensive market of Nippon Crankshaft. American and Nippon would be dividing up the labor of making a car, Nippon specializing in making crankshafts and American in assembling and selling them (with a few thousand other parts attached). The principle involved was enunciated by Adam Smith two centuries ago, namely, "that the division of labour is limited by the extent of the market."¹⁶ It is an answer to the question posed in earlier chapters of why firms are constituted as they are, that is, why they buy some things and make others themselves, why they choose the marketplace for one task and the command economy inside their walls for another.

Summary

The marginal cost of different activities can guide the firm to the allocation that minimizes total cost. A steel firm with four different plants and hundreds of alternative suppliers (external and internal) of coal, ore, repairs, transport, accounting, metallurgical analysis, and so forth will do best to equalize the marginal cost of each. For minimum cost the marginal cost of steel from the Gary, South Chicago, Pittsburgh, and Birmingham plants must all be equal. And the marginal opportunity cost of, say, coal from the company's own inventories on hand must be equal to the marginal cost of bringing more from the company's mines and also equal to the marginal cost of purchasing the coal from another company. At some low outside price relative to a high inside price the steel company will buy all its coal from a coal company. That is, it will divide the labor, such division being limited by the extent of the coal company's market. In this way do economic cells divide and multiply.

EXERCISE FOR SECTION 11.4

1. State the way the equimarginality principle applies to each of the following decisions, allocating:
 - a. Eight hours of study time per day between Latin and Economics.
 - b. A certain output of electricity to be gotten from two power plants at minimum cost.
 - c. A certain large sum to be borrowed by Brazil from ten different banks, each of which will charge more if too much is borrowed from it alone.
 - d. A certain number of soldiers to be deployed along a front line.

PROBLEMS FOR SECTION 11.4

- o 1. Show minimum cost for two plants on an allocation diagram with *total* cost curves. (*Hint:* Use curves with progressively steeper costs—that is, rising marginal cost; remember that each plant's total cost curve is added to the other to get *total* cost.) Prove the equivalence of this point to the point of equality of marginal costs. Where is the point of equality of average costs?

¹⁶ *The Wealth of Nations* (1776), title of Chapter III. The analysis here relies on George Stigler's essay of the same title, *Journal of Political Economy* 49 (June 1951): 185–193, reprinted in many places, for example, William Breit and Harold Hochman, eds., *Readings in Microeconomics*, 2nd ed. (New York: Holt, Rinehart and Winston, 1971).

2. Suppose that Houston Power and Light Company has two power plants, one a steam plant with rising marginal costs of a kilowatt of electricity produced and the other an atomic plant with falling marginal costs. *True or false:* It should concentrate all production in the atomic plant if it wishes to produce a given output at the lowest cost.

True or False

3. A given output of wheat will be produced at least cost if the output is divided among farms in such a way that the average cost of production is exactly the same on all farms.

CHAPTER

12

Cost Curves of the Firm

12.1 Production and Input Supply: Why Marginal Cost Is as It Is

What to Read For **What are the two fundamental messages of the idea of the production function? What are concavity and diminishing marginal returns to scale? What is the factual reason that economists assume diminishing marginal returns to scale? How do you derive a total cost curve from a production function and a price of the input? How do you derive marginal and average cost from the total cost? What is the necessary relation between marginal and average cost curves? How does a firm behave if its marginal cost curve slopes down? How can delivery costs and rising costs of inputs offset a perversely shaped marginal cost curve? What is duality? Why is it a consequence of the equality of cost and revenue?**

Output Depends on Input For some questions it is enough to know about the firm's marginal cost curve. But for others one wants to look behind the curve to its causes. Its causes are, put briefly, the constraints imposed on the firm by the state of markets and of knowledge. The firm combines costly ingredients according to the best recipes it knows to produce a given output at minimum cost. A steel company produces outputs of bars, angles, sheets, rails, structural shapes, and so forth with inputs of coal, iron ore, marketing managers, insurance, limestone, blast furnaces, soaking pit crane operators, computers, file clerks, rolling mills, and thousands of other distinct entities, whether purchased from other people or owned by the steel firm itself. The set of recipes for combining the inputs is called the *production function*—an idea that was used earlier and will be used later still again. A firm might have dozens of outputs and inputs. For present purposes, however, one output and one input will do. This means adding up steel sheets and cold rolled bars into one output, in amount Q , and adding up soaking pit crane

operators and rolling mills into one input, in amount I . That is, some function F connects output to input; thus, $Q = F(I)$. A function, for example, might be a constant elasticity one, such as $Q = 5I^{0.8}$. The simple one-input production function conveys the idea that output depends on input. (A many-input production function conveys the additional message that there is more than one way to skin a cat; that is, there are different combinations of inputs that can be used to produce the same level of output.)

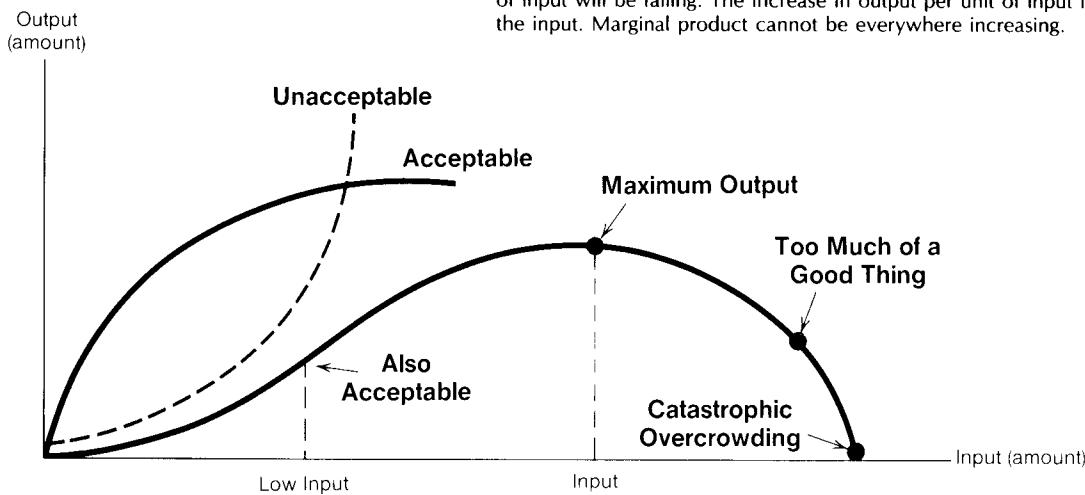
Diminishing Returns Is, for One Thing, a Fact Directly Observed

A feature this particular function shares with other functions that might describe a firm's recipes is *concavity*. It is the shape marked Acceptable in Figure 12.1. What is acceptable about it is that it exhibits *diminishing marginal returns to scale*, which is to say that, as the firm attempts to produce more output by pushing in more input, each additional dose of input produces less additional output. The slope diminishes.

Unlike the man in the street, economists do not believe that bigger is always better. Early in the application of fertilizer, labor, machinery, and so forth to a given plot of land (or all these and land to a given farmer), there may well be "economies of scale" in producing corn. But eventually the firm is overwhelmed by the additional inputs. The 50 tractors crowded onto the 2 acres of Mr. Craft's farm smash into each other and explode; the fertilizer left in tons on each square yard buries the young corn plants to a depth of 6 feet; the thousands of laborers become a mob and take to sleeping, fighting, and card playing on company time, then turn to Craft's house and burn it down for sport. The point of Catastrophic Overcrowding in Figure 12.1, needless to say, will never be reached by a rational firm. In fact, it will never reach beyond

Figure 12.1
Eventually Diminishing Marginal Returns to Scale Are True and Necessary

An empirical law of production is that at high levels of input the marginal product of input will be falling. The increase in output per unit of input is lower the greater the input. Marginal product cannot be everywhere increasing.



Maximum Output. Less colorful stories apply to the earlier parts of the curve, but the moral is the same. Along Acceptable, marginal returns diminish continually; along Also Acceptable, they at first increase, then diminish. The critical point is that the functions do not end like Unacceptable, the upward-curving dashed line. The reasons are two, the first being fact: When they can be measured well—as in the case of farming—production functions of firms do not exhibit continuously increasing returns to scale.

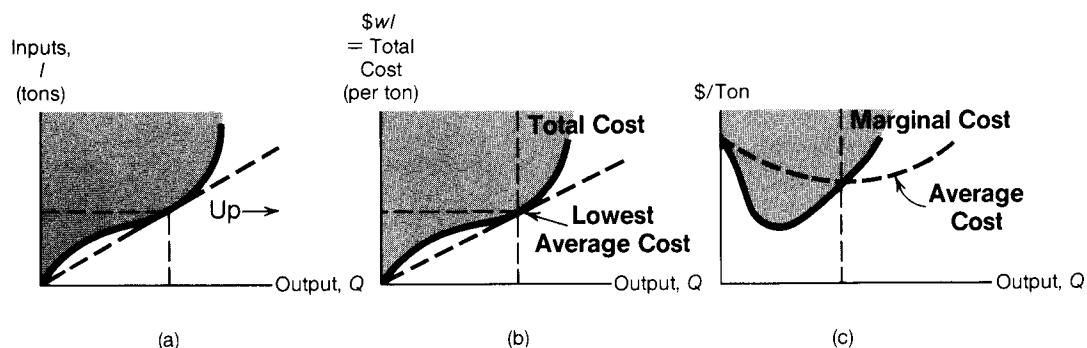
How to Reason from the Shape of the Production Function to the Shape of the Cost Functions

The second reason the slope declines will become clear as the argument goes forward. Suppose that the all-purpose input has a fixed per-unit price of w (for wage, the price of the input that is most often the subject of thought). The price is one of the constraints facing the firm. Given w , it is easy to get from the production function to the cost curve. Look at Figure 12.2. The left panel has the production function laid on its side, with a shape like Also Acceptable in Figure 12.1. Look at it with the arrow pointing Up. You will see that it merely reverses the direction in which input is measured. If w is fixed, the total cost is of course simply wI (that is, the input price times the input amount). The vertical axis can be stretched by the factor w . The stretching clearly will leave the general shape of the curve unaltered. The result is the Total Cost curve in the second panel of the figure. The third (right) and final panel simply plots the *slope* of the Total Cost curve against quantity. The slope is the rise in total cost per unit of a rise in quantity; that is, the slope is Marginal Cost.

The diagram contains or implies everything you will ever need to know about cost curves. So read the previous paragraph again, slowly: It's worth the investment. Notice the way in which the little tangencies along the Total Cost curve follow as they should the falling and rising shape of the Marginal Cost curve. Notice, too, that the Average Cost (the dashed curve) is the slope of a ray

Figure 12.2
The Shape of the Production Function (a) Determines the Shape of Total (b) and Marginal (Average) (c) Cost Curves

If inputs can be bought at a constant price, the total cost curve and the production function are identical except for scale. The slope of the total cost curve at a given quantity is of course marginal cost at that quantity. The slope of a line through the origin intersecting the Total Cost curve is average cost at that quantity.



from the origin to the total cost curve. Such a slope measures total cost divided by output, and therefore when the ray reaches a minimum in slope at Lowest Average Cost on the Total Cost curve (middle panel) it is also identical to the *slope* of the curve. That is, marginal cost is equal to average cost once, slicing up through the average cost at its minimum point. This amusing and surprising fact is useful for drawing self-consistent diagrams of average and marginal cost.

A Numerical Example It will help in driving home the diagram to work through a numerical example. Try to do it without peeking at the answer.

Q: Marcello de Cecco hires labor to harvest his olive trees and press out the oil at a fixed wage of \$4 per hour. His production of olive oil as a function of hours of labor is:

Labor Input (hours)	Olive Oil Output (gallons)
100	30
200	100
300	200
400	350
500	800
600	1200
700	1400
800	1575
900	1700
1000	1800
1100	1850

1. List the total costs as a function of output from 30 to 1900 gallons. Sketch its shape.
2. Calculate the marginal cost at each output. (To be definite, use, so to speak, a forward-looking definition of marginal cost: At 1200 gallons ask what additional costs are incurred *for each gallon on average from 1200 to 1400*.) Sketch it.
3. Calculate the average cost at each output. What, approximately, is the output at which average cost is at a minimum? Is marginal cost approximately equal to average cost there?

A: 1. Simply multiply the hours of labor input by the \$4 cost per hour and list the result next to the corresponding output level (ignore the last two columns):

Output (gallons)	Cost	Marginal Cost per Gallon	Average Cost per Gallon
30	\$ 400	\$5.71	\$13.33
100	800	4.00	8.00
200	1200	2.67	6.00
350	1600	.89	4.57
800	2000	1.00	2.50
1200	2400	2.00	2.00
1400	2800	2.28	2.00
1575	3200	3.20	2.03
1700	3600	4.00	2.12
1800	4000	8.00	2.22
1850	4400		

The sketch is just like Figure 12.2, middle panel, appropriately labeled.

2. At 30 gallons an additional 70 gallons costs an additional \$400. So marginal cost is on average $\$400/70 = \5.71 per gallon. At 100 gallons, an additional 100 gallons costs an additional \$400. So marginal cost is on average $\$400/100 = \4.00 . Going through the table in like fashion gives the third column above, the marginal cost. The sketch is the right panel of Figure 12.2.

3. Average cost is just total cost divided by output. For instance, at 1400 gallons of oil it is $\$2800/1400 = \2 per gallon. The last column fills in the rest. The minimum average cost is somewhere around 1200 or 1400 gallons. Marginal cost is approximately equal to average cost there. In other words, just as Figure 12.2 claims, the marginal cost curve, starting below the average cost ($\$5.71$ is less than $\$13.33$, $\$4.00$ is less than $\$8.00$), is equal to the average cost when the average cost hits bottom. Thereafter, the marginal cost is higher ($\$3.20$ is higher than $\$2.03$, and so forth).

**Other Things Equal,
Diminishing
Returns Imply
Rising Marginal
Cost**

The crucial result is the mere upward slope of the marginal cost curve. The upward slope comes directly from the diminishing returns to scale. Put verbally, the lower increments to output forthcoming from additional units of input mean that given increments to output will require larger and larger doses of inputs. That is, they will require higher and higher costs. With marginal cost rising, the firm will arrive at a particular finite output if it applies the rule (of rational life) "to maximize profit, set output such that marginal cost equals marginal revenue."

T or F: If on the contrary the production function implied *higher* increments to output from additional units of input, the firm will expand without limit.

A: With increasing returns to scale (and a fixed factor price, w) the marginal cost curve slopes *downward*. A competitive firm would be minimizing, not maximiz-

ing, profit if it stopped at the output that equalizes marginal revenue and marginal cost (see Figure 12.3). The firm makes a positive marginal profit on each quantity sold beyond the point $MC = MR$ and will therefore continue moving to the right indefinitely. Therefore, it would expand indefinitely. Therefore, true.

Again, consider the following.

T or F: Although increasing returns to the scale of a perfectly competitive firm is not consistent with equilibrium, constant returns is.

A: By the reasoning just given, constant returns implies a flat marginal cost curve—by contrast with the rising

one of decreasing returns and the falling one of increasing returns. If a flat demand curve—a price given to a price-taking firm—is the marginal benefit, then there are three possibilities, none of which is attractive (see Figure 12.4). Therefore, false.

The usual case is that marginal cost eventually rises:

Q: If de Cecco in the earlier problem can sell his olive oil at \$2.30 a gallon, where does he produce? How much profit does he make there?

A: He produces where marginal benefit (\$2.30 a gallon) equals marginal cost: that is, at an output of 1400 gallons, approximately, where marginal cost is \$2.28. His

**Figure 12.3
Why Marginal Cost Must Slope Upward**

A firm facing a horizontal demand curve cannot have a marginal cost curve that is falling and that lies below the demand curve. The optimal size of the firm would be infinitely large.

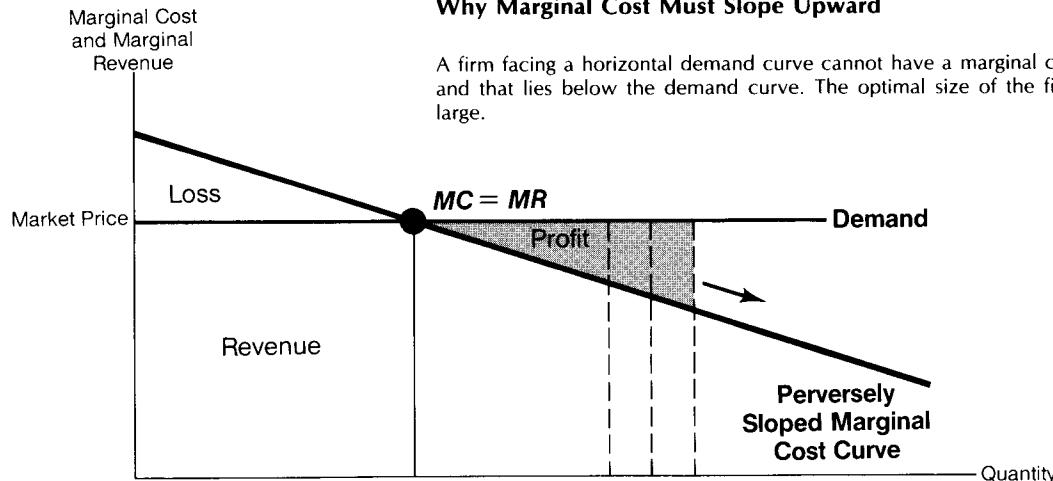
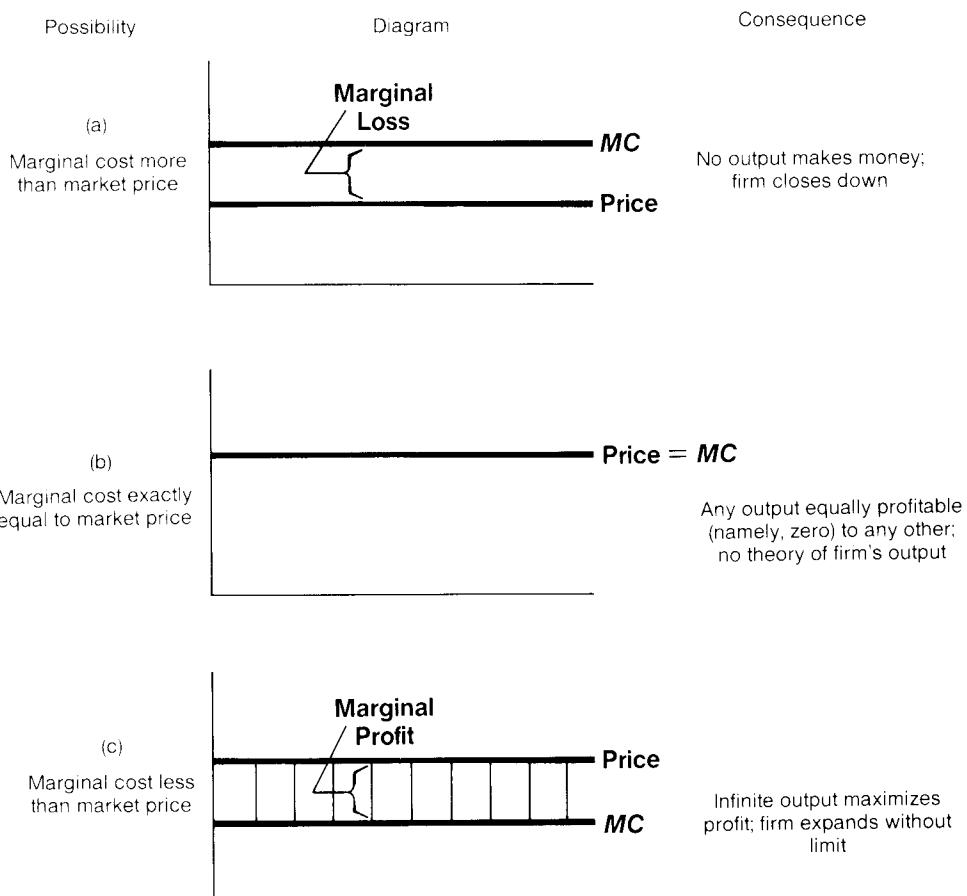


Figure 12.4
Constant Returns Is Inconsistent with an Equilibrium Scale for the Firm

A firm facing a horizontal demand curve and having a horizontal marginal cost curve would produce either nothing or an infinite quantity or would be indifferent to the amount it produced.



profit is his revenue on 1400 gallons—namely, $(1400) \times (\$2.30) = \3220 —minus his total cost, which at

1400 gallons is \$2800. So he makes $\$3220 - \$2800 = \$420$.

Transport Costs Might Not Be Equal

Such considerations demonstrate that a flat marginal cost is inconsistent with a *price-taking* firm—a firm that is so small in its market that it faces a flat demand curve. But a firm that is big in its market, and that faces therefore a downward-sloping demand curve (because it faces such a big part of the entire demand curve), could rationally stop at a finite size though its marginal cost of production were flat. The reason, to be elaborated in Chapter 17 on monopoly, is that selling too much could carry the firm so far down the demand curve that the fall in price would overbalance the rise in quantity. A monopolist, in

other words, faces diminishing returns *in selling* that can play the same role of limiting the scale of the firm as do diminishing returns *in production* or (as we shall see in a moment) in buying inputs.

Another case of a single firm facing a downward-sloping demand curve is that of a firm that can get more customers only from farther afield, with higher costs therefore of delivering the product. Costs may be constant or even declining as output rises *at the plant*, but the rising cost of *delivery* from the plant leads to a finite equilibrium. Big hardware stores, for example, are better, since they can stock a wider range of hardware and assure that each visit by a customer is more certain to yield the kind of paint, nail, or tool the customer wants. Should Chicago therefore have one big store? No. It has in fact over 400, because a local store is convenient; the trip to the One Big Store would be expensive.

Q: A careful study of economies of scale in the generation of electricity by privately owned American utilities in 1955 concluded that they were very strong: In the constant-elasticity form $Q = I^\alpha$, the α was 2 or higher for the firm (any α greater than 1, recall, is an unacceptable elasticity in Figure 12.1).¹ Yet there were over 400 privately owned electrical utilities in the United States in 1955. Why?

A: Transporting electricity is expensive, and the farther it is transported the more expensive it is. Therefore a flat or downward-sloping marginal cost at the plant is consistent with a rising marginal cost at the place of use. Other *natural monopolies* limited by transport costs are said to include grocery stores, schools, drugstores, banks, and the like.

Diminishing Returns Can Limit the Size of the Firm

Unless it is offset by transport costs, then, a flat or downward-sloping marginal cost curve for a competitive firm is a great theoretical nuisance. For many firms the costs of transport are insignificant. Clothing manufacturers on Seventh Avenue or automobile dealers in Los Angeles or wheat farmers in Kansas are not isolated from each other's competition by economically significant distances, yet they are finite in size. To have a theory of such finite firms actually observed in the world, one is forced to abandon the common notion that firms always reap endless economies of scale or even the less common notion that costs are constant.² The argument is one of survival, in the style of biology since Darwin. A firm that is "fit" in the size it has chosen will survive by being profitable. Unfit species will vanish. If sizes of firms in the retail women's clothing trade cluster around one size, the economist seeks reasons for the apparent optimality of the size. She uses the results of selection to guide what would otherwise be an impossibly difficult inquiry into the cost curves of firms.³ If there were no presumption that the sizes and other characteristics of firms that survive are in fact the least-cost characteristics, economic studies would be crippled. For

¹ The facts, the question, and the answer are taken from Marc Nerlove, "Returns to Scale in Electricity Supply," in C. F. Christ et al., eds., *Measurement in Economics* (Stanford, Calif.: Stanford University Press, 1963), reprinted in Arnold Zellner, ed., *Readings in Economic Statistics and Econometrics* (Boston: Little, Brown, 1968).

² The various attempts to actually observe the cost curves of firms, however, show with embarrassing frequency that cost curves do appear to be flat. The observations can often be rationalized as mismeasurements, compare Milton Friedman, in National Bureau of Economic Research, *Business Concentration and Price Policy* (Princeton, N.J.: Princeton University Press, 1955), pp. 230-238. The best case for constancy of marginal cost and against the conventional theory of the firm as presented here and in other texts is J. Johnston, *Statistical Cost Analysis* (New York: McGraw-Hill, 1960), especially Chapters 5 and 6. His view is that many firms are monopolists.

³ The idea was made explicit by George Stigler, "The Economics of Scale," *Journal of Law and Economics* 1 (October 1958): 54-71.

the same reason, so would ecological studies that could make no presumption that the cowardice of wolves in the hunt or the falling of leaves in autumn are in fact valuable for the survival of wolves or of broadleaf plants.⁴

Chapter 14 will exploit the argument from survival to its limit. For the present the argument serves merely to buttress the belief that marginal cost curves rise. It is the second reason for the upward curvature: Not only is the upward curvature in fact observed when observation is possible, but it would have to be observed in a rational and finite-sized firm.

**An Inelastic Supply
of Inputs (Like a
Downward-Sloping
Demand) Can Also
Limit the Size of the
Firm**

It is the shape of the cost curve, not the underlying production function, that matters in the end for the firm. Consider the following:

T or F: The production function of Bethlehem Steel might exhibit constant or increasing returns to scale, yet the resulting cost curve could have the normal shape if the supply curve of iron ore facing the firm were inelastic instead of perfectly elastic.

the ore (if, as assumed, the ore is supplied inelastically to the firm). The rise in the price of the inputs can offset the advantage gained from increasing returns in needing less and less inputs per unit of output. Therefore, true.

A: The reasoning is simply that as Bethlehem produces more, it buys more iron ore, which raises the price of

**The Dual and the
Primal Problems**

In short, the production function and the cost function are linked. The link is called *duality*, a thought so powerful in economics that in the higher reaches of abstraction in the field it dominates much thinking. The word comes from the jargon of *programming*, that is, the branch of applied mathematics that deals with maximizing something under many constraints. In the present context, the thought is that the production function is the answer to one question—what is the most quantity, Q , the firm can attain for a given input, I ? And the cost function is the answer to a related question—what is the least cost, wI , the firm can attain for a given Q ? And that the two questions are two sides of one question—what is the best thing the firm can do? They are *dual* to each other; that is, each is one of a pair. The two solutions are solutions to the same problem, but one solution is to maximize and the other is to minimize. In doing well an executive for General Motors can either start with a work force, a plant, and some raw materials and produce as many cars as possible or, equivalently, start with a certain number of cars and produce them with the given work force, plant, and raw materials as cheaply as possible. The two are not really two solutions, but the same, single solution looked at two ways. In other words, maximizing output for a given input is the same as minimizing input (and therefore input costs) for a given output; that is, Q/I , or the physical productivity of the inputs, is maximized when $w(I/Q)$, or the input cost per unit of output, is minimized. Elementary though it is, it is easy to get confused

⁴ The examples comes from Paul Colinvaux, *Why Big Fierce Animals Are Rare: An Ecologist's Perspective* (Princeton, N.J.: Princeton University Press, 1978), pp. 58, 153.

about the matter, and confusion about it is a reliable indicator of economic ignorance.

T or F: In a nonexpansive market, managers will attempt to minimize costs rather than maximize output.

A: Taken literally, "minimizing costs" would occur at an output of zero, and "maximizing output" would occur at an output of infinity. Read more sympathetically, the assertion must refer to minimizing (input

costs for a given output and maximizing output for a given input. But these are merely alternative ways of expressing the same thing. Staying on the cost curve (minimizing costs) is identical to staying on the production function (maximizing output). Therefore, false or meaningless.

Similarly, consider the following.

T or F: The link between new farming methods and productivity is weak. The new methods *may* show up in higher output per unit of input (which is higher productivity) but alternatively may be used merely to pay higher wages, with no increase in output.

A: The second sentence is true, but it is false evidence for the first. The confusion is again between the amount produced and the efficiency with which it is produced.

New, higher productivity methods will raise output per unit of input (which is what higher productivity means). They will therefore reduce input per unit of output. They will therefore permit higher pay for each unit of input, for less input is wanted per unit of sellable output. The argument illustrates the strength of the link between new methods and productivity, not its weakness.

The link is so very strong because it is so very obvious. A firm must earn as much as it pays. That is, the price of its product times the quantity produced must equal the price per unit of inputs times the quantities of them used. The assertion that $PQ = wI$ implies, by the magic of elementary algebra, that $Q/I = w/p$. The new equation says that output per physical unit of input equals the price of the input divided by (or "deflated by") the price of the output. In other words, there are two ways of measuring productivity, one with quantities and another—the price "dual" of the quantity "primal" (*primal* means "the original problem")—with prices. The two are equal. The productivity of inputs is equal to their real price (that is, their money price deflated by the price of output). The higher real wages, rents, profits, and so forth that Americans get compared with Ethiopians are not something apart from the higher physical productivity of the American economy but identical to it, because $Q/I = w/p$. The failure of incomes (which are input prices) to "keep up" with inflation (which is the output price) is not a consequence of an economic footrace but, rather, the necessary consequence of falling productivity unrelated to inflation, because the rate of change of w/p must equal the rate of change of Q/I . A similarity between a measure of productivity change based on quantities and one based on prices is not a happy accident but, rather, a necessary correspondence, because again $Q/I = w/p$.

Summary Behind the marginal cost curve is the production function. Either diminishing returns to scale in the function or rising delivery costs or sufficiently inelastic supply curves of inputs can cause the marginal cost curve to slope upward. If it slopes downward, the firm will expand without limit, which suggests that it must slope upward. Somewhere in the constraints facing the firm, then, more gets you less.

The relationship between costs and production is one of *duality*. As the production function is the most output for a given input, so the cost function (the "dual" of the production function) is the least cost for a given output. The link between the two is the budget equation, $PQ = wI$, which is to say that revenues equal expenditures. The equation implies that $Q/I = w/p$, which is to say that physical productivity equals the real reward of the input.

EXERCISES FOR SECTION 12.1

1. Which if any of the following production functions have acceptable shapes?

a. When Labor Is:	Quantity Is:
100	200
200	500
1,000	3,000
10,000	40,000

- b. $Q = L^2$
 c. "As labor input rises by 1 percent, output always rises by more than 1 percent."
 2. Suppose that de Cecco's olive grove produced output this way:

Labor Input (hours)	Olive Oil Output (gallons)
100	30
200	120
300	250
400	370
500	500
600	600
700	675
800	700
900	710
1000	715

He still pays \$4 an hour for labor. Calculate and list against the various outputs from 30 to 715 gallons the total cost, marginal cost, and average cost.

3. In Exercise 2, what approximately is the output with minimum average cost? If de Cecco sells oil at \$16 a gallon, where does he produce?
 4. Contrast the cost curves of Exercise 2 with those in the text. Which represents a small-scale firm? Which has more sharply diminishing returns to scale?
 5. What would de Cecco in the text produce if the price of olive oil were \$1 a gallon? \$3.20 a gallon? \$8? What curve, then, is de Cecco's *supply curve*, that is, the curve showing how much output he supplies at various different prices?

PROBLEMS FOR SECTION 12.1

- 1. The McGouldrick Textile Mill of Manchester, New Hampshire, is planned as a cube of height (and width and depth), h . Suppose that the construction and maintenance costs of the mill itself are proportional to the surface area of the mill, because the cost is determined by the extent of brick and glass ("Factory windows are always broken / Somebody's always throwing bricks."). Suppose too that the output, Q , of the mill is proportional to its cubic volume, because the larger the volume the more spindles, looms, and other producers of output can be crammed in. *True or false:* Total construction and maintenance costs will vary as $cQ^{2/3}$, where c is some constant; that is, marginal construction and maintenance costs will decline as planned output rises, two thirds being less than one. (*Hint:* Use the output equation to solve for h as a function Q ; then substitute into the cost equation.)
- 2. A lone dealer in the London *Times* newspaper living at Isca at the end of the Fosse Way can get as many copies as he wants at 10 pence a copy (that is, his marginal cost at Isca is constant). All his customers will pay only 15 pence a copy. At each mile marker north of Isca (the Fosse Way runs north from Isca to Lindum, you see, and is the only road), there are 1000 potential buyers of the *Times*, beginning at mile 1. Each mile of transport cost a quarter of a penny per copy. How many copies will he sell in this case? Draw a diagram illustrating the general case, with distance in miles along the horizontal axis (distance being equivalent to numbers of copies) and cost and selling price along the vertical.

True or False

3. If the production function for the airframe industry is characterized by constant returns to scale, the supply curve of the industry will be infinitely elastic in the long run.
4. If total costs = cQ^α and α is less than 1.0, marginal cost is declining (the firm is experiencing increasing returns to scale).

12.2 Cost Curves in Use: The Long Run and the Short Run

What to Read For

What is the principle that adding constraints hurts? How does the short-run cost curve of a firm illustrate the principle? What are the two reasons a firm will tolerate being on a higher-cost, short-run curve? Would it tolerate it if there was a secondhand market in equipment? Why might a firm prefer an expensive but flexible plan to a cheap but inflexible one? What is the envelope of all short-run cost curves?

The Production Function or Cost Curve Is the Best

That the production and cost curves measure the most output and the least cost implies that the firm can do worse. Indeed it can. It can take up positions anywhere in the upper area of Figure 12.5, such as the point Error, using excess inputs in amount ΔI to attain Q_0 , which will therefore cost more than it should, by the amount ΔI multiplied by the cost of each unit of input.

A firm having full knowledge of its costs and managed by a genius would never be in error. Knowledge and genius, however, are expensive, more expensive for some firms than for others, with the result that actual firms depart more

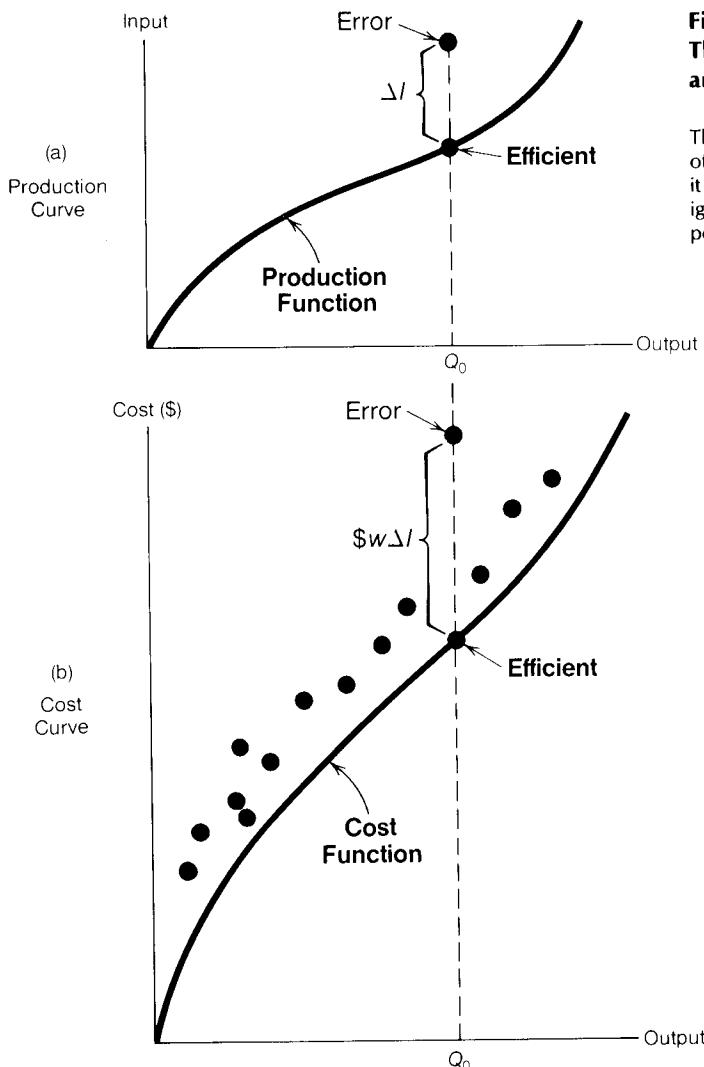


Figure 12.5
The Firm Can Do No Better than the Production (a) and Cost Curves (b), but It Can Do Worse

The production and cost functions show the least costly methods of production. If the profit-maximizing firm had full information, it would always choose points such as Efficient. In practice, ignorance, which is costly to remove, makes the firm choose points like Error.

or less from the most output or the least cost. The departures from the total cost curve, shown as dots in the bottom panel, can be viewed either as mere errors or, more fruitfully, as the larger costs imposed by smaller amounts of inputs not explicitly measured. In either case the deviations make it difficult to find out what "the" cost curve is.

A Constrained Firm Does Worse than the Best

The key point is that real firms spend much of their time above "the" cost curve. The point can be put in the form of another one of those stunningly obvious but frequently misunderstood principles of economics: namely, *adding constraints hurts*. The principle was first stated in the chapter on utility functions. A firm that is constrained to use dull managers because of the cost of finding better ones will have higher costs than one free to use the best. Likewise

a runner constrained to leap over hurdles while running 100 yards will have worse times than one free to run straight. The women's 100-meter freestyle record in swimming is never slower than the 100-meter backstroke, since freestyle means that the swimmer could choose the backstroke if it were the fastest way of covering 100 meters.

T or F: The costs of operating the World Trade Center building in New York will be higher with a new federal rule that air conditioning must be no cooler than 78°F than without it.

A: If constrained to shift from 72°F to 78°F, the building might save on energy costs (as a matter of fact, because existing equipment is designed to operate best at the old, low temperature, it might not) but will lose on

other costs, such as the efficiency of the office workers who are using the building or the life expectancy of stored paper. The conclusion would be inevitable if the building were being operated at minimum cost before the federal rule was imposed, because if the higher 78°F temperature were the optimum, it would have been chosen anyway, in the absence of the federal rule. Therefore, true.

Constraints Put the Firm on the Short-Run Cost Curve

The most important class of examples of how firms can do worse and how constraints make them do so is that short-run costs are always above long-run costs. By *short run*, the economist always means "constrained to use the equipment in place, that is, unwilling to adjust fully and immediately to a new scale of operation." Since the short run is a situation with more constraints than is the long run, it is clearly more costly than the long run. Constraints always hurt.

There are two reasons why a firm might suffer the constraints of the short run. The first is that, if a change in output is known to be short lived, the firm will not want to buy a long-lived piece of equipment to service it. For example, suppose (as was the case) that the output of skateboards was very large during the time that children were just discovering them, but fell to the replacement output as soon as every house in the land had a couple of skateboards in it. An intelligent firm wishing to cash in on the temporary boom would not buy specialized factories, retail outlets, and so forth suited to a long persistence of the boom output, because it knows the boom will not last. The specialized factories will last for 20 years. If the boom lasts as long, the per-skateboard cost of the factories will be low. But if the boom lasts only two years, the cost per skateboard will be high. At such costs it may well be better to rent a factory than to buy, to get a factory suited to general use rather than one specialized in the production of skateboards, and to pay premiums to old workers to work overtime rather than to hire and train new ones. The first principle, then, is that one does not use a cannon to kill a fly.

The second reason why a firm might suffer the short run is that, even if the change in output is known to be permanent, it may be cheaper to adjust to it slowly. For example, if the college of the University of Chicago decides to expand from 2500 to 3500 students, it will eventually want to have more housing, more classrooms, and more teachers to accommodate the increased number of students. But only "eventually." It will not immediately knock down the old buildings and rebuild them to better suit the higher scale; it will not immediately hire new teachers. The costs of adjustment are high. Haste makes waste. It will, instead, house the students in rented hotels and hire visiting teachers on temporary assignment. These are expensive and inconvenient devices but cheaper than excess haste in adjusting. Only over the long run will it take

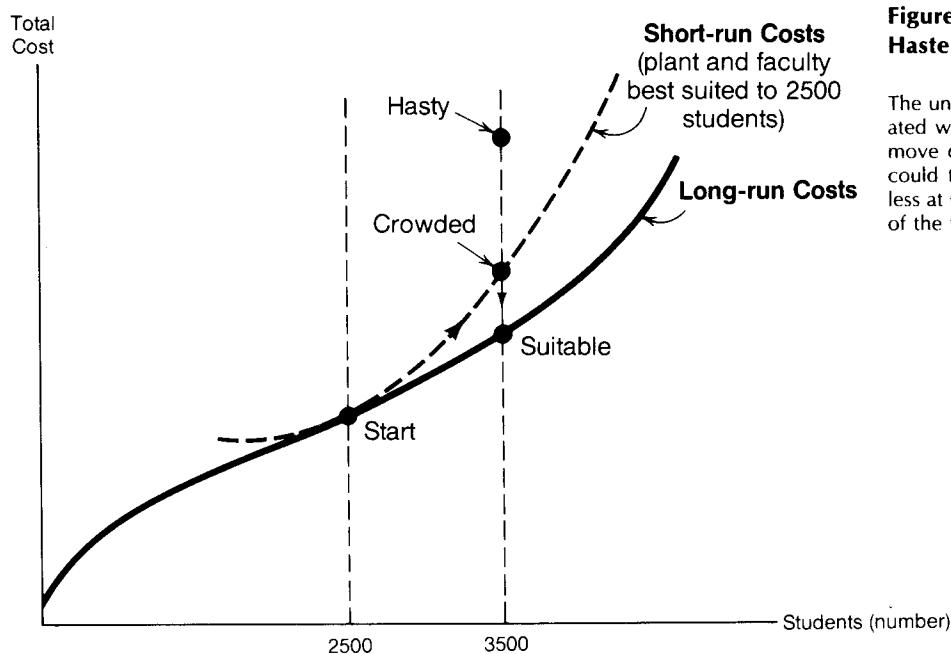


Figure 12.6
Haste Makes Waste

The university could pay the costs associated with point Hasty for a short time to move quickly to the point Suitable, or it could take longer to adjust while paying less at Crowded. It chooses the less costly of the two alternatives.

advantage of the wearing out of buildings and the wider pool of good teachers available from a leisurely and thorough search to remake the plant and faculty to suit the higher scale.

The long-run costs will be lower than the short-run costs, which are in turn lower than the Hasty costs. See Figure 12.6. A rational university wanting to increase its size will move from point Start to Crowded to Suitable. The dashed short-run cost curve is exactly suitable only to the output of Start. Everywhere else—at student numbers below as well as above 2500—it is higher than long-run cost. The second principle then is that haste makes waste.

The two principles are widely applicable. A sudden rise in the number of students in grade school should not inspire an immediate remaking of whole schools, even if the rise itself is permanent and especially if it is not. A rise in the demand for automobiles should inspire Ford to hire people on overtime at old plants before constructing new plants. And it applies to falls in output as well as to rises. A fall in attendance at Boston Red Sox baseball games should not inspire the management to tear down Fenway Park and build a new, smaller stadium at once. Firms spend some time on their short-run cost curves.

Why Firms Choose to Spend Time off the Long-Run Curve

Both principles would be false, however, if there was a frictionless secondhand market in university buildings, grade schools, auto plants, and baseball stadiums. Unless getting into and out of owning such things were expensive, the distinction between long- and short-run costs would be empty, because the firm would never be stuck with a short-run expedient.

T or F: A shirtmaker who owns easily sellable space on Seventh Avenue, rents his sewing machines, and buys labor by the week is never off his long-run cost curve.

A: He can resell all these things to the market at a moment's notice, repurchasing them at the next, with no

significant cost of making and unmaking the deals. He will therefore never have to suffer the inconvenience of an unsuitable set of equipment. That is, true.

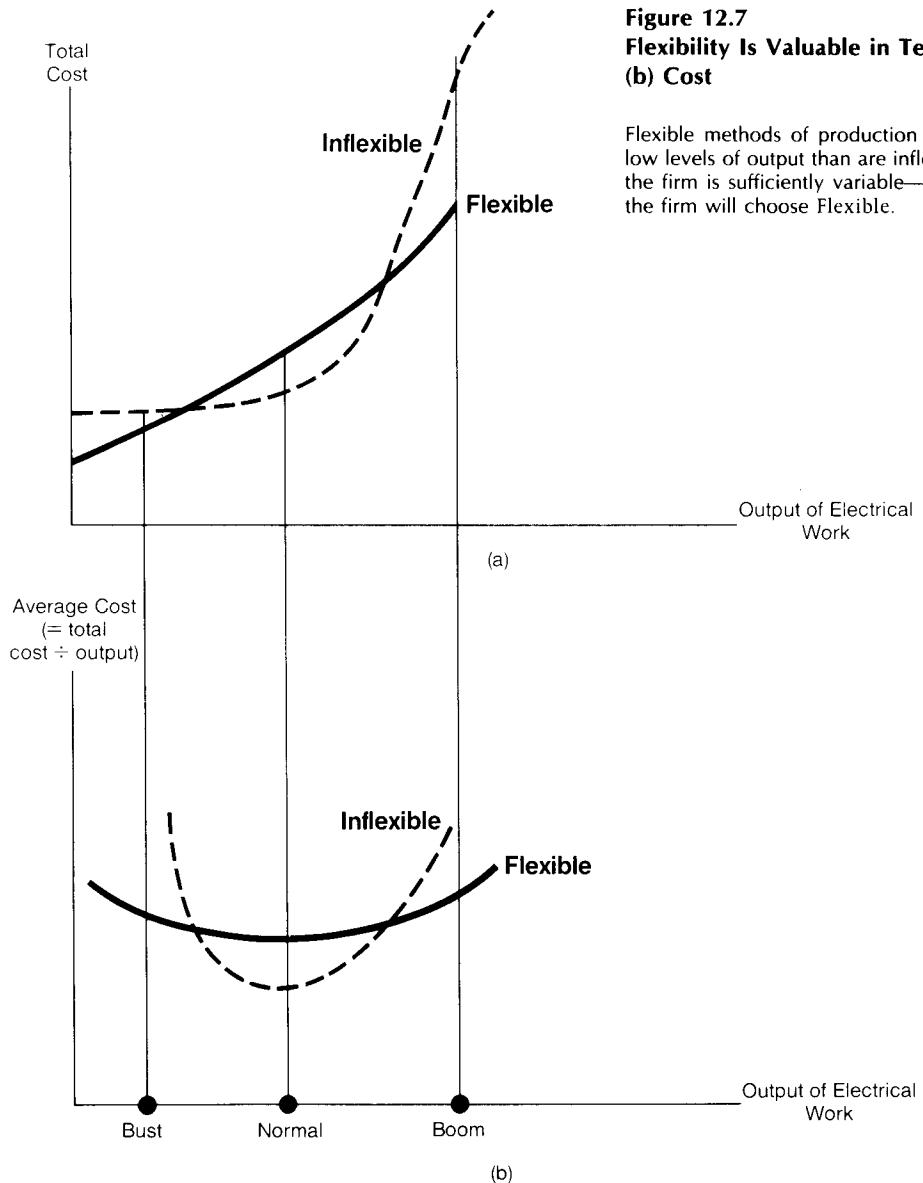


Figure 12.7
Flexibility Is Valuable in Terms of Total (a) and Average (b) Cost

Flexible methods of production are less costly at very high or very low levels of output than are inflexible methods. If the demand facing the firm is sufficiently variable—if there is often a boom or a bust—the firm will choose Flexible.

The distinction between the long- and short-run cost curve, then, is a crude way of allowing for the costs imposed by inconstancy of output, by ignorance of its duration, by haste in adjusting. In a word, the distinction allows for the cost of transacting. The costs of transacting are not given to the firm by God, for a firm has some choice of the set of transaction costs it will face. For example, it can choose either to rent or to buy equipment, such as sewing machines for shirtmaking. Renting has different sizes and types of transaction costs from ownership, and the balance of advantage of one over the other is not obvious. In particular, to mention an elementary point, owning one's machines does not make them costless.

T or F: Friedman, the sweatshop shirtmaker, who owns his sewing machines outright, has a cost advantage over Schwartz, who must pay rent on her machines.

A: One way of seeing the point is to note that Friedman had to borrow \$500 (or whatever) per machine to become an owner and therefore must pay to the banker

interest equal to the annual rental Schwartz pays to an owner. Another and deeper way is to note that the real "cost" of a machine to Friedman is what it could earn outside his own factory, which is precisely (in a rental market without transaction costs) the rental that Schwartz would pay. By either argument, false.

A firm must often make a choice between one or another short-run cost curve (that is, between one or another set of transaction costs keeping the curve above the long-run curve). Stueland Electric in St. Joseph, Michigan, for example, must choose between arrangements such as owning outright its own trucks or supplies of cable that give the least cost when the firm is running at normal output and arrangements such as depending on a truck-renting firm in Baroda or a supplier of cable in South Bend that give higher costs at normal output but lower costs at very high or low outputs. If Stueland faces a predictable, stable demand, it will choose the inflexible but (perhaps) cheap arrangement of outright ownership. If it faces a highly variable demand, it will choose the flexible arrangement, for this will give the lowest cost over high and low outputs taken together. The choice is between the dashed total cost curve and the solid total cost curve in the top panel of Figure 12.7. The bottom panel displays the same choice in terms of average cost (that is, cost per unit), bearing in mind that if one cost curve is higher than another in total cost for a given output, it must also be higher in average cost for that output. If the firm faces outputs such as boom and bust frequently enough, it will prefer the curve Flexible to Inflexible.

The Long Run Is the Envelope of All the Short Runs

The engine of analysis can at the end be thrown into reverse, deriving the long-run cost curve from the short instead of viewing the short-run cost curves as deviations from the long. There are *infinitely* many short-run cost curves, some high, some low, some flexible, some inflexible. What is true of all of them is that the long-run cost curve is by definition below them and in fact consists of their combined lowest borders. The long-run total cost curve shows the lowest cost at which the firm can produce any given level of output. Speaking technically, the long-run curve is the *envelope* of all the short-run curves. In mathematics a curve that just touches all of a number of curves is known as the *envelope* of the curves, a term you can remember by thinking of the curve's

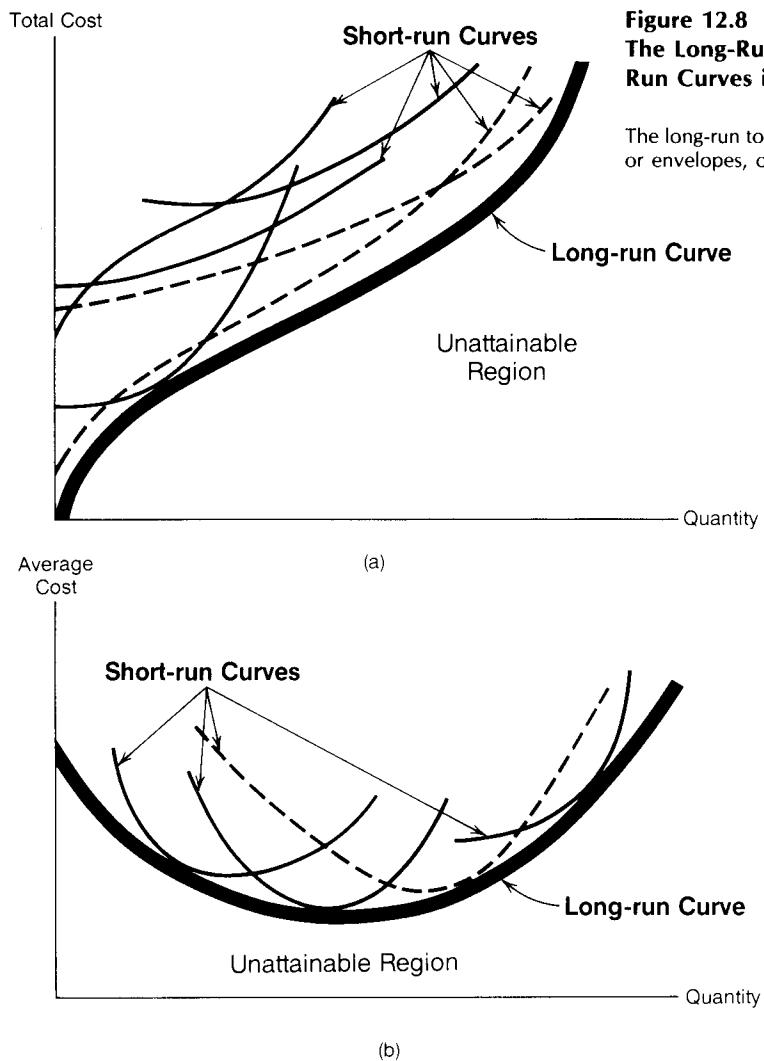


Figure 12.8
The Long-Run Cost Curve Is the Envelope of All Short-Run Curves in Terms of Total (a) and Average (b) Cost

The long-run total and long-run average cost curves are lower bounds, or envelopes, of all the points on the short-run curves.

enveloping (that is, surrounding) the other curves. Evidently, the short-run cost curves are always above their envelope, since being always below is how the envelope is defined (see Figure 12.8). The diagram simply restates the opening theme of the section: The best that a firm can do is its long-run cost curve, but it can do worse.

Summary The production function can be viewed as best practice, in which case the corresponding cost curve is lowest cost. Adding constraints hurts, a principle applicable in the present case to cost curves. When additional constraints drive firms off

the lowest curve, the cost observed is not a good estimate of the lowest cost. An important example of additional constraints is the transaction costs that force firms to operate along short-run cost curves. Firms do not fall blindly into such curves. On the contrary, they choose them, choosing, for example, a flexible plant that has lower costs calculated over episodes of boom and bust relative to an inflexible one that has low costs only if output is steady. In any event the long-run cost curve is the lowest attainable for a given output, which is to say that it is the envelope of all the short-run curves.

EXERCISES FOR SECTION 12.2

1. Suppose that you had a set of observations of cost and output such as the scattered dots in Figure 12.5. If you wanted to know the Cost Function, would you run a line through the points?
2. Explain in each case the relevance of the adding-constraints-hurts principle:
 - a. The 55-mph speed limit on balance hurts.
 - b. Energy conservation cannot make the nation better off.
 - c. If without railroads a shipment of wheat had to follow the pattern it in fact followed with railroads, the shipping cost would be higher than if the pattern was allowed to adjust to the absence of railroads.
3. *True or false:* Sears, Roebuck is better off owning the land under its stores than renting it, because then it doesn't pay rent.

PROBLEMS FOR SECTION 12.2

- 1. You are Jan deVries, the fabulously wealthy owner of a fleet of oil tankers. As a common carrier you are required by law to accept any shipment of oil demanded of you; that is, you have no control over your output (contrary to the assumption in the last few chapters that output is the one thing a firm *can control*). Initially your output of oil shipped is Old Output, and you have chosen a collection of ships that minimizes the cost of producing Old Output (see Figure 12.9). Output and total cost are those that will persist into the indefinite future. You are on the Old short-run cost curve.
 - a. Look at the vertical distance called Transaction Cost: It is the cost (all costs that persist into the indefinite future) of shifting from the Old to the (dashed) New short-run cost curve. Assume that it is the same length no matter where on the diagram it is moved. If output changes permanently to New Output, will you find it worth your while to adjust your fleet to get the New cost curve?
 - b. Suppose, however, that the old fleet deteriorates a little each year, driving the Old cost curve up. When will investment in an adjusted fleet take place?
 - c. Suppose that the New cost curve *falls* a little each year, reflecting the improvements in technology that permit a fleet built to embody the technology cheaper to operate. When will investment in an adjusted fleet take place?
 - d. Would you expect an industry with rapidly growing output, equipment that wore out quickly, and experiencing rapid technological change to have few or many occasions to invest in changing from one cost curve to another?
- 2. Marie and Brownie contemplate running an all-season skiing and camping resort near Montpelier, Vermont. They have two possible designs, one a general design that serves both for skiing in the winter and camping in the summer and the other a specialized design that makes skiing cheaper but camping more expensive. The costs expected into the indefinite future are as follows:

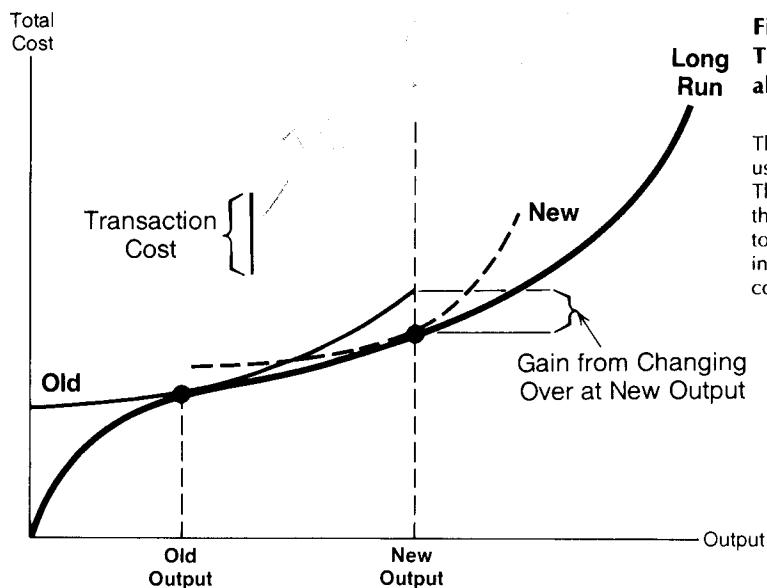


Figure 12.9
Transaction Cost Is a Vertical Distance, Placeable Anywhere

The vertical distance transaction cost tells the cost of using the market to change the method of production. The change shifts the firm from the Old cost curve to the New. The firm that must increase output from Old to New output will make the shift if the gain from changing over at New output is greater than the transaction cost.

	Winter	Summer
Number of guests	10,000	5,000
Cost per guest, general	\$10	\$10
Cost per guest, special ski	\$5	\$30

Suppose that the number of guests is given (the common carrier assumption) and that it is invariant to the choice of design. What is the unit of time over which costs should be measured? Which design has the lowest cost?

True or False

3. Compulsory conservation of energy will save the nation's resources.
4. Since the long-run cost is the lowest-cost way of producing each output, the long-run cost curve runs through the minimum points of short-run cost curves.

12.3 Cost Curves in Use: Fixed and Variable Costs

What to Read For	How do the total cost curve and the total revenue curve determine how much a firm wants to produce? Are past, sunk costs relevant? Why not? Are fixed costs relevant? What are the two meanings of "fixed costs"?
-------------------------	---

The Point of Maximum Profit Revisited

After so much attention to the domestic affairs of the firm, it is now time to return to its foreign affairs, that is, its decision of how much to produce for sale. The story is really just a review, because you already know how a firm arranges its foreign affairs—it sets output so that marginal benefit equals marginal cost. The goal of making profit is served by two steps. First, plan how to produce any output at least cost. Second, choose the output that maximizes the excess of revenues over costs. Lars Sandberg, a farmer in Ohio producing soybeans for sale to a market that will buy all he produces at a fixed dollar price, is in such a situation. His total cost has the usual stretched Z shape. His total revenue is a straight line out of the origin. This is because each additional ton brings him a fixed additional number of dollars (that is, the total revenue curve has a *constant* slope).

The top panel of Figure 12.10 shows what he does. He produces the output

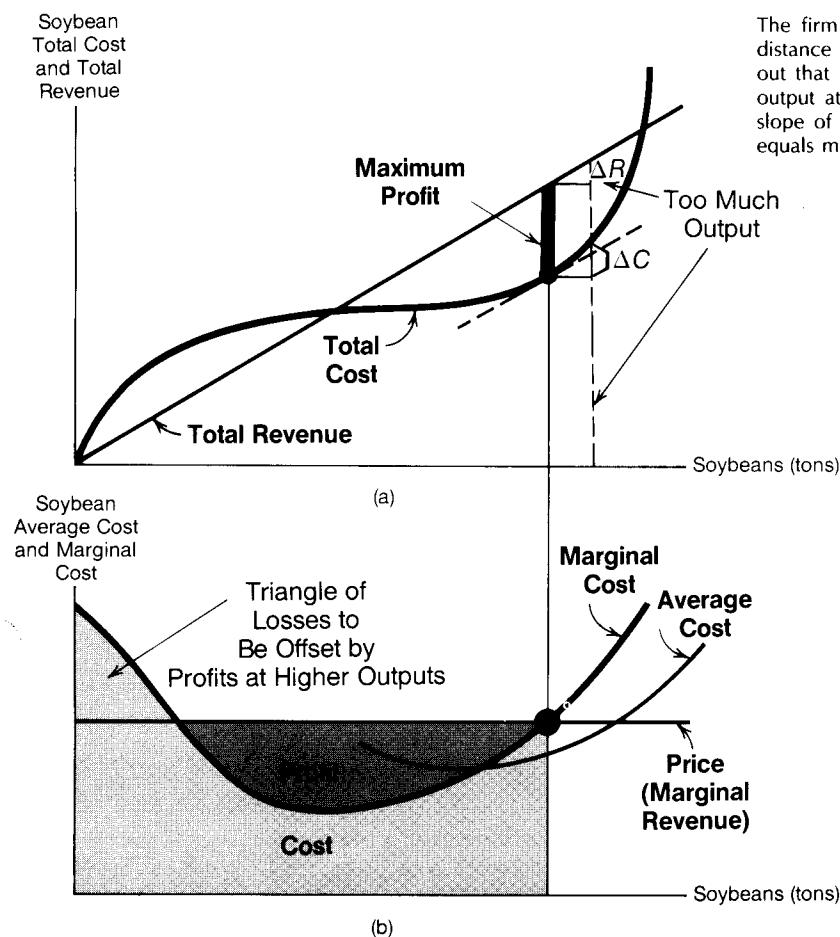


Figure 12.10
The Gap Between Revenue and Cost Is Profit

The firm maximizes profit by maximizing the vertical distance between total revenue and total cost. It turns out that it can always do this by choosing the rate of output at which the slope of total revenue equals the slope of total cost, that is, at which marginal revenue equals marginal cost.

that makes the gap between the total revenue curve and the total cost curve as large as possible, because the gap is profit.

The bottom panel gives the usual correspondences between the totals on the top and the averages and marginals on the bottom. As usual, in the bottom panel the Maximum Profit point is at the output that equalizes marginal cost and price. And as usual the amount of Profit is the area of revenue under the Price line (equal here to the two shaded areas) minus the area of Cost under the Marginal Cost curve. The heavy line in the top panel is an alternative picture of the same Profit. It will not come as a complete surprise that the Maximum Profit is at the output at which the slope of the total cost curve (look at the dashed tangent) is equal to the slope of the total revenue: These slopes are exactly the marginal cost and the price, and by a well-known argument equalizing them brings bliss. You can overcome any surprise by translating the argument into the terms of the total cost diagram. At Too Much Output, for example, total revenue has risen above what it is at Maximum Profit by the amount ΔR , but total cost has risen even farther, by ΔC . Clearly it will be better to move back to Maximum Profit.⁵

Past Costs Are Irrelevant

The central message of the diagram—and of the last two chapters—is that in the pursuit of profit a rational firm is influenced by its costs. What costs? *Future costs that can be altered by the amount of output.*

Q: A jeweler in Harper Court leaves the prices of her gold necklaces at their price when made even though the price of gold has since doubled. She says to an amazed and incredulous but grateful customer that "I make money even at the old prices." Does she?

A: The "cost" relevant to a fully rational firm is not yesterday's outlay of money but the cost from the moment the sale is made *into the future*. To take an even more extreme case, if gold prices were going to triple tomorrow and if God had informed the jeweler, then

it would be madness to make any sales today. Likewise in the present case: The true "cost" is the opportunity cost, that is, what the necklaces would bring now or later, at another time or place. A doubling of the price of gold sharply increases the cost (though by less than doubling, for the labor and capital in fabrication and in marketing would not double in price and are also part of the cost). Selling a necklace worth \$70 for \$50 is to incur on this account a \$20 loss of opportunity, not a gain.

The point is that historical costs are not directly relevant to the forward-looking decisions of a rational firm. Historical costs are, as the vivid word in accounting has it, "sunk." The proverb is "let bygones be bygones." Forget about the past and look to the future, for purposes of moneymaking if not of other sorts of wisdom. That Manhattan once cost \$24.00 does not mean that for present purposes owners should care. If the owner of the land under the World Trade Center foolishly sells the land for \$24.50 he cannot comfort himself by thinking, "Well, at least I made some profit: Once the whole island sold

⁵ First-year calculus reduces the matter to routine. Remember, though, calculus is not necessary. Suppose that the cost function were some function $f(q)$, where q is the output chosen. Total revenue is of course pq , where p is the given price. Profit is then revenue minus cost, $pq - f(q)$, and to find the condition for maximizing this profit function, one sets its first derivative with respect to q equal to zero. Well, its derivative is $p - f'(q)$, which when set equal to zero implies $p = f'(q)$. This says that the maximum gap of profit occurs when the slope of the total revenue function (in this case the slope is just p) is equal to the slope of the cost curve, $f'(q)$. Profit is maximized at the output level for which marginal revenue is equal to marginal cost.

for less." Firms in regulated industries, to be sure, must often pay close attention to historical costs, because the accounting principles demanded by the regulations do. A regulated telephone company constrained to earn no more than 5% on the acquisition costs of its assets must calculate the acquisition costs historically, not at the cost in future that is relevant to their economic use. Furthermore, the past acquisition cost of even a nonregulated firm, such as the Evans Slipper factory, may be a useful estimate of its present cost of replacement if not much has happened since it was built to alter the cost of building or if what has happened is measurable. But these uses aside, historical costs do not determine present and future costs and are therefore useless for measuring profit.

Unavoidable Fixed Costs Are Past

A closely related point is that fixed costs do not affect the output of the firm. There are two distinct types of *fixed costs*. One type is fixed costs that are "fixed and unavoidable." These have as much effect on present and future decisions as do historical costs: none. Fixed costs in this sense, for example, are often identified with the repayment of debts incurred to invest in equipment. If you borrow \$10,000,000 from your bank to invest in a ship, you must pay it back with interest. Suppose that your monthly payment to the bank is \$150,000. You would be pleased to earn more than this amount from the ship, and surely you expected to when you made the investment. But whether or not you actually do so is irrelevant to your behavior as a shipowner. *Whatever you do*, the bank each month presents you with a bill for \$150,000. The bank could care less if you are still a shipping magnate or if business has been good. For your part, that you face a bill for \$150,000 affects your business no differently than would a bill for \$150,000 for something unrelated to your business, such as child support. The way in which you run the shipping business is unaffected by the burden of debt "on" it. The fixed and *unavoidable* cost has no effect on your business decisions--whether or not you keep the ship and how much a year you run it.

Q: British owners of old coal-powered (as distinct from oil-powered) ships after World War I earned enough in revenue to cover the captains, crew, supplies, fuel, and the like for operating the ships, but not enough to also cover the interest and repayment costs on the loans taken out before the war to buy the ships. *True or false:* The shipowners were probably acting irrationally when they bought the ships and were certainly acting irrationally when they continued to operate the ships despite the losses.

A: What matters to a judgment on the rationality of the initial investment is the *expectation* of future profits.

The shipowners could well have made investments in 1910 in expectation of making money yet could find later that their expectation was mistaken, as it in fact was. And having made the initial mistake, the decision is bygone, irrelevant to the decision a shipowner faces in 1930 of whether or not to keep the ship afloat. So long as the ship earns enough to pay all the *variable* costs (the costs that do vary with output), the rational decision is to carry on. Therefore, false.

Avoidable Fixed Costs Determine Whether or Not to Produce Anything

There are two distinct types of *fixed costs*. The first has just been discussed: fixed and *unavoidable*. The second is fixed and *avoidable* if one closes down entirely. Payment on a loan used to open a restaurant is fixed and unavoidable even if the restaurant closes down. Payment of a license to operate the restaurant is fixed (that is, it does not vary with the output of the restaurant) but avoidable if it closes down and no longer begs permission of the sovereign power to exist.

To look at it from the other side, there are two kinds of variable costs: costs that are zero when the output is zero and costs that are *not* zero when the output is zero. The significance of the distinction arises from problems of the following sort.

T or F: Since a tax of \$1000 per firm imposed on the existence of firms does not affect marginal cost, it does not affect the output chosen by each firm (unlike a sales tax or an employment tax) and therefore does not affect the output of the industry.

A: The successive assertions travel from plain truth to probable error by small steps. The first assertion is plain truth and deserves a lot of discussion.

It is true that a tax of \$1000 per firm does not affect marginal cost. The \$1000 is a fixed cost, invariant with output, given that *some* output is to be produced. The total cost (including here both fixed and variable cost) will move up by \$1000 when the tax is imposed, thereby cutting profits by \$1000. But because a uniform rise of \$1000 in the cost curve does not alter its slope, the corresponding marginal cost curve is not moved (see Figure 12.11). That the shaded area of the Profit on Variable Cost in the bottom diagram does not change after the tax even though the Profit Gap in the top diagram shrinks is due to its exact definition, now made explicit for the first time. It is profit on *variable* cost alone, because it is calculated by subtracting from revenue the area of the costs under the marginal cost curve (that is, all and only those costs that do *vary* with output).⁶ Because the marginal cost does not move from its old position, the optimal output does not change. The economic common sense here is that the firm always wants to have as large a gap as possible between revenue and cost, whether the gap is big or small. The tax makes the gap smaller, and may even make it a loss, yet the firm still does as well as it can under the circumstances by setting marginal cost equal to marginal

revenue. It is often said that low profits spur the firm on to greater effort and larger outputs; and it is also often said that they discourage the firm, causing less effort and smaller outputs. The middle ground claimed here, by contrast, says that the size of the profit earned has no effect whatever on the output chosen, given that some output is to be produced. The assertion is not self-evidently true, which is to say that it is a hypothesis about how people behave that might in some cases be false. In any event, it is the assumption made in the usual theory of the firm.

The later assertions in the question, however, do not follow from it. True, the \$1000 tax per firm does not affect output *given* that some output is to be produced. But there's the rub. That some output is to be produced is *not in the long run given* but is a choice made by the firm. If the tax is so large that profits on variable costs are offset entirely, then the firm will close down. And *in the long run all costs are variable*; that is, all costs are avoidable. In the long run, then, the tax does affect output, for it affects whether or not the firm produces at all. Therefore it can affect the output of the industry. As always, an extreme example helps. Suppose that the tax were \$100,000 instead of \$1000 and that it were imposed on each hot dog stand. Clearly, none would exist at such a high price of existence. To take the other extreme, suppose that the tax were \$1. Clearly, the effect on the decision to exist as a hot dog stand would be trivial. The effect on the industry running hot dog stands is the more powerful the larger the effect on profit. In short, the answer is false; a \$1000 tax *can* affect the output of the industry.

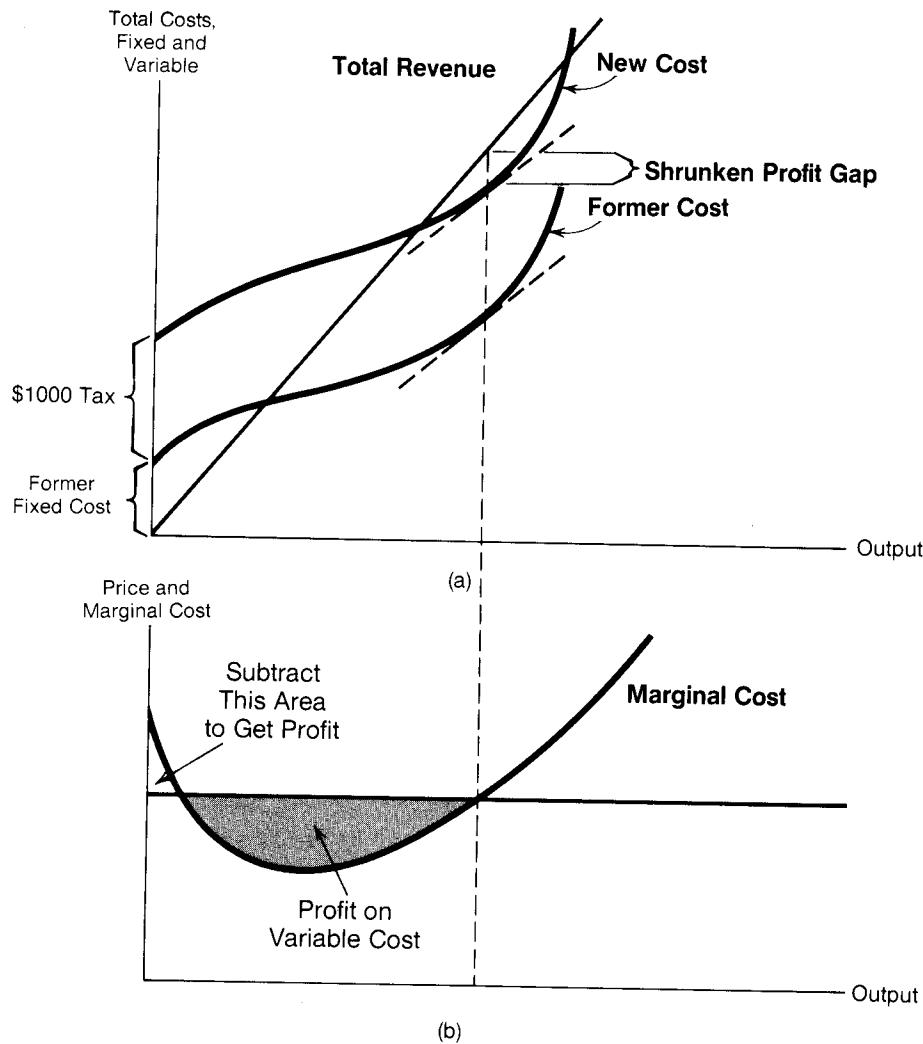
⁶ Fixed cost, then, is a constant of integration in the calculus problem to find $\int_0^q MC(q) dq$, in which MC is the marginal cost function. Or, to look at it the other way, the total cost is variable cost (a function of q) plus fixed cost (a constant), meaning that marginal cost is the derivative of variable cost alone,

$$\begin{aligned}\frac{dT C}{d q} &= \frac{d[V C(q) + F C]}{d q} \\ &= \frac{d V C(q)}{d q}\end{aligned}$$

since FC is a constant.

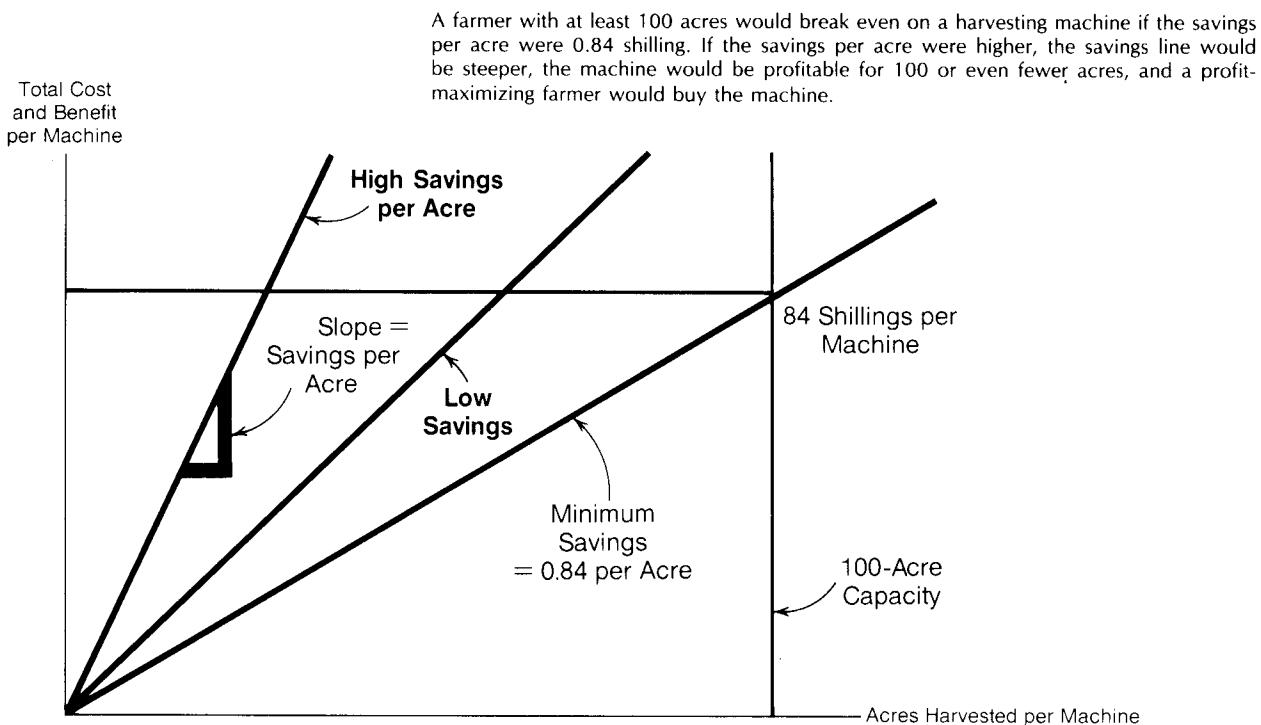
Figure 12.11
A Fixed Tax per Firm (a) Does Not Move Marginal Cost (b)

A tax that does not vary with output increases total costs by the same amount at all levels of output and therefore does not change marginal cost. As long as the firm produces some output the tax will have no effect on how much the firm produces.



Average Cost Is Only Relevant to the Decision Whether to Produce at All The logic of fixed and average costs, then, though poorly suited for questions of what output should be, is well suited for questions of whether any output should be produced at all. For example, the investment in equipment that a firm makes is analogous to a tax. Profit over variable cost must cover the cost

Figure 12.12
The Economics of Lumpy Investments



of the investment in the long run if the firm is going to embark on it. The analogy can be used to second-guess the decisions made by managers.⁷

Q: By the 1870s a reaping machine to replace hand harvesting of wheat and barley had been available for decades, yet in England (unlike the United States) less than half the harvest was done by machine. It has been argued that this was a result of the small size of English farms, the high cost of machines, and the inconvenient nature of the English landscape (plowed for purposes of drainage into a "ridge-and-furrow" configuration, which made the cumbersome reaping machines less effective).

1. If a farmer bought a machine for 660 shillings on which he had to earn 12.7% a year to pay back the banker, what in shillings does the machine have to pay back every year to justify its purchase? If the

labor saving on flat land *not* plowed into ridges and furrows were 3.3 shillings per acre harvested per year, what would be the lowest annual acreage harvested (the "threshold" acreage) at which a machine would become profitable?

2. The labor saving due to the introduction of the reaping machine was lower on ridge-and-furrow land. Illustrate in a diagram of total cost against acres harvested per machine per year how the threshold acreage varies with the per acre labor saving. If the maximum annual capacity of a machine were 100 acres harvested, what is the minimum labor saving that will make the machine profitable?

3. If the labor saving on a ridge-and-furrow farm were

⁷ The example is drawn from Paul A. David, "The Landscape and the Machine," in D. N. McCloskey, ed., *Essays on a Mature Economy* (London and Princeton, N.J.: Methuen and Princeton University Press, 1971), reprinted in David's book, *Technical Choice, Innovation and Economic Growth* (Cambridge: Cambridge University Press, 1975).

2 shillings an acre, what would you conclude about the rationality of *not* adopting the reaper in view of the fact that the wheat and barley acreage on a majority of English farms (mostly ridge-and-furrow) was less than 50 acres (with a good deal of variation around this average)?

4. How does this conclusion change if farmers can share a machine, if a market in machine time exists, if the size of farms can be altered, or if the percentage of a farm devoted to wheat and barley can be altered?

A: 1. The annual opportunity cost of the investment is $0.127(660) = 84$ shillings. This is a fixed cost with respect to the number of acres harvested. You pay it regardless. So if the machine saves 3.3 shillings an acre, you had better have at least $84/3.3 = 25$ acres to harvest with your machine.

2. The diagram is given in Figure 12.12. The minimum labor saving is simply $84 = 100(S)$, that is, $S = 0.84$.

3. The threshold acreage for 2 shillings an acre saving is $84/2 = 42$, so a good many farms in England harvested acreages below the threshold (some harvested only, say, 20 acres). These farmers were not being irrational to

ignore the charms of mechanized harvesting. One could build from these elements a full theory of the adoption of the reaper: As the benefits varied over time or as the size of farms varied, the percentage of the distribution of farm sizes that adopted the reaper would vary.

4. The conclusion of (3), however, changes radically if machines can be rented, loaned, or shared. Farmers could club together to buy a machine, for example, and harvest 100 acres with it each year, even if they all have pitiful 15-acre stands of wheat and barley. Or a farmer with a reaper on 70 acres could notice that his machine stood idle for the last week or so of the harvest and could rent it out to his reaperless neighbor. Or a company could be formed specializing in reaping. The threshold idea no longer has force, and the failure to adopt reapers remains a puzzle, as long as the labor saving on ridge-and-furrow land was more than 0.84 shilling an acre. Furthermore, the threshold could be changed if the size of farms or the percentage of a farm devoted to wheat and barley could be raised. The acreage to be harvested, in other words, is not given by God but is alterable, in which case it might be altered to take advantage of the machine.

Summary A firm achieves the most profit by equalizing marginal revenue and marginal cost or (equivalently) finding as large a gap as possible between total revenue and total cost.⁸ The total cost in question is future, variable cost. It is not past, fixed costs. Firms, like all rational folk, let bygones be bygones, do not cry over spilt milk, and look to the future. A tax levied in a lump or a machinery cost that can be avoided by refraining from any output, of course, is subject to choice by the firm. Costs of doing business do not affect marginal cost and therefore do not affect output if some is forthcoming. But if the cost imposed is too large, the firm will shut down entirely. The question—taken up in detail in Chapter 14—becomes whether or not profit on variable cost (such as the revenue over cost in a restaurant or the savings on labor cost in buying a machine) is large enough to offset the cost of licenses, buildings, or reaping machines. To be or not to be, that is the question.

EXERCISES FOR SECTION 12.3

1. This exercise is meant to convince you that where marginal cost equals marginal benefit is where the gap between total cost and total revenue is greatest. The diagram you construct is used later, so do it with care. Get a piece of graph paper (or simply use lined notebook paper, imagining where the vertical lines would come). Draw some axes without units, labeling the vertical one “total cost and total revenue” and the horizontal “quantity.” Leave

⁸ It might as well be admitted here that the phrase “total cost” is ambiguous in economics. It sometimes means “all cost, including fixed and variable costs,” but sometimes, as here, it means “cost in total, as distinct from marginal or average cost.”

room below them for other axes. Now draw freehand a stretched Z curve of Total Cost in the style of the Total Cost curve in the top panel of Figure 12.10. Make sure it is a Z, not an S. Draw a straight-line Total Revenue curve cutting through the cost curve. Now mark off the distances on your axes, dollars on the vertical, and, say, tons on the horizontal (choose easy-to-work-with numbers, like 100, 200, 300, and so forth). No tricks, right? You've supplied all the data so far, except the general shape of the curves. So if it turns out that the output where marginal cost in your diagram equals marginal revenue *is* the output where the gap between total cost and revenue is greatest, you'll have to believe that it's a property of such curves in general, not some accident or some sleight of hand by economists.

All right. Now locate and mark where the vertical gap between revenue and cost is greatest. If your curve is curvy enough you can do it with your eye; otherwise you can move a straight edge of paper around until the greatest gap is marked off on it.

Draw axes exactly below these, again in the style of Figure 12.10. Mark off the (horizontal) quantity axis just as it is in your top panel. Calculate the slope you have implicitly given the Total Revenue curve. It will be, of course, the rise over the run, so many dollars per ton. In fact it will be the price; and since the line is straight it will be one price. Choose a convenient point on the (vertical) average-and-marginal-cost axis of your bottom panel to represent this level of price, and draw the Price line (consult Figure 12.10).

Now the delicate part. Using a straightedge pushed up against various points on the Total Cost curve, find the numerical values of the slopes of the various tangents to the curve. Remember: rise over run, so many dollars per ton. These will be, of course, marginal costs: Plot them in your bottom panel, making sure to line them up with the same tonnages and making sure that their heights are exactly as they should be relative to the Price line. Do enough points to sketch in the Marginal Cost curve. At what tonnage does Marginal Cost intersect Price? That's right: the same tonnage as the one that makes the gap between total revenues and costs a maximum. Q.E.D.

2. By way of review of cost curves, use the diagram constructed for Exercise 1 to prove that Marginal Cost intersects Average Cost at the minimum of Average Cost. Average Cost at some output, you recall, will be the slope of a ray from the origin out to that point on the Total Cost curve.

3. Again by way of review of cost curves, use your diagrams to show that the size of the area analogous to that marked Cost in the bottom panel of Figure 12.10 is the same as the *height* of the total cost curve. To do this you will need to divide up the area under Marginal Cost into ten or so little rectangles, each with some easy-to-reckon base marked off along the tonnage axis and each with a height running through that part of the Marginal Cost curve. The area of all the rectangles will not be *exactly* equal to the total cost because the Marginal Cost curve is not *exactly* a series of little steps like the tops of your rectangles. But it will be pretty close. The equality proves the crucial point about marginal diagrams: Areas in marginal diagrams correspond to some vertical distance in the corresponding total diagram.

4. Draw a *third* diagram with tonnage on the horizontal axis and Total Profits on the vertical. This can be done by measuring the vertical gaps between revenue and cost and then plotting the result (if revenue is below cost, of course, the profit is negative). Where is profit a maximum? Define the idea of *marginal profit*. (*Hint:* Look at your second, marginal-cost-and-revenue diagram. What is the profit earned from the first unit produced? From the second?) *True or false:* When marginal profit is zero, total profit is at a maximum.

5. Which of the following are fixed costs? Distinguish those that are fixed and avoidable from those that are fixed and unavoidable:

- Heating expenses of the college buildings.

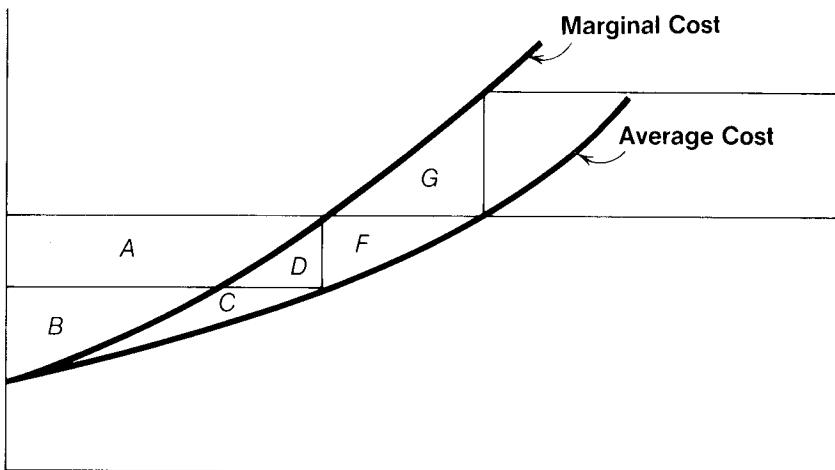


Figure 12.13
Areas of Profit When Costs Rise

- b. Payments by the Wretched of the Earth Mining Company of Pretoria to the college (which loaned Wretched money by buying its bonds).
- c. Payments by the college to tenured faculty. [“Tenure” is alleged to mean that one cannot be dismissed from one’s job except for gross moral turpitude or gross dereliction of duty (Look those up).]
- d. Payments to untenured faculty.
- e. Fees paid to the state by the college to license its dining facilities.
- f. Fees paid by the college to be a college.
- g. Purchase of exam booklets by the college.

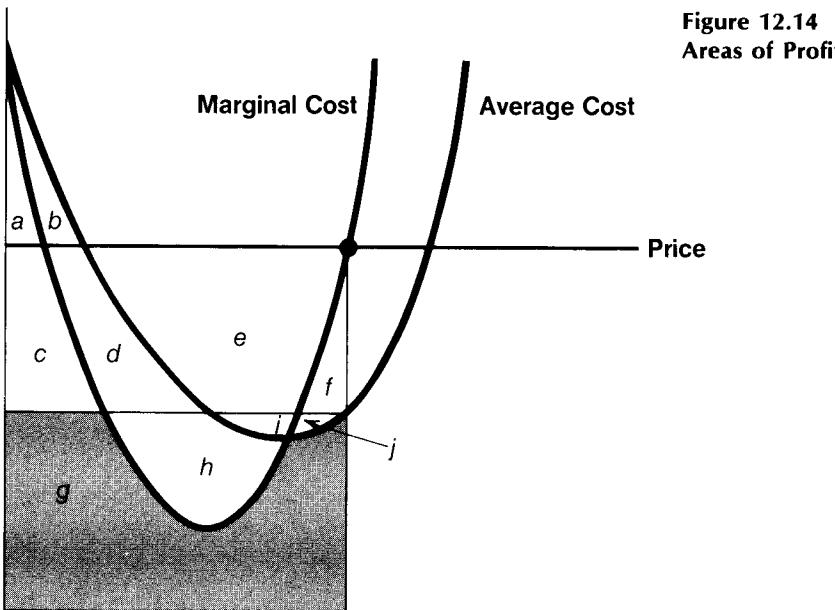


Figure 12.14
Areas of Profits When Costs Fall, Then Rise

PROBLEMS FOR SECTION 12.3

- **1.** Label and explain two alternative and equal areas of profit in Figure 12.13.
- **2.** Do the same for Figure 12.14.
- **3.** In the diagrams of total revenue drawn in the text, it is assumed that price (average revenue being the slope of a line from the origin out to the place on the total revenue curve) does not fall with larger quantities sold. If Saudi Arabia produces substantially more oil, however, the price *will* fall. Only oil at a lower price will find new demanders.
 - a. In such a situation, what is the shape of the total revenue curve?
 - b. Suppose that Saudi Arabia has the usual stretched Z for a total cost curve and faces the total revenue curve described in (a). Where would Saudi Arabia set its output to maximize profit? Would the resulting price be equal to marginal cost?
- **4.** An Iowa City councilman argued that "We spent \$30 million refurbishing the downtown: It would be silly not to build a highway now to make it easy for people to get to the downtown." Criticize.

Competitive Industry

13.1 Industry Supply with a Fixed Number of Firms and Costs Independent: Optimality and Upward Slope

What to Read For

How is the marginal cost curve for a competitive firm related to its supply curve? How do the curves of marginal cost for a bunch of individual firms add up to make the supply curve for the industry? Why are profits necessary to get the individual firms to produce the efficient amount? Would it be good if some of the firms did not pursue profit? Why does the supply curve of the industry always slope upward?

The Supply Curve of the Industry Adds Up Marginal Cost

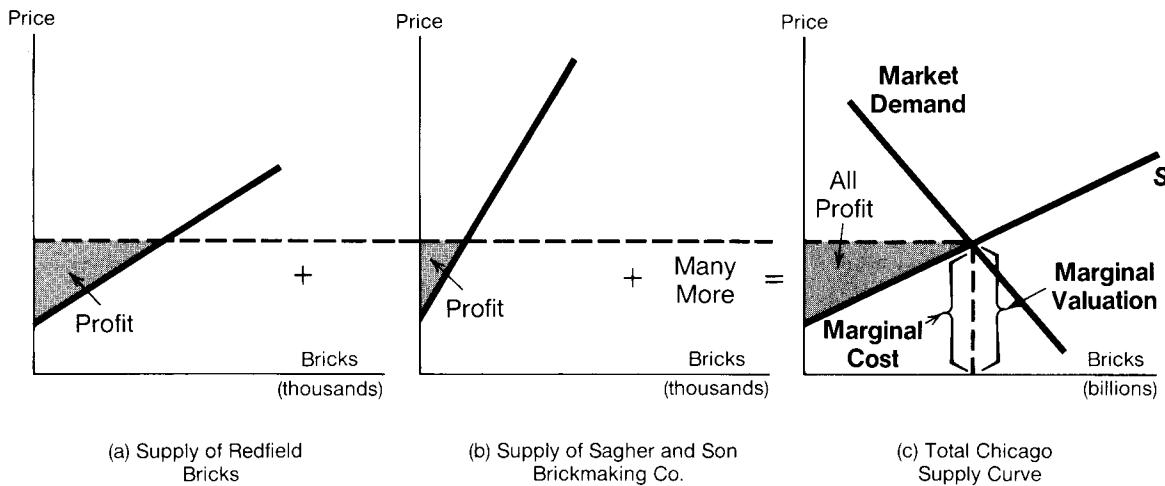
Recall that the theory of the typical consumer—of utility functions, budget lines, and all that—is useful in itself, but it is even more useful when the consumers are added up into market demand. Likewise, the theory of the typical firm—of production functions, cost curves, and all that—is useful in itself, but it is even more useful when the firms are added up into market supply.

The first step in the adding up is to notice that a firm need only be a price taker for its marginal cost curve to be its individual supply curve. Look at Figure 13.1. If a brickmaking firm is such a small part of the whole supply of bricks that it cannot change the price by its own decisions on how much to produce, then it takes its demand curve as flat. That is, it takes the price of bricks as given and maximizes profit subject to the given price. Maximizing profit leads it to produce the output of bricks that brings marginal cost up to the marginal benefit (which is the price). In other words, for price takers the marginal cost is the firm's individual supply curve.

The next and final step is to the brick industry as a whole. In the same way as one adds up each person's demand curve horizontally when each takes price as given, one adds up each firm's supply curve of bricks to get the market supply. Many firms making bricks yield the Total Chicago Supply Curve of bricks.

Figure 13.1
Add Curves of Marginal Costs to Get the Curve of Supply

The supply curves of price-taking firms are summed parallel to the quantity axis to give the industry supply curve. When each firm is at its optimum and the market is in equilibrium, the marginal cost of a unit supplied equals the marginal valuation of a unit consumed.



The Industry Produces the Best Output

The first of several lessons from this construction is the familiar one announced by Adam Smith in 1776 (Chapter 8) that profit maximizing leads to efficiency. One part of the meaning of "efficiency" made plain in the diagram is that the whole output is the socially correct size. Look at the shaded areas of profit for each firm. These are maximized, subject to the condition that market supply equals market demand and that the firms believe themselves individually powerless to alter the price. No other output of the Redfield company, say, suits it better; no other gives higher profits. Since the industry as a whole must supply only the amount demanded, the marginal valuation (the height of the demand curve) in equilibrium will be equal to the marginal opportunity cost (the height of the supply curve). Like an earlier diagram showing this argument, the scrambling for profit leads to the point on the production possibility curve at which the curve touches an indifference curve. At this point the slope of the production possibility curve is equal to the slope of an indifference curve. Without profits to guide them, capitalist firms would not produce the best output of bricks.

Q: A bumper sticker reads "People, Not Profits: Smash Capitalism." Comment.

A: The money-grubbing, selfish, penny-pinching, greedy, avaricious capitalists yield in their struggle with each other the highest income for the people, the correct

output of bricks, the highest attainable point on the society's opportunities. Private vice is public virtue. We the people are the capitalists, as owners or employees. We want the size of output to be efficient and can attain it by unleashing the dog of capitalism.

Each Firm Produces the Best Output

Another part of “efficiency” is also plain in the diagram, namely, that the output is allocated among firms in such a way that it is produced at least cost. The reason is simple. Each firm faces the same price. Therefore, each ends with the same marginal cost. The principle is that of equimarginality, as discussed in Chapter 11 for the allocation of students between the two branches of the University of Maryland. In terms again of the diagram of an earlier chapter, the economy is brought to the production possibility curve. That’s least cost: producing as many bricks for a given amount of all other goods as possible.

Q: Look at the diagram of the Redfield and Sagher brick firms. Suppose that the two companies make up the whole market supply. Suppose too that they take the price as given to them by the market. Use straight-line supply curves.

1. If the allocation of output between the two is as portrayed in the diagram, what is the *marginal* opportunity cost of bricks (that is, the cost of the “last” brick) in terms of all other goods for each firm? What is the *total* (variable) opportunity cost (that is, the amount of all other goods the society gives up to get all the bricks, not just the last one)? (*Hint:* Add up marginal cost.)

2. Now suppose that Redfield Bricks, Inc., does not maximize profit but instead produces for any given price 10% *less* than its profit-maximizing output. Compared with the profit-maximizing case, what will happen to the industry supply curve? Suppose—merely to keep output the same as before to facilitate comparison—

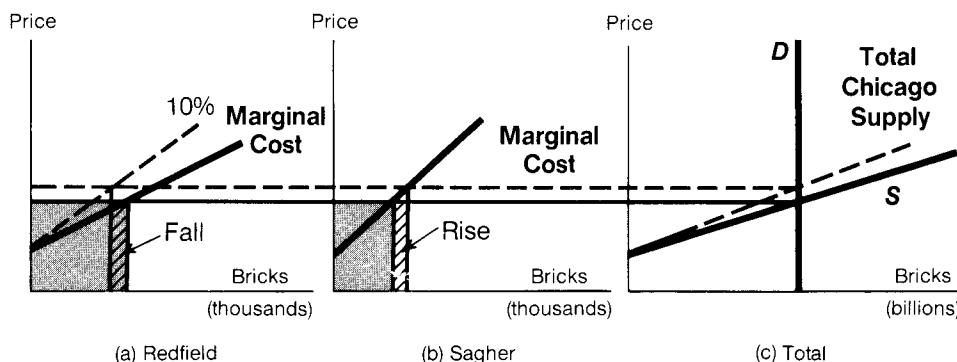
that the industry demand curve was perfectly inelastic. What happens to Redfield’s output? Sagher’s? (Remember: inelastic demand.) In equilibrium, what now is each firm’s marginal cost? Does equimarginality hold? What is the industry’s total cost of the output produced? How does it compare with the total cost when Redfield Bricks, Inc., *does* maximize profits? Is profit maximizing good for society?

A: 1. The marginal cost is the same for each firm, equal to the going market price. Recall that a competitive firm that is operating efficiently will produce the output level at which market price is equal to marginal cost. The total variable cost is the area under its marginal cost curves out to the quantities produced by each firm, the shaded areas in Figure 13.2 (ignore all the dashed lines for a moment).

2. With Redfield’s supply swiveled back 10%, the whole Chicago supply swivels back by the same absolute

Figure 13.2
Failure to Be Greedy Puts the Society Below Its Production Possibility Curve

If Redfield produces less than the profit-maximizing quantity of bricks, market price rises, Sagher expands output, and total industry costs increase because the marginal costs of production at Redfield and Sagher are no longer equal. Equimarginality is violated.



amount. So the market price will rise. At the higher price Sagher will be induced to make more bricks than in the profit-maximizing case. The price cannot rise so much that Redfield also would be making more bricks. If it did, both would be making more and the market output would rise, which is impossible with a higher price unless the demand curve has an upward slope—an absurdity. That is, Redfield necessarily makes fewer bricks than he did in the profit-maximizing equilibrium. Indeed, with the demand curve perfectly inelastic, the reduction in his output is precisely the same as the rise in Sagher's.

Now consider the costs. That Redfield ignores his true marginal cost does not mean it is not socially relevant. It is. The true marginal costs are now different,

Redfield's being below Sagher's (look at the equilibria marked by the dashed lines). *Since the principle of equimarginality is violated, one would expect costs to be needlessly high.* One would be right. The total cost for the unchanging output rises by the emphasized area in the Sagher diagram marked Rise and falls by the area in the Redfield diagram marked Fall. The Rise is larger than the Fall. The conclusion is that *if Redfield does not pursue profit, the total opportunity cost of producing the given output is higher.* If the firms in an industry do not follow the Rule of Rational Life, the society is pushed inside its production possibility curve. The society has to sacrifice for a given amount of bricks a larger amount of all other goods than necessary.

A central planner could achieve the same allocation as do rational firms, but he would need the information on profits they have. It is for these reasons that many socialist states embrace the profit motive, Yugoslavia being the most successful example. The state lets firms keep profits, adopting the capitalist system of profit maximization and the capitalist prices that give it meaning in order to achieve efficiency. Profits are good for you, even when you don't earn them.

The Industry Supply Curve Slopes Upward

Another general point that comes out of the construction of an industry supply curve from marginal cost curves is that the supply curve slopes upward. At the quantity a firm produces to maximize profits its marginal cost curve must be sloping upward, not downward.

T or F: If a firm's marginal cost curve is U-shaped and if the market price cuts across the U, the upside cut is at the quantity of maximum profit, the downside cut at minimum profit.

A: True. The area under the price out to some output is the revenue from that output; the area under the marginal cost is the (variable) cost (see Figure 13.3). The difference between the areas is profit. At the Worst Output, for example, revenue minus cost is $R - (R +$

$L)$ or the top left area L , an area of *loss*, not profit. The loss is evidently greatest at Worst. No firm would operate on the falling portion of its marginal cost curve. At Best, however, revenue minus cost is the rectangle $I + (R + C)$ minus the area under the marginal cost curve $L + (R + C)$, or $I - L$, with the loss offsetting some of the gain. No rational firm would operate at a point such as Worst. Only the upward-sloping portion of the marginal cost, then, is the firm's supply curve.

The Lower Portions of the Marginal Cost Curve Do Not Give Enough Profit to Be Relevant

In fact, only a portion of the upward-sloping portion is the supply curve. This is because *the price must be high enough to make the profits on variable cost positive.* A shoe factory facing such a low price for its shoes that it cannot pay even for its variable costs of labor and materials, not to speak of the fixed and unavoidable short-run costs of its buildings and machines, does better to close its doors entirely. See Figure 13.4. The lowest short-run supply price is the price that just equalizes the area of loss, L , with the area of surplus, S , giving zero profits. Another and exactly equivalent way to put it is to say that

Figure 13.3
Price Equal to Marginal Cost Minimizes Profit if Marginal Cost Is Falling

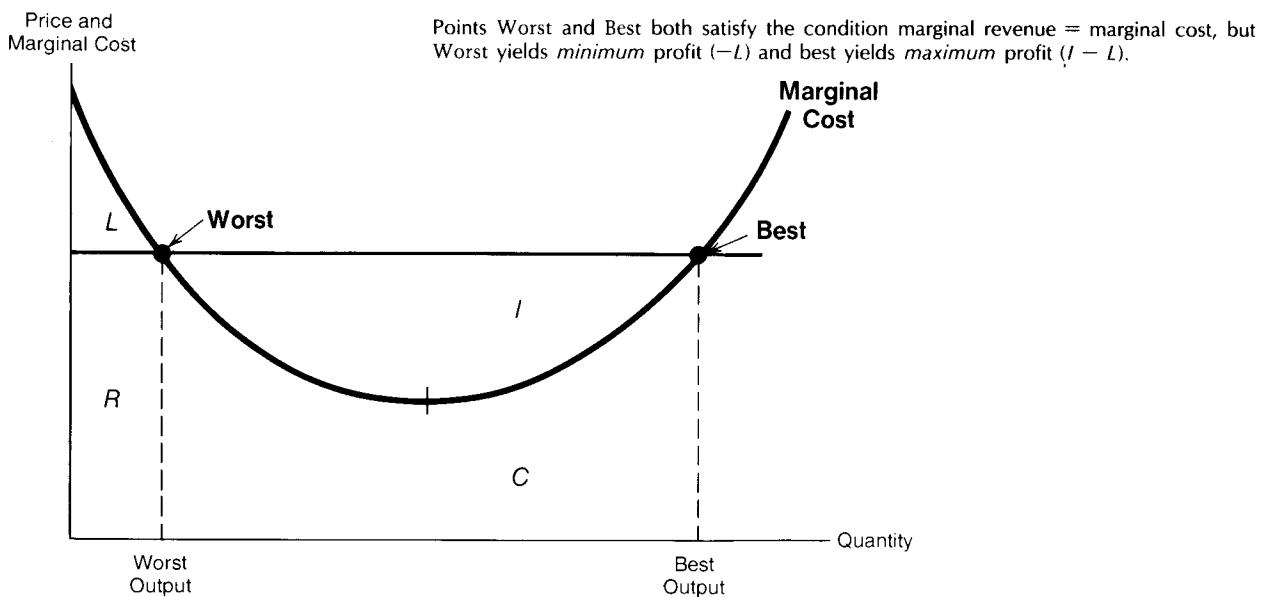
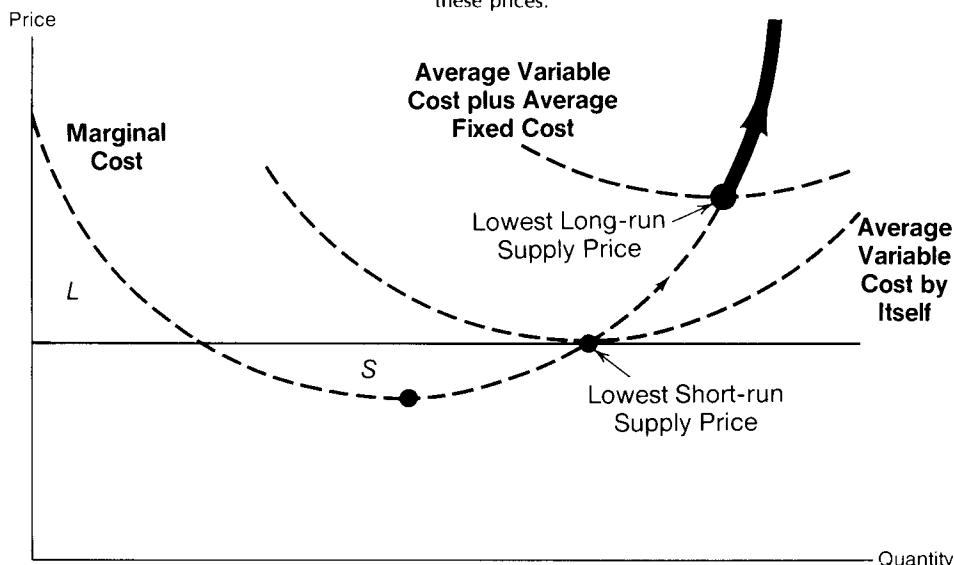


Figure 13.4
Only a Portion of the Rising Marginal Cost Curve Is the Firm's Supply Curve

The firm will produce nothing even in the short run unless price is at least equal to minimum average variable cost and, in the long run, will produce only if price is at least equal to minimum average *total* cost. So the firm's short-run and long-run supply curves lie above these prices.



average variable cost just equals the going price.¹ For a time the firm will be willing to supply along the portions of marginal cost marked in the diagram with arrows. In the longer run the fixed cost of investment must be covered, which is to say that the area of surplus, S , must be large enough to offset not only the area of loss, L , but also the fixed cost. Again, an equivalent way to put it is to say that average *total* cost just equals the price. In the long run, then, the supply curve is only the upper, heavy line.

In any event, the supply curve of each firm slopes upward. Because it does the sum of them all does, too. The industry's supply curve normally slopes upward, not downward.

Summary

The marginal cost curve of a competitive firm is its supply curve. Therefore, the industry supply curve is the horizontal sum of all marginal cost curves, if the number of firms in the industry is fixed. The equilibrium of demand and supply that results will satisfy the rule of rational life, because the marginal valuation of, say, wheat is equal to the marginal cost from each farm. Wheat farming as a whole is at its efficient amount and the growing of wheat is allocated among farms in the most efficient way. A market organizes its affairs as though it were one rational individual or one perfectly knowledgeable planner. This is why many of the socialist governments of Eastern Europe use the market with such capitalistic enthusiasm.

The industry supply curve slopes up because each firm's supply curve slopes up. And each firm's supply curve slopes up because only the upward-sloping portion of its marginal cost curve is a rational place to be. In the short run only the portion above average variable cost is rational, and in the long run only the still more limited portion above average total (fixed and variable) cost. The cost curves of the firm, in sum, have implications for the supply curve of the industry.

EXERCISES FOR SECTION 13.1

1. Imagine 100 brick firms in Chicago, all about the same size, competing for the brick dollar. What share of the whole demand for bricks will Kruskal and Co., a typical firm, supply? (*Hint:* Note that he's a "typical" firm among 100). Suppose Kruskal makes a major change in his plans and increases his output 33%. How much will the *industry's* output increase in percentage terms? What do you conclude about the ability of one supplier to influence the industry's output?

2. If the elasticity of demand for bricks in Chicago is 1.0, how much will Kruskal's 33% increase of his own output cause the price in Chicago to fall? (*Hint:* It's not 33%. Why not?) The price elasticity of demand facing Kruskal is, of course, the percentage change in his quantity associated with a 1% change in price. It is just a fraction. What is Kruskal's elasticity of demand (note that it is the elasticity of demand, not supply, that is at stake here)?

3. In view of Exercises 1 and 2, how would you justify the "price taking" assumption in the construction of the brick industry's supply curve? (*Hint:* What sort of elasticity of demand does a supplier face who takes price as given, nearly unalterable by his actions?)

¹ It is "exactly equivalent" because one way subtracts from the same revenue the area under the marginal cost (which is variable cost) and the other subtracts from revenue the quantity multiplied by the average variable cost (which is also variable cost).

4. Seeking profits is usually considered to be bad. But consider the example in the text of Redfield not looking for profits as much as possible—that is, producing less than the profit-maximizing amount. What is the bad consequence of Redfield's behavior?

PROBLEMS FOR SECTION 13.1

1. "A cost-saving tactic that some hospitals embrace and others approach gingerly is plain old competing to raise occupancy rates. Mainly, of course, one hospital gains at another's expense." W. Daniel Barker, administrator at Crawford W. Long Hospital in Atlanta and chairman-elect of the American Hospital Association, says, "I'm not sure competition does anything for the system. It's like drilling in someone else's oil field." If competition reduces the price of hospitals to customers, is it like drilling in someone else's oil field (fixed in amount)? What about new customers to hospitals as a whole?

- **2.** Suppose that you are the president of Progressive Pen, Inc., which has three branch plants run by Richard Craven, Samuel Farrington, and Robert Paul. The Craven plant has a total cost function $\text{Cost} = 100 + (2/1000)(q^2)$, in which q is the output from the Craven plant and cost is expressed in dollars. Using calculus, the marginal cost from this plant is therefore

$$\begin{aligned}\frac{d\text{Cost}}{dq} &= \frac{d(100)}{dq} + d(2/1000)q^2 \\ &= 2(2/1000)q \\ &= (4/1000)q\end{aligned}$$

The Farrington plant has $\text{Cost} = 200 + (2/1000)(q^2)$, the Paul plant $\text{Cost} = 100 + (1/1000)(q^2)$. The branch plants send their output of pens to you and you in turn sell them in a competitive market at \$1 per pen.

a. Without doing any algebra, what is the condition for the cheapest production of a given output of pens? That is, what in general can you say about it? Given the cheapest production, what is the condition for the profit-maximizing output of pens if pens sell for \$1 each?

b. You could tell the branch plants how much to produce. What would you need to know to decide what to tell them? Notice that it's a lot to know. If you knew it, how much would you tell each to produce (do the algebra)?

c. Suppose that you did not have this knowledge but that the managers of the branch plants did. Suppose that you bring them together and say "Gentlemen, we are going to play a game. I am going to 'pay' you \$1 for every pen you produce. Your job is to maximize 'profits' from each of your plants (handing the money over to me in the end, of course). Go forth and maximize." What outputs will this decentralized system yield?

d. What scarce resource does the decentralized, market-imitating system save?

3. a. What is the total profit earned by Progressive Pen, Inc., at its optimal output?

b. Suppose that you make a mistake, telling the managers that \$1.10 is the "shadow price" they face rather than the true market price of \$1.00. What output will they produce? How much will it cost? (Do it a couple of times: It is easy to make an arithmetical error.) How much will it sell for? What, then, is the loss in profit if you make the mistake?

c. Describe a procedure for finding the correct shadow price when you do not know what it is but *do* know what profit is earned.

- **4.** The median number of slaves per slave-holding farm in the South rose from 20.6 in 1850 to 23.0 in 1860. The distributions by size of holdings were as follows:

Number of Slaves per Holding	Percentage of All Slaves on Holdings of This Size		Increases or Decreases as a Percentage of 1850
	In 1850	In 1860	
1–9	26.6	25.6	−3.8
10–19	22.8	21.6	−5.3
20–49	29.0	27.6	−4.8
50–99	13.1	14.9	+14.0
100–199	6.3	7.6	+21.0
200–299	1.3	1.4	+7.0
Over 300	0.9	1.0	+10.0
All sizes	100.0	100.0	

What were the unsuccessful sizes? The successful sizes? With this evidence, what curve of average total cost would you draw for 1850 over the range of sizes of slaves per holding? Were there economies of scale in slave owning? Formulate the general principle of “survivorship” as a test for economies of scale.

True or False

5. If individual brickmaking firms do not face a flat demand curve, there is no individual or industry supply curve.

13.2 Industry Supply with a Fixed Number of Firms but Costs Interdependent: Externalities

What to Read For

What causes the prices of specialized inputs for the petroleum industry to rise? Does the increase of production by one firm do it? How does the simultaneous increase in production by all firms reinforce the law of upward-sloping supply? What happens when the price of specialized inputs falls as the industry's output rises? Can the price fall offset the forces of upward-sloping supply? What is wrong with the idea that the market will fail to exploit the few downward-sloping supply curves that exist?

Inelastic Supplies of Inputs Make the Supply Curve Even Steeper

The argument has so far supposed without comment that the cost curves of the firms are independent of each other and can therefore be added up. In certain simple and common cases this is true. The cost curves are truly independent whenever one firm can expand or contract without affecting in the slightest the costs of another firm. One producer of jet fuel can expand without raising

costs to other producers; one producer of brass belt buckles can contract without lowering costs to others.

There are cases in which the supply curves are *not* independent, also simple and common. These mostly reinforce the argument that the industry supply curve slopes upward. The cases are those in which the individual firms all draw on a common resource or input supplied inelastically to the industry as a whole. If producers of all petroleum products expand (in contrast to the expansion of jet fuel alone, only one minor product among many), the price of the input, oil, will rise, since it is supplied inelastically *to the industry as a whole*.

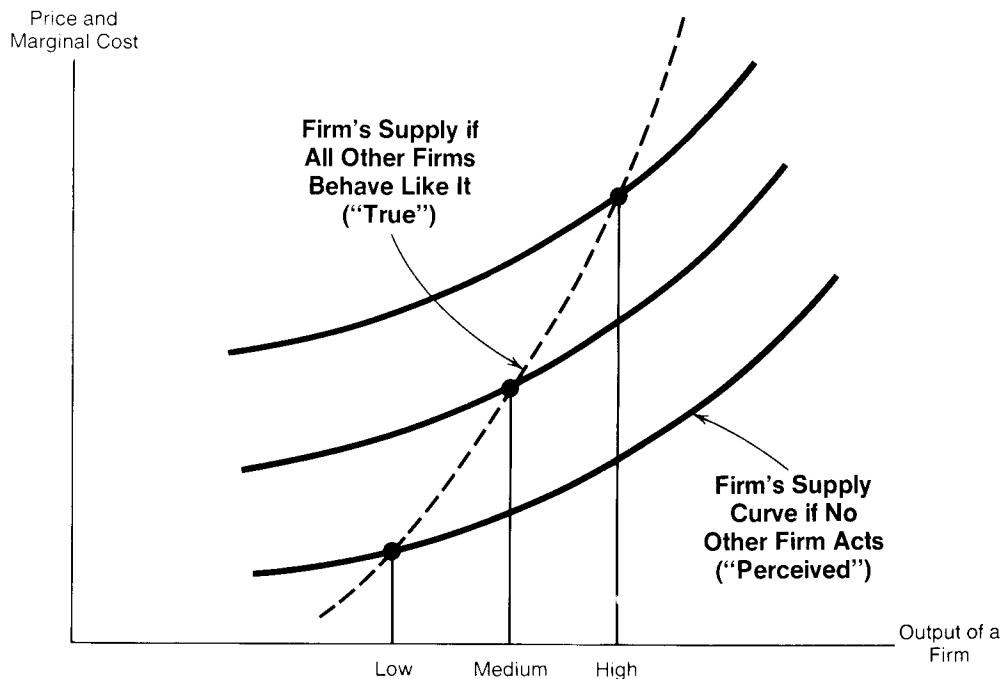
T or F: The elasticity of each ranch's supply curve is normally greater than is the elasticity of the beef growing supply curve.

A: The larger the share of total supply of an input used by a firm (cowboys, pickup trucks) the lower the elastic-

ity of supply of it. The lower elasticity of supply that all ranches face if they expand all at once makes their supply curve less elastic than that of any one ranch if it expands in isolation. If all ranches at the same time expand output, each rancher will perceive the inelasticity of supply as an upward shift in his marginal

Figure 13.5
The Industry Supply Is Less Elastic Because It Cannot So Easily Take Specialized Factors Away from Other Uses

The Perceived supply curve of a single firm, assuming that other firms will hold output constant, is more elastic than the true supply curve. The True supply curve reflects the fact that if one firm increases output in response to an increase in demand, then other firms, experiencing a similar increase in demand, will also increase output. The other firms will bid up the price of inputs used by the industry more than would be the case if they held output constant.



cost curve (see Figure 13.5). The dashed locus is a supply curve of a firm that can be added up with others to yield the industry curve. A firm would in fact view such a curve as the supply curve if it correctly anticipated that all other firms would respond as it did to a change in price. If it believed itself to be brighter or luckier than the others, it would view the flatter curve as its supply, believing that no other firms would re-

spond to the rise in price as it did. The dashed curve is in any event the single firm's share of the market supply, the curve along which the firms must in fact finish if they all rationally pursue profit. No other curve allows correctly for the impact of their combined outputs on the cost of inputs and thence on marginal cost. And it is less elastic. Therefore, true.

Economies of Industry Scale: Partial Equilibrium View

An exception to the law of upward-sloping supply is the unusual case in which a rise in industry output *lowers* rather than raises the cost of an input. A rise in the output of cotton textiles, for example, might lower the cost of textile designers or humidifiers. It would do so only if these inputs had themselves downward-sloping instead of upward-sloping supply curves, perhaps because of strong economies of scale in the schooling of designers or the making of humidifiers. Such a happy event—a case of the more the merrier—is called a *pecuniary external economy of industry scale*. It is “pecuniary” because it affects money costs of inputs instead of affecting the production function directly. It is “external” because it runs between firms instead of within one. It is an “economy” because it is cost *reducing* (as opposed to a *diseconomy*, which is cost increasing). And it refers to “industry scale” because it arises from the rise in the output of the whole industry, not of merely one firm.

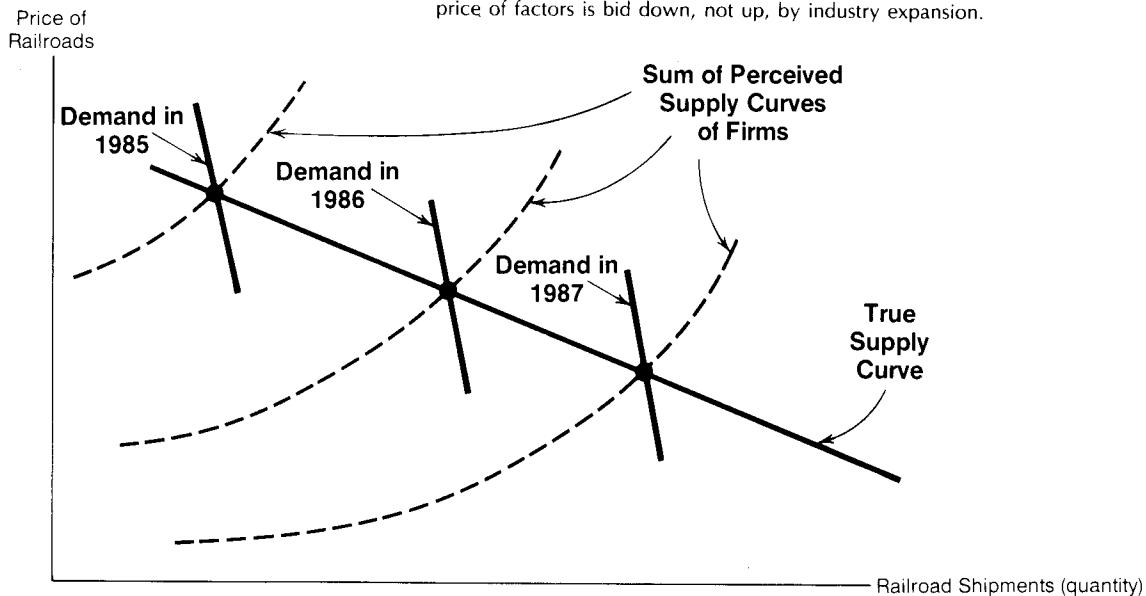
Just as an upward-sloping supply curve of iron ore has the effect, as it were, of contaminating every product using iron with the upward slope, likewise a downward-sloping curve of, say, railroad engineering contaminates with a downward slope every product using railroad engineering (most notably, railroads themselves). In the extreme the cost curve of each railway might fall rather than rise as the output of all railroads rose, and the locus could slope downward (see Figure 13.6). Notice the demand curves in successive years 1985, 1986, and 1987. The outward march of the demand curve cheapens the product. By demanding more railroads, one gets more and cheaper railroads. Ask, and it shall be given you.

Such external economies between industries, and the economies of scale within an industry that give rise to them, are considered to be among the major failures of *laissez-faire* capitalism. They justify proposals to supplement or replace capitalism by subsidies or central planning. The market is said to be unable to reap the advantages. This is said to be particularly true in poor countries lacking “infrastructure” or “social overhead capital” in the form of railroads, schools, distribution networks, irrigation systems, telephones, and the like, such as America in 1837 or Eastern Europe in 1943 or Latin America in 1959.²

² The origin of this view is Tibor Scitovsky, “Two Concepts of External Economies,” *Journal of Political Economy* 17 (April 1954): 143–151, Section II, heading (b), which is an extension of P.N. Rosenstein-Rodan, “Problems of Industrialization in Eastern and South-Eastern Europe,” *Economic Journal* 53 (June–September 1943): 202–211. The idea was extended further by Hollis B. Chenery, “The Interdependence of Investment Decisions,” in M. Abramowitz et al., *The Allocation of Economic Resources* (Stanford, Calif.: Stanford University Press, 1959), pp. 82–110. The essays by Scitovsky and Chenery are reprinted in K. J. Arrow and T. Scitovsky, eds., for the American Economic Association, *Readings in Welfare Economics* (Homewood, Ill.: Irwin, 1969).

Figure 13.6
Powerful External Economies of Industry Scale

The True Supply Curve of a firm can be more elastic than the Perceived Supply Curve, and can even be downward sloping, if there are sufficiently strong economies of scale in the production of one or more "factors" (inputs) used by the industry. In such a case the price of factors is bid down, not up, by industry expansion.



The Case Against Intervention

As a logical possibility, the main concern here, the argument has merit. It should be pointed out, however, that even on a logical level the notion has a flaw. It *assumes* that the market cannot see the opportunity. It assumes, as the economic historian responsible for one of the few serious factual inquiries into the matter put it, "not only the absence of perfect foresight, but virtual exclusion of foresight altogether."³ It assumes that investors in railroads, say, look fixedly at presently existing prices, not expected future prices, an assumption that ignores the forward-looking nature of economic rationality. But, to turn to the facts in the American case,

expectations do exist . . . and in the instance of American railroad investment [before the Civil War], they were typically optimistic. . . . The first annual report of the Norwich and Worcester Railroad [in 1837] . . . recognizes that "in regard to transportation of merchandise, etc., there is abundant reason to be satisfied that it will equal our highest anticipations, and *like all similar undertakings greatly increase the business upon its line.*" . . . America before the Civil War thus did not suffer from a deficiency of railroad investment despite the private nature of its provision.⁴

³ Albert Fishlow, *American Railroads and the Transformation of the Ante-bellum Economy* (Cambridge, Mass.: Harvard University Press, 1965), p. 309.

⁴ *Ibid.*, pp. 309–310.

One of the few factual studies of the matter suggests that the premise of nearsightedness may be false.

It remains to be seen whether the presumption in favor of central planning in the case of economies of scale will survive other factual studies. The theorizing thus far has been built on few facts. For example, the prior fact of the existence of economies of scale at the level of an industry is often supported by arguments such as the following.

T or F: Since we see that industries that grow fast are commonly industries with falling prices, economies of scale at the level of industries must be common, and rising demand will cause prices to fall. (*Hint:* Suppose that the fall in prices were caused by the cost curves shifting downward.)

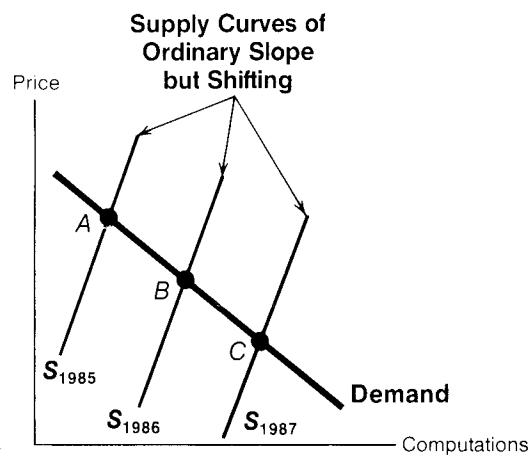
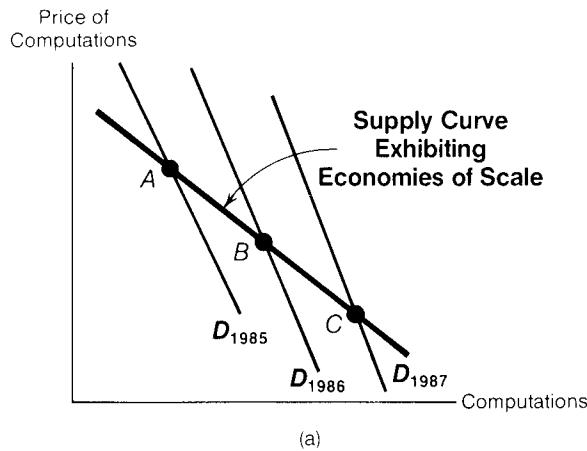
A: Taking the hint, if cost curves fall because of improvements unrelated to industry scale (wider use of steam engines in nineteenth-century factories or of electric typewriters in twentieth-century offices), the lower price will cause the quantity demanded to increase. In other words, fast-growing industries such as the computer industry will have falling prices because falling prices of computation cause fast growth, not necessarily

because fast growth causes falling prices. Therefore, false. Both directions of cause are possible (see Figure 13.7). The three observed points *A*, *B*, and *C*, in Figure 13.7 are the same in both cases, requiring one to bring other evidence (another example of the "identification problem" mentioned in Section 6.1). The extent of rise required in the demand curve, for example, is often too large to fit known income elasticities and rises of income. In such a case one would find in favor of the second panel, rejecting economies of industry scale. The great importance of economies of industry scale in the ruminations of English and American economists over the past 60 years should not be taken as a measure of their importance in the world.

Figure 13.7

The Evidence for Economies of Scale Is Inconclusive: Fast Growth (Shifts in Demand) Causes Falling Prices and Economies of Scale (a) Whereas Falling Prices (Shifts in Supply) Cause Fast Growth and No Economies of Scale (b)

Price often declines in rapidly growing industries. This can be caused by economies of scale resulting in a downward-sloping supply curve or by technological progress that shifts the supply curve outward. You can't tell easily which is which.



Summary The cost curves of firms in an industry are not independent if some input is supplied inelastically to the industry as a whole. The expansion of all firms together will in such a case raise costs more than will a similar expansion in one firm. In other words, the industry supply curve is less elastic than the individual marginal cost curves, the effect being more pronounced the more important is the inelastic input to costs. Normally, then, allowing for interdependence reinforces the conclusion that supply curves slope upward.

Pecuniary economies of industry scale is an exception. If the supply curve of an input into railroading slopes downward, not upward, it may transmit the downward slope to the supply curve for railroading itself. Downward slope spoils the optimality of the market. Since the reduction in costs is external to each firm, no firm takes account of the effect to its own expansion in reducing the cost of the input. Too little is produced, and there is room for government intervention to encourage balanced growth. Yet there is room for doubt that this leading objection to market capitalism is factually important.

EXERCISES FOR SECTION 13.2

1. Oil is used for things other than gasoline; it's used for plastics, heating oil, and so forth. If oil were *specialized* in gasoline making (that is, if gasoline were its only use), how would an increase in the demand for gasoline affect the price of oil? What if oil was heavily but not completely specialized?
2. If oil were itself *elastically* supplied—that is, if more could be pumped out of the ground or out of large reserves at a nonrising marginal cost—would the reasoning of Exercise 1 be true?
3. Oil is a big input for gasoline. If it were small—that is, a small share of the total cost of gasoline—would a rise in its price cause the price of gasoline to rise much?
4. In view of Exercises 1, 2, and 3, describe the sort of input (for instance, oil) that would contribute a lot to making the supply curve of some output (for instance, gasoline) inelastic.
5. Name an input for each of the following products that might cause its supply curve to slope upward *more sharply* than the summed marginal cost curves of the individual firms. (*Hint:* Look at Figure 13.4 again; and your answer to Exercise 4.)
 - a. Ground beef.
 - b. Milk.
 - c. Cut stone, such as ornamental stone for buildings.
 - d. Paper.
 - e. Economics teaching.
 - f. Professional football.

PROBLEMS FOR SECTION 13.2

1. Since 1885 many observers of British railways have believed that coal cars in Britain are too small. Half the size of American cars, they are not owned by the railways themselves but by the coal mines. To make larger cars worthwhile would have taken large investment in loading equipment and sidings, and the railways (not the coal mines) owned the loading equipment and sidings. The division of ownership is similar to the case of pecuniary externali-

Persistent
behavior has
a reason
The roots
around us.

ties. It is alleged to have the same result: The "interrelatedness" results in less than optimal investment, such as less than optimal investment in large coal cars.⁵

a. Supposing that the larger coal cars were in fact socially desirable, what could the railways do by way of buying up coal cars that would eliminate the "externality"?

b. Short of this, are there bribes by which the railway could induce the coal mines to replace old cars with new ones at the socially optimal time?

c. By 1947 both railways and coal mines in Britain had been seized by the state. Still supposing that the larger cars were desirable, would this nationalization solve the problem? What do you make of the fact that at present the coal cars are *still* small?

True or False

- o 2. Suppose that there is an industry whose output fluctuates more over the business cycle than does a number we shall call K , defined as the fluctuation in national income multiplied by the income elasticity of demand for the product; the industry probably experiences economies of industry scale.

⁵ Charles P. Kindleberger, *Economic Growth in France and Britain, 1851-1950* (Cambridge, Mass.: Harvard University Press, 1964), pp. 141-145.

The Long-Run Supply Curve and the Principle of Entry

14.1 The Long-Run Supply Curve

What to Read For Can an “industry” be defined by the techniques it uses to make its products? Are railways “an industry”? Which suppliers are in it? What about potential suppliers? What is a *normal return*? What is an *entrepreneur*? What is an *entry price*? If an entrepreneur is getting a price higher than his entry price, what is the excess called? What is *economic rent*? The margin of cultivation?

The Mere Definition of an Industry Can Affect One's Argument The analysis of the industry given so far starts to unravel if you ask insistently, “What is the industry?” For example, what is the industry in which a railway competes? The answer seems obvious: track and wheel, locomotives and cars, stations and tunnels, executives and porters. For purposes of a book on railway engineering or of a discussion of make-work rules by railway unions or of agitation in Congress for taxes on trucks this technological definition makes sense. But for purposes of analyzing the market for the services of railways, the technological definition misses the point entirely. The Norfolk and Western Railroad and the Santa Fe Railroad do not supply substitutes: They are 2000 miles apart, supplying completely different customers. No single customer faces a mutually exclusive choice between shipping cattle from Norfolk to Roanoke by one railroad or from Santa Fe to Albuquerque by another. And looking at the matter from the railways’ point of view, within a technologically defined “industry” the parts in different locations will have different interests.

Q: Like the airlines today, railways in the late nineteenth century divided into eastern and western companies

at Chicago. There were a half-dozen companies from the East Coast to Chicago and about ten from Chicago

westward. The Interstate Commerce Commission (ICC) was formed in 1887 to regulate this commerce, which it continues to do. The ICC says that it acts on behalf of the general public, but most economists and historians believe that it was in fact captured by the railways soon after its formation, if not actually created by them, and has been used since then to raise rather than lower rates. The simplest version of the "capture theory" ignores the geographical divisions within railways, arguing that the ICC is a simple conspiracy of all railways against the public.

1. Were the eastern and western companies in the same "industry"? How would a ton of wheat get from Iowa City to New York?

2. A ton of wheat cost \$6 in 1890 to transport from Iowa City through Chicago to New York. All railways would clearly like to have seen the rate be higher than \$6. But given a \$6 cost in total, why was it in the interest of eastern railroads to get the ICC to reduce western rates?

3. Not all goods transported to or from Iowa City came from or went to New York: Some came from or went to Chicago. Why would this fact have made western (for instance, Iowan) customers of railways also enthusiastic about reducing western rates?

4. Restate the capture theory in a form that takes account of the two different industries making up "the" railroad industry.

A: 1. No. The two groups did not compete in the same market. They carried the same ton of wheat, but at different stages of its travels. A ton of wheat would travel on one of the western railroads (on the Rock Island as it happens) to Chicago and then to New York on one of the eastern railroads, say, the Pennsylvania Railroad.

2. If the Pennsylvania could get the ICC to reduce rates on the Rock Island, the Pennsylvania would receive a larger share of the \$6 total.

3. Since some of their business was not through to New York, the *customers* of the Rock Island would join the Pennsylvania in appealing for lower western rates. In fact they did, arguing that the high rates for the Iowa City to Chicago leg of a journey constituted unfair discrimination.

4. The ICC can be viewed in 1890 as a conspiracy of eastern railways allied with western customers against the western railways and eastern customers. It was not "the" railway industry against customers, but a blend.¹

How to Choose the Definition of an Industry Wisely: First, Substitutability

The trivial-looking task of defining the industry, then, is difficult and important. One wants a workable definition of the industry in which railroads compete that means "all suppliers in the same market." Therefore one must pick a market such as "the transport industry from Norfolk to Roanoke," including in it not only the Virginia Railroad and Norfolk and Western Railroad, but also the buses, trucks, autos, airplanes, barges, ships, and mule trains competing with the railroads over the route. If one defines the *people transport industry* as "automobile-assembling companies, most of whose stockholders are Americans," then the automobile industry will appear to be nearly a monopoly, with four major companies, one much larger than the other three (General Motors, Ford, Chrysler, and American). But if one recognizes that Toyota, Volkswagen, Honda, Mazda, Subaru, Fiat, Peugeot, Mercedes-Benz, Renault, and others compete with the Big Four in some markets and that public transport, motorcycles, bicycles, and shoe leather compete with them in others, it becomes less obvious. It is not so obvious that the people transport industry is a case of competition among the few, by which the few can rob their customers at will.

T or F: Since the U.S. Postal Service has a legal monopoly of first-class mail, people wishing to communicate be-

tween San Francisco and Los Angeles are at the mercy of the service.

¹ So argues David Haddock in an unpublished doctoral dissertation at the University of Chicago, 1978, entitled "The Regulation of Railroads by Commission."

A: False, if "at the mercy of" means "unable to find any alternative to." The industry is not the mail between the two cities but communication in general,

with all the possibilities available of telephone, parcel post, telex, microwave, telegraph, radio, messenger, and personal travel.

The decision whether or not to include the telephone in the same industry with the mail depends on the size of their cross elasticities of demand and on the demanders one wishes to emphasize. Cross elasticity, you recall, measures how sensitive telephone consumption is to the price of mail. That is, it measures *substitutability in consumption*. Business may well view a telephone call or a typed letter as close substitutes for many purposes. The courts, however, are required by law to do their official business by letter; and some lovers prefer the phone to the mail. The magnitude of cross elasticity one chooses will also vary from question to question. The only guide is consistency. If you bundle together the varied products of "the housing industry" for some question, you must admit a similar level of variety, and similar cross elasticities of demand, when asking a similar question about another part of the economy.

**How to Choose the Definition of an Industry Wisely:
Second, Profitable Supply**

The root definition of industry is "all firms that *can profitably supply* goods some group of consumers view as *close substitutes*." *Close substitutes* is a feature of the utility functions of consumers. "*Can profitably supply*," it will now be shown, is a feature of the cost curves of firms. The industry contains everyone who can make money at the going price. *At a higher going price more will enter.*

Q: Tell what is wrong with the following argument used to explain the cheapening of cotton textiles in the early nineteenth century: "The expansion of demand for cotton textiles allowed each firm to expand, to spread fixed costs over a larger output, and therefore to lower prices."

A: Everything is wrong. To take the last point first. If the fixed costs were in fact fixed over such a long period, then they would not determine price, since fixed cost is no part of what does determine price, namely marginal cost. Furthermore, marginal cost for each firm must have been upward sloping, not downward sloping. If it were downward sloping, then each firm could have profited more by expanding at once, which is to say that the firms would not be in equilibrium at the earlier, small output. The argument needs at the least a supplement explaining why firms did not seize an opportunity for profit in front of their noses. Equilibrium and there-

fore an upward slope of marginal cost would imply in fact that expansion of the industry would have raised, not lowered, cost. The most important error, however, is the first, "the expansion of the industry allowed each firm to expand." This would be true only if the firms in the industry were unchangeable, that is, only if the expanded quantity had to be shared among the *existing* firms. But in the long run it does not have to be shared. At the higher price from expanded demand, additional firms will enter (and at a lower price exit). The size of the whole industry does not fix the size of firms within it, only the number of firms. A permanent rise in the demand for cotton textiles increases the number (not the size) of firms, each of optimal size. So too a permanent rise in the attendance at Busch Stadium increases the number of ice cream vendors (not their daily capacity) and a permanent fall in the number of newspaper readers reduces the number of newspapers (not the circulation of each).

The long-run supply curve, in other words, includes *potential* entrants. The point is the same as that involved in defining the railroad industry, for in the business of transporting people from Boston to New York the buses, airplanes, and autos are all potential or actual competitors with the New Haven Railroad. The correct supply curve of an industry adds up horizontally the supply curves of everyone who can at the various prices profitably supply the good.

The Meaning of Normal Profit

The remaining question is simply who can profitably supply the good—say, rental housing for students. The answer is, anyone who can cover long-run total cost. The answer is straightforward, but it contains one piece of fancy footwork that should be noted. The definition of “cost” of rental housing for students includes of course all the costs of materials such as gas for heating and of labor for janitors. Slightly less obviously, it includes the cost of capital. The cost of capital is the interest cost on the loan taken out to buy the building *plus* the annual cost of replacing it as it wears out under the constant assault of late-night parties and the weather *plus* the expected annual capital loss from a fall in price of a building of given quality (or if a rise in price, minus the expected capital gain). Least obviously of all—and this is the fancy footwork—it includes the “normal profit” or “normal return” to taking the bother, knowing the market, seeing the opportunity, assuming the risk. In a word, it includes the normal return to the “entrepreneur.”

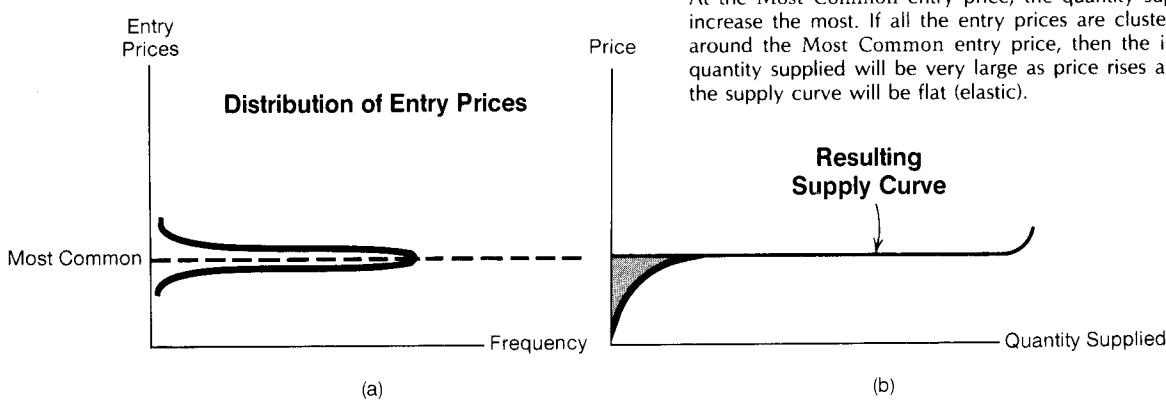
The word *entrepreneur* is French, meaning “contractor” in the American sense of “building contractor,” that is, someone who takes the considerable trouble of bringing together the carpenters, lumber, painters, and paint to remodel your house. The entrepreneur (or, if female, *entrepreneuse*) coordinates the other factors of production. Since she could hire a manager as one hires a janitor, it is not routine management that is her skill. Indeed, entrepreneurship could be defined as all inputs that cannot be hired. If she can buy insurance against some hazard in the business, for example, she is hiring “security,” and only uninsurable risks remain with her. The buck stops with her.

It is easy to point to entrepreneurs in this sense, for they are the heroes of our economic mythology: Eli Whitney, Thomas Edison, John D. Rockefeller, Andrew Carnegie, Henry Ford, Edwin (Polaroid) Land. It is much less easy to point to a line in the actual income statement of a real firm, or in the tax returns of the leaders of the firm, and say, “Ah, I see that the price of entrepreneurship has gone up this year.” The income of an Edwin Land will include income as a normal research scientist and as a living advertisement for Polaroid, but mixed into it will be his income for the qualities of brains and energy that make him one of the company’s entrepreneurs. The reason one needs a line for “entrepreneurship” in the hypothetical accounts of a firm is precisely because the brains and energy are scarce and have alternative employment. If entrepreneurship is not rewarded at Polaroid, it will leave to join Kodak, or to join another industry entirely. The normal amount of entrepreneurship is rewarded at the normal rate and is included in normal costs. Building contractors are not earning “profits,” really, when their price is above the costs of hired factors. Given the free entry to and exit from the industry, it is reasonable to suppose that they are earning merely normal profit, that is, the return to entrepreneurship necessary to keep it in the industry. The supply curve, in short, includes normal profit.

The Distribution of Entry Price Determines the Shape of the Supply Curve

The final step in allowing for entry is to add up the output supplied at each price. When the expected, long-run price goes up, more firms find that price exceeds their long-run average costs (including normal profit). They enter, and output jumps up. If all the potential entrants are clustered around one entry price (as they might well be if the skills required to enter the industry are widespread), the resulting supply curve at that price is elastic (see Figure 14.1).

Figure 14.1
An Industry That Anyone Can Enter Easily at the Same Price (a) Has a Flat Supply Curve (b)



At the Most Common entry price, the quantity supplied will increase the most. If all the entry prices are clustered tightly around the Most Common entry price, then the increase in quantity supplied will be very large as price rises a little, and the supply curve will be flat (elastic).

A case in point would be the restaurant industry. Because the skills of purchasing, cooking, and serving food are widespread (that is, themselves elastically supplied), an increase in demand for ordinary restaurants will attract an increase in their number at roughly constant cost. Almost all restaurants at the equilibrium of supply and demand would be earning virtually zero economic profits (that is, they would be earning nothing over normal profits), because almost all would have entered (thus revealing their own point of zero profit) at or very slightly below the prevailing price. *The supply curve for such an industry, with identical cost curves for all firms and no interdependence in the curves, is flat.* It is a common case.

Producers' Surplus Is Economic Rent Is Quasi-rent Is Supernormal Profit

Notice in the diagram, however, the little shaded area. It is the sum of all the excess of the actual price over the minimum price necessary to bring forth various different supplies. In other words, some of the firms are earning profit above normal profit. The profit is called in this context *rent* (or *economic rent* to distinguish it from a "rent" payment to a landlord, which may or may not be a true economic rent). If the profits are temporary, it is called *quasi-rent*.

It is larger when the entry prices are more varied.

Q: Suppose that there are only two kinds of oil land, Persian Gulf with a very low marginal cost curve and entry price and North Sea with a very high marginal cost curve and entry price. View the two as two firms, each of which (however) takes world price as given.

1. If world demand for oil is such that the price is exactly at the Persian Gulf Entry Price, what is the amount supplied? If the price fell 1 cent below the Entry Price in the long run, what would happen?

2. If the price is above the Persian Gulf Entry Price but *below* the North Sea Entry Price, how is the amount

of supply determined? Draw the diagram and show the rent earned at some price.

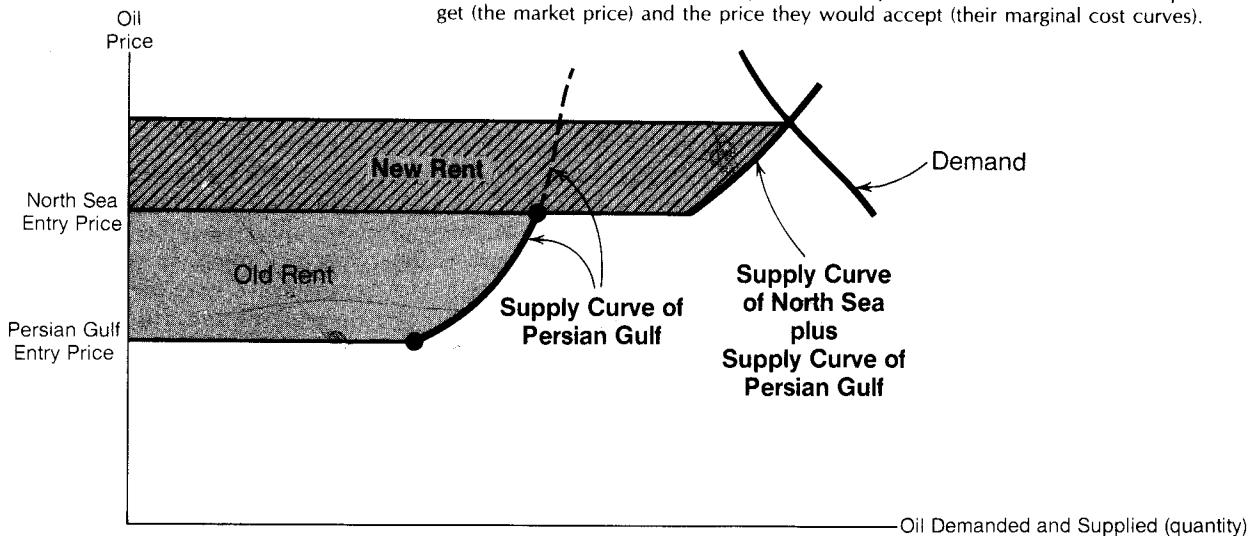
3. If the price is slightly above the North Sea Entry Price, what happens? If it is still higher, what are the areas of rent?

4. What, therefore, is the area above a supply curve, the area called producer's surplus in earlier chapters?

A: 1. Consider Figure 14.2. If the price is at the Persian Gulf Entry Price, the amount supplied is the amount on the Persian Gulf's marginal cost curve at the entry

Figure 14.2
An Industry That Few Can Enter at a Low Price Has a Steep Supply Curve and Much Rent

Because firms will enter only if it is profitable to do so, they will earn more than the minimum price at which they would enter. That is, they will be profitable. The profits—also called rents or quasi-rents or producers' surplus—are the area between the price they get (the market price) and the price they would accept (their marginal cost curves).



price, that is, at the zero-profit (zero-economic-rent) point. Evidently, then, rent is zero. If the price fell 1 cent below the entry price, the Persian Gulf would exit the oil industry, at least in the long run. World output of oil would go to zero.

2. Since the North Sea is still out of the picture, the supply curve when the price is between the two entry prices is simply the Persian Gulf's marginal cost curve. In Figure 14.2 it is the marginal cost curve between the Persian Gulf Entry Price and the North Sea Entry Price. The rent is the area behind the marginal cost curve, being all the profit in excess of zero profit. If the price were 1 cent below the North Sea Entry Price, the Persian Gulf would supply out to that point and the rent would be the area Old Rent. The North Sea would be beyond the "margin of cultivation" (margin in this phrase means "edge" or "outer limit"). At some

lower price the quantity supplied and the rent would be the lightly shaded area.

3. If the price is slightly above the North Sea Entry Price, the North Sea enters and quantity supplied jumps out. If it is still higher, the area of rent is the area New Rent plus Old Rent.

4. The area above a supply curve is the sum of economic rents earned in the industry, justifying the notion of producers' surplus used in earlier chapters. The supply curve in the right-hand panel is a scalloped curve because there are only two types of cost curves. If one generalized the argument to many types, ranging from Saudi Arabian through Venezuelan to North Sea oil, the scallops would look smoother and the supply curve more conventional. In any event the producers' surplus is the supernormal profit which is the economic rent.²

² By defining the rent as costs paid to the owners of the scarce resources earning them (such as very fertile oil land), one could eliminate the rent and declare all cost curves flat. Competition among owners of firms would result in their paying over to landlords all their "profits." This maneuver, which protects the zero-profit condition of equilibrium from any conceivable criticism, is sometimes performed in theoretical works. It lacks point, for the rent is only known as what is left over from revenues when the other (opportunity) costs have been paid. Rent is a residual. To define it as a cost would merely encourage one to think that all cost curves are in fact flat when they have been made "flat" merely by definition.

The argument returns in the end to its beginning. The supply curve is the sum of each firm's marginal cost, with the condition that price be above the point of zero long-run profit. It is this condition that forces one's gaze outward, from the existing firms toward all potential entrants. What one sees when one gazes outward, as the next section will show, are the most typically economic pieces of reasoning.

Summary

The industry cannot for many purposes be defined as the existing firms or as the firms sharing a certain technology of supply. It must be defined as all potentially profitable firms sharing a set of consumers who view the products as substitutes. An industry such as American railroads or rental housing violates the definition by including too much in geography and too little in variety of product. The Boston and Maine Railroad does not share customers with the Chicago and Northwestern Railroad, but it does share customers with Eastern Airlines and the Maine Turnpike. Furthermore, an industry must include all firms that can share the customers profitably, whether presently in existence or not. The existing firms grow and shrink in number (not necessarily in size) as the demand grows and shrinks. As the price rises above and falls below the point of minimum average total cost including normal profit, the firms enter or exit. When all firms happen to have the same entry price the resulting supply curve is flat. When they do not have the same entry price the supply curve slopes upward, and the firms inside the margin (the edge) of cultivation earn economic rents.

EXERCISES FOR SECTION 14.1

1. Decide whether the following are industries. Name the sorts of firms who compete for the same customers (remembering that the only guide is consistency with other uses of "industry," such as the "people transport industry"):
 - a. Movie theaters, TV, Betamax.
 - b. Universities.
 - c. Grocery stores in the United States.
 - d. American producers of bicycles.
2. Is the manager of a store in a grocery chain an entrepreneur? Is the banker who provides loans to run the business? Is the food technologist hired to invent new ways of packaging meat?
3. Draw the supply curve of typists in Madison, Wisconsin, that corresponds to the following distribution of "entry prices" into the "industry" of typing:

At a Price (Wage) per Hour of:	This Many Enter the Industry:
\$3	100
4	1,000
5	10,000
6	1,000
7	1,000

What shape does it have?

4. What shape would the curve of Exercise 3 have if the distribution were

At a Price (Wage) per Hour of:	This Many Enter the Industry:
\$3	100
4	1,000
5	10,000
6	10,000
7	10,000
8	10,000

5. Suppose that the going price of typists is \$5 an hour. What do you call the difference between \$5 an hour and the \$3 an hour at which some typists would nonetheless enter?

PROBLEMS FOR SECTION 14.1

1. In explaining why British steel companies before World War I were slower than German or American companies to introduce large-scale plants, one historian wrote, "Great new plants could emerge in Germany and the States without other plants being stationary, let alone shrinking or disappearing. . . . German and American markets had far larger and more expansive home markets."³ Comment.

- o 2. Overheard at a restaurant in Paris at the height of the battle of Verdun in 1916: "Do you know their latest demand? That the market [for arms] should be closed not only to middlemen, but to all businesses started since the war—yes, to all those manufacturers who were not operating in the same branch of industry before 1914!"⁴ Supposing that the demand for arms were completely inelastic, would the closure reduce or raise the profits of arms manufacturers generally? Is closure an effective policy against profiteering on arms?
- o 3. The attempt by the British after 1807 to stop the slave trade from Africa to the New World has been viewed by historians as noble but futile. The reason they believe it futile is that the main instrument of British policy in this matter was to use the Royal Navy to seize slave ships off the coast of Africa and to confiscate their cargoes (to free the slaves on the ship seized). Yet of the nearly 2 million slaves imported into the New World from 1811 to 1870, the Royal Navy intercepted only 160,000, or only 8%.
 - a. Is the smallness of this figure persuasive evidence of the futility of Britain's policy?
 - b. How would you go about estimating in a supply and demand diagram the true effect on the number of slaves imported? (Hint: Ask yourself how many slaves would have been imported in the absence of a chance of a ship's being seized.)⁵

³ D. L. Burn, *The Economic History of Steelmaking, 1867–1939* (Cambridge: Cambridge University Press, 1961), p. 240.

⁴ Jules Romains, *Verdun*, trans. Gerard Hopkins (New York: Knopf, 1939), p. 384.

⁵ The analysis is from J. Phillip LeVeen, "A Quantitative Analysis of the Impact of British Suppression Policies on the Volume of the Nineteenth-Century Atlantic Slave Trade," in Stanley Engerman and Eugene Genovese, eds., *Race and Slavery in the Western Hemisphere: Quantitative Studies* (Princeton, N.J.: Princeton University Press, 1975).

Unintended
consequences.
Counterfactuals.
Superficial story
lies

- 4. The Bureau of Land Management rents public grazing land in the West to the same few ranchers year in and year out. The rent is low and secure—so low and secure that when a ranch with public grazing rights is sold the bureau offers the same rights to the new owner. Despite this subsidy to the lucky ranchers, an inquiry by the National Association of Feedlot Operators into alleged excess profits in ranching found no difference in return between subsidized and unsubsidized ranches. Why?
- 5. One observes that steamboats on the Mississippi before the Civil War earned in any one year widely differing rates of return. *True or false:* This observation is inconsistent with an assumption that steamboating was competitive and had free entry.
- 6. A tariff is imposed on shoes into the United States. The shoe industry, composed of many small firms, has a constant-returns-to-scale production function. The tariff, therefore, will increase the size of the shoe industry, but not the rents to factors employed in it. *True or false:* In light of this fact, it is difficult to see why members of Congress from shoe-producing districts support such a tariff.

True or False

7. As the number of children in Philadelphia falls, the number of children per day-care center will fall.

14.2 The Uses of Entry and Exit

What to Read for	What happens when there is an opportunity to earn supernormal profits? Can supernormal profits be earned in the long run? Do supplies of expensive goods like diamonds make a lot of money? Should cab drivers support higher cab fares, considering that the drivers are hired by cab owners? What is the "principle of entry"? The "Tiebout effect"? "The American Question"? Why are many American communities segregated by tastes and income? Do these ideas apply only to industries, towns, and stock markets? What is the principle of "rational expectations"?
-------------------------	--

In the Long Run, Profits Are Normal

The idea that firms will enter at the smell of supernormal profit is among the half-dozen leading ideas in economics, ranking with the budget line, the utility function, the Rule of Rational Life, equilibrium, the production function, and the one competing among many. If Americans suddenly develop a taste for carbonated water the Perrier Company will suddenly become rich. But in the long run its riches will be dissipated by the entry of others, and the many new makers of carbonated water will earn only normal profits. If Chicagoans develop a distaste for living in the center of the city the owners of property there will become poor. But in the long run their poverty will be alleviated by exit, and the few surviving landlords will earn normal profits. The assertion is that in the long run profits are normal, and in many shorter runs they are on their way to becoming normal. Notice that the assertion is not that super- or subnormal profits never happen. They do. That in the long run the weather is summer and that from December 22 onward it is on its way to becoming summer does not mean that winter is a mirage. But the tendency is nonetheless worth keeping in mind.

The principle of normal profits can be viewed as a suggestion for analysis, like the suggestion in the analysis of supply and demand that one look hard at equilibrium or the suggestion in the analysis of a maximizing individual that one look hard for marginal costs and benefits. *A piece of economic analysis is not complete until everyone is earning only normal profits*, or at least until the analyst has identified a reason why not.

T or F: If people have a great demand for diamonds, diamond sellers will earn much money.

whenever there is supposed to be supernormal profit there should be doubt.

A: If entry to diamond selling is free (that is, not obstructed by law), then the sellers will earn only normal profits in the long run. The only doubt is whether the long run obtains. If the "great demand" is of long standing, then the entry of sellers seeking profit will already have taken place. Unless it is a recent development, the intensity of demand for a product (or its essentialness or basic-ness or whatever) is irrelevant to its profitability. In short, false. To put the principle another way,

T or F: Since a patent on an invention makes the owner the only seller of it, the owner can charge much and therefore earn more than a normal return on investments in making inventions.

A: Apply doubt. If there were supernormal profits, then potential inventors would enter the inventing game until the return was normal. False.

Normal Profits in Various "Industries"

The principle does not apply only to literal firms. A license to operate a taxicab in most cities is very expensive, a mere pittance at the city licensing office to be sure, but \$30,000 in the open market because the city does not issue new licenses when the demand goes up. The actual drivers of the cab, therefore, are not usually the owner of the license, or even of the cab itself. They are college students, moonlighting firefighters, ambitious immigrants. The city sets the fares, raising them from time to time at the request of the owners. The owners are often able to recruit the drivers to the cause.

T or F: Although they are sometimes tricked into supporting it the drivers, as distinct from the owners, have no interest in a higher fare.

of the drivers in other occupations. It cannot rise above the return in other occupations without attracting more drivers to, uh, drive it back down. Drivers get the average hard-working-but-unskilled wage no matter what the fare is. In fact, they will get it whether or not the cab industry exists. So, true.

A: There is free entry to *driving*, as distinct from owning. The return to driving, therefore, is pegged to the return

A spectacular example of the principle of entry in action is entry to the industry of crime. The logic can be used to count the cost. The cost of crimes such as arson and vandalism that destroy property is straightforward, namely, the value of the property destroyed. A building burned down for fun or money is lost to the uses of society. But at first the cost of crimes such as burglary, robbery, fraud, extortion, counterfeiting, and blackmail that *transfer* property from one person to another appears quite different. After the theft, Al Capone and Bugsy Moran have the property instead of you, but society as a whole (which after all includes Capone and Moran) appears to have no fewer commodities. The value of goods stolen seems a mistaken estimate of the *social cost*.

Q: If there is entry to the theft industry, however, what relation would there be between the value of goods stolen and the value in alternative employment of the

resources employed in theft? What, then, is a good way of estimating the social cost of theft (leaving aside police, locks, and so forth)?

A: The zero-profit condition implies that crudely speaking the value of the burglar's time, the counterfeiter's engraving skills, the robber's boldness, and the con man's guile, as well as the value of the services supplied by related industries (fencing the stolen property, legal advice, and so forth) *must equal the value of the loot*. The equality is only "crudely" true because the loot is usually less valuable to the thieves than to the victims

and because the theft industry might not always be in long-run equilibrium. In any case the skills and other resources used to reallocate property (that is, by theft) are lost to the task of making more property. Crime does pay the individual criminal, but in this case does not pay society. The argument that theft is merely a transfer is true, but the value of the transfer is a reasonable estimate of another, *social cost*.⁶

The eventual equality of revenue and cost, then, is the principle of entry, and is fundamental to thinking economically. It is also fundamental to thinking ecologically: A population of plants or animals grows into an ecological niche until the "profit" is zero, and no more entry is encouraged. So it is with human populations, which inspired economists at the birth of economics to formulate the subsistence theory of wages:

The natural price of labour is that price which is necessary to enable the labourers . . . to subsist and to perpetuate their race, without either increase or diminution. . . . The market price of labour may deviate from its natural price. . . . [But] when . . . by the encouragement which high wages give to the increase of population, the number of labourers is increased, wages again fall to their natural price. . . .⁷

The principle of entry applies to politics as well.

Exit, Voice, and Loyalty: Normal Profits in Politics

Q: Laws, such as laws about jogging on the streets, are pure public goods, enjoyed by all if by any, and supplied collectively. Among any fixed group of people, there will be disagreements over what laws are best, disagreements that simple majority voting, as we have seen, will not solve. Suppose, for example, that the cities of Wakefield and Reading are neighbors and have initially identical populations, *exactly* half who love jogging, half who detest it. Ordinances either to turn the public roads into jogging paths or to instruct the police to shoot joggers on sight are rejected in both communities. Jogging is tolerated but not encouraged. Now suppose that Francis McGrath of Wakefield converts from

detestation to love of jogging, George Evans of Reading the other way.

1. If people cannot move their residence from one city to another, what happens to the proposed ordinances and to the (new) minorities against whom the ordinances are directed?

2. If the minorities can and do move to escape the oppression by a majority, what is the final composition and character of each town?

A: 1. Joggers are now a (bare) majority in Wakefield and vote to turn all roads into jogging paths, reducing the (bare) minority who detest jogging to ineffectual rage.

⁶ The analysis is that of Gary Becker, "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76 (March–April 1968): 169–217, note 4. Destructive crime, such as arson, therefore, has a double cost: the goods or lives lost and the resources lost in plotting and executing the loss.

⁷ David Ricardo, *On the Principles of Political Economy and Taxation*, first published in 1817 (Harmondsworth, England: Penguin, 1971), pp. 115–116. A tendency to entry and zero profits, however, is not its achievement. He discusses at length the possibility that "notwithstanding the tendency of wages to conform to their natural rate, their market rate may, in an improving society, for an indefinite period, be constantly above it" (p. 116).

Likewise, joggers are a bare minority in Reading and must dodge bullets if they venture out as they once did.

2. Clearly the situation of (1) is not an equilibrium. There is still "profit" to be made by entry and exit.

And this is the solution to the problem of providing a public good pleasing to all residents: Change the residents. All the joggers will move to Wakefield and all the antijoggers to Reading. The initially identical towns become quite different.

The argument is called the *Tiebout effect*.⁸ It asserts that people can vote about public goods with their feet, moving to communities having the bundle of public goods they want (schools, parks, police protection, jogging ordinances). The problem that people disagree about the amounts of public goods to be provided is solved by moving the people. The result is a segregation by taste and income such as that which characterizes, say, American suburbs, some being rich and vulgar, others poor and refined, others poor and vulgar. It is a consequence to entry to and exit from political communities.⁹

If no one can *enter*, the existing firms continue to enjoy *positive* profits, as do citizens of the United States in view of the barriers to entering the country. If no one can *exit*, the existing firms can experience *negative* profits, as do citizens of the Soviet Union in view of the barriers to exiting the country. Effective barriers are in fact evidence that there is a profit or loss. Barriers to exit are essential for exploitation, which explains why governments such as the German Democratic Republic or South Africa, intent on forcing their subjects to do their will, erect barbed wire and guard towers at their borders.

T or F: If sharecroppers in the South after the Civil War were mobile, no single country store making loans to sharecroppers could exploit them.

A: The "exploited" person under such conditions could leave. The factual question is "How mobile were the

sharecroppers?" An answer to it determines the interpretation of the economic experience of many black Americans after slavery. If the sharecroppers *were* mobile, they were not exploited. Therefore, true.¹⁰

Exit to Avoid Economic Harm

Entry and exit mean that hurts and helps do not always stay where they are placed. The idea is *incidence*, that is, "on whom a burden ends up," which the next chapter will pursue in detail for the case of taxes. For the present, the lesson is to use the zero-profit condition.

Q: To encourage victims of price rigging to come forward, a federal law gives triple damages to the victims. The "victims" in the famous General Electric case of long ago were buyers of large-scale electric generators; the

"damages" paid in triplicate to the existing buyers were the markups caused by the price rigging of GE and friends.

1. *True or false:* Because of exit, the firms driven

⁸ Charles Tiebout, "A Pure Theory of Local Expenditures," *Journal of Political Economy* 64 (October 1956): 416-424. An up-to-date discussion of this and related matters is Dennis C. Mueller, *Public Choice* (Cambridge: Cambridge University Press, 1979), Chapter 7.

⁹ If exit is impossible, agitation ("voice") becomes the only remedy to an evil policy or evil times, as Albert O. Hirschman puts it in his brilliant and readable application of economic thinking to politics, *Exit, Voice, and Loyalty: Responses to Decline in Firms, Organizations and States* (Cambridge, Mass.: Harvard University Press, 1970).

¹⁰ Two contrasting views of the matter are Robert Higgs, *Competition and Coercion* (New York: Cambridge University Press, 1977); and Roger Ransom and Richard Sutch, *One Kind of Freedom* (New York: Cambridge University Press, 1977).

out of business by the price rigging were not around to collect the damages.

2. *True or false:* Because of exit, the firms that survived the price rigging earned normal profits while it was going on, just as they had before.

3. *True or false:* Because of (2), only the ultimate consumers of electrical power, not the firms buying the generators, were hurt by the price rigging.

4. *True or false:* Because the damages paid were a windfall to the buyers, they enriched the owners of the buying firms but did not compensate the people really hurt (that is, the ultimate consumers and the owners of the firms driven out of business).

5. *True or false:* Therefore, although it might be an efficient way of getting private companies to help the government pursue price riggers, the provision for triple damages has no basis in fairness.

The problem looks to the past history of entry and exit; the next problem looks to the expected future history of entry and exit.

T or F: Because it will discourage new entry, a law preventing businesses from exiting from Ohio may, contrary to its intent, reduce the amount of business in Ohio.

A: The ability to move if a location proves mistaken is valuable. Businesses moving to Ohio will have to forgo

A: The firms driven out were gone, and because their competition was gone the surviving firms could raise the price of electricity to cover the rigged price of electrical generators. That the surviving firms survived implies that their profits were unaffected by the conspiracy. The ultimate consumers, who could not exit, bore the burden in higher electricity prices. When the money damages were paid after the successful suit, the owners of the surviving firms kept the money, since the damages were a fixed (negative) cost that did not affect marginal cost or price. The price of generators was reestablished at its competitive level, but the true victims were not compensated for past damages. The true victims were not compensated: That may be efficient, but it is not fair, and fairness is one purpose of the law. Therefore, true on all counts.

the possibility of leaving. Fewer will come to Ohio in the first place under these terms. What Ohio gains by forcing old businesses to stay it may lose by failing to attract new businesses. Therefore, true.

If entry and exit does not protect you in the short run, it may do so in the long run.

T or F: Because a passenger is stuck in a cab once he gets into it, cab drivers would in the absence of regulation exploit passengers; the market in this case gives the passenger no protection.

A: This is the usual justification for regulation of cabs and other things (for instance, hotels). Were such behavior widespread—advertising \$0.50 per ride, say, and

then actually charging \$5, backed up with physical threats—companies would form that based their appeal on a guaranteed fare. A roadside diner has the same incentive to serve bad food to nonrepeat customers; Howard Johnson's, Stuckey's, and the like have made fortunes guaranteeing a minimum quality. So would Reliable Cab, Inc., and No Rip-off Rented Cars, Inc. Therefore, false.

COMMENT The argument is not that the market protects everyone all the time. It is that the market might in some cases be as effective a protector as regulation. Regulation is not perfect, either. As a Supreme Court justice once said, we must always remember that regulations are enforced by ordinary people, not by all-knowing saints. But we must also remember, as J. M. Keynes once said, that in the long run we are all dead. That fraud in cabs or restaurants will encourage nonfraudulent firms to enter in the long run is little comfort to the robbed passenger and poisoned eater in the short run. The argument for regulation is often that we cannot wait for the long run. And often it is persuasive.

Entry to Knowledge

When she sees no entry the economist's presumption is that entry is unprofitable. We are accustomed to complaining about the wretched apartments, the awful restaurants, and the crummy stores in our neighborhood. The economist argues that after all what exists might possibly be the best that can be done, at least without someone else subsidizing our apartments, restaurants, and stores. The reflection that opportunities for profit have already been seized is merely a presumption, not a law of nature. Sometimes it will be true that you could indeed make money opening a restaurant to compete with the overpriced hash houses in the town. If you could do so, then your complaint that no one else has done so deserves respect. But it is always worth asking The American Question: "If you're so smart why aren't you rich?"

The question is serious. An economist who claims to see an unexploited opportunity for profit is supposing that business executives have not already seen it, entered it, exploited it, and therefore eliminated the profit in it. But it is the business of the executives, after all, to find opportunities for profit. To believe that economic theory or observation can see better what executives daily strain every nerve to see is arrogant. To announce the alleged sighting to the world without charging dearly for it is foolish. The plainest case in point is advice on investments. Your aunt hears that you are studying economics and asks you for advice on investing in the stock market. What do you reply? You reply that if you were so smart as to know which stocks to pick you would be rich; and anyway (if you do not particularly care for the aunt) you would not tell her. To tell other people is to spoil the opportunity, because those other people will drive up the present price of the "bargain" in stocks, making it a bad bargain. The reasoning should make you suspicious of cheap investment advice, such as the stock tips available for the price of a newspaper or for the price of tuning the television to "Wall Street Week." And even expensive investment advice is suspect.¹¹

The if-you're-so-smart principle applies more widely. In fact, it is a profound limitation on knowledge of the economic future. You will hear economists and others making predictions about the coming boom or the coming fall in interest rates. You should ask whether such knowledge would if it were true create opportunities for profit. If Paul Samuelson's prediction of the coming fall in interest rates printed in his column in *Newsweek* were in fact true and not generally known, then he or you could make a fortune of immense size using it. Markets exist in which you and he could bet against other, less bright investors. That he is only well-to-do, not possessed of a fortune of immense size, and that the very announcement of the prediction eliminates the opportunity, by encouraging entry, should lead you to value his prediction at its price.¹²

The Ultimate in the Logic of Entry and Exit: Rational Expectations

The implication is unsettling for any attempt to predict the economic future. The econometric forecasts sold to the government for making policy and to businesses for making money appear to be just like weather forecasts. If I am sure that it will rain tomorrow, I can make money buying up umbrellas today.

¹¹ The case against security analysis—predicting which stocks will make money—is by now overwhelming, though the industry still prospers. See James H. Lorie and Mary T. Hamilton, *The Stock Market: Theories and Evidence* (Homewood, Ill.: Irwin, 1973).

¹² As Samuelson himself demonstrated mathematically in his "Proof That Properly Anticipated Prices Fluctuate Randomly," *Industrial Management Review* 6 (Spring 1965): 41–49.

It is not a meteorological principle, however, that people will enter until the weather forecast is worth only the cost of acquiring it. It *is* an economic principle. The very science that makes the economic forecast possible suggests that it is only worth its cost. And the mere announcement of the scientific finding changes the conditions on which the prediction was based. Cold fronts do not change their behavior when they hear the weather forecast; when they hear the economic forecast, people do. On both counts the economic forecast should have built into itself the economics of how people acquire and use forecasts. This subtle and reflexive obstacle to economic speculation, both intellectual and money-making speculation, is called *rational expectations*.¹³ It is a version of the if-you're-so-smart principle, which is itself a version of the pervasive principle that entry drives supernormal profits to zero.

Summary

Profits draw business managers as flowers draw bees, with the result that in the long run profits are zero (or normal). An economist will do well to doubt any assertion that activity *X* earns super- or subnormal returns, correctly measured. The doubt will not always be warranted. Sometimes people do earn supernormal profits, do get to the gold first, do spot some undervalued stock. That they do, indeed, is the reason for the tendency toward normal profits, for their success attracts further entry. The principle of entry and exit is very widely applicable to businesses, to political communities, to economic regulation, to the stock market, and to every other situation in which movement can earn the mover profit. It appears and reappears throughout economics: as the zero-profit condition, as the condition that revenues equal cost, as the subsistence theory of populations, as the Tiebout effect, as the exploitation condition, as the principle of incidence, as the natural protector of consumers, as the if-you're-so-smart principle, as rational expectations.

EXERCISES FOR SECTION 14.2

1. Criticize each of the following statements in view of the principle of entry and related ideas:

- a. "Furs are luxury items in great demand, so furriers must be rich."
- b. "The fur industry has recently experienced a big increase in demand and has not had time to expand by entry. Therefore, file clerks in the offices of furriers will be well off, at least for a while."
- c. "I have an economic model that says that the fur industry will make a lot of money soon. Invest in fur stocks."

¹³ The most recent uses of the argument are in the mathematical theory of booms and busts, for example, Robert E. Lucas, Jr., "Econometric Policy Evaluation: A Critique," in Karl Brunner and Allan H. Meltzer, eds., *The Phillips Curve and Labor Markets* (Amsterdam: North-Holland, 1976). It was in fact anticipated in the writings of so-called "Austrian" economists, for example, Ludwig Von Mises, *Human Action: A Treatise of Economics* (New Haven, Conn.: Yale University Press, 1949), p. 867:

If the end of the boom could be calculated according to a formula, all businessmen would learn the date at the same time. Their endeavors to adjust their conduct of affairs to this information would immediately result in the appearance of all the phenomena of the depression. It would be too late for any of them to avoid being victimized. . . . What people expect from the economists is beyond the power of any mortal man.

2. Make the entry-exit argument *against* governmental regulation of the following. Imagine in each case why a firm would have the incentive to offer a good deal.
- Hotels, to prevent them from charging more than they advertise, or raising rates on a snowy night.
 - Airlines, to prevent them from investing less than the best amount in safety.
 - Drug companies, to prevent them from selling drugs that do not work.
 - Sausage makers, to prevent them from selling dirty sausage.
3. *True or false:* The profits from providing a known, safe product, the threat of loss if one's reputation becomes bad, the threat of legal suit, all backed up perhaps by governmental requirements that firms *tell honestly* how safe their products are, would be a substitute for governmental regulation.
4. Is Exercise 3 a conclusive argument in favor of deregulation? Is entry and exit a *perfect* substitute for regulation? What about the people who die in the crash or from using the bad drugs?

PROBLEMS FOR SECTION 14.2

- 1. It has been observed that innovations in industrial or agricultural or any other sort of technique do not travel with lightning speed throughout an industry. When these innovations are very expensive (mechanical harvesters in nineteenth-century agriculture, computers in contemporary industry), as many innovations are, and when firms vary in size within an industry, it is possible to develop a simple theory that accounts for this phenomenon (conceived by P. A. David of Stanford University), as follows.
 - Suppose that there is substantial variation in the size of firms before the innovation comes along. Suppose also that the adoption of the innovation reduces average cost by some fixed absolute amount at all scales of output and for all firms. Why, then, will larger firms be more likely to buy the innovation?
 - If we observe, as we often do, that small firms that do not adopt the innovation immediately do not disappear, what assumptions about entry into the industry are required to make the theory work? If the small firms were the more profitable ones, would this assumption be necessary?
 - What if the innovation could be rented out? What assumption about the costs of renting versus those of owning are required to make the theory work?

d. *Philosophical postscript:* Economies of scale at the level of individual firms are widely neglected by economists. Why?
- 2. In the middle of the nineteenth century, laws of incorporation were passed in England, France, and the United States that limited the liability of owners of companies in case of bankruptcy to the amount they had invested in the company (under the earlier law of partnership there was no such limit—any partner was liable for any debt of the company, no matter how large). The many opponents of the law argued that it shifted risk to, say, suppliers of a bankrupt company, who could not now be sure of collecting on all the credit they had extended to the company. *True or false:* Even if this were true, the suppliers would not be hurt.
- 3. Crème-de-la-Crème Canned Soup is usually tasty but it sometimes kills its eaters with botulism; its eaters, you may suppose, are fully aware of the probability of death. *True or false:* Unless the company selling the soup is made liable for the deaths, it will have no incentive to improve the soup (in healthfulness if not in taste).
- 4. Consumer groups want to prohibit the practice of "repricing," that is, marking up prices of products already on the shelves if a new shipment costs more. *True or false:*

This will cause supermarkets to stock lower average amounts on their shelves and to raise prices.

→ **5. TV editorial:** [Wide shot of a carnival] The Jones Carnival arrived back in town yesterday. Before you send your child to it, you may wish to know of the shocking condition of the machinery [angle shot of dizzying ride, zoom in to tight shot of dirty machinery]. We tried to talk to the owner about the problem [shot of owner slamming a door in the face of the camera; distance shot of owner escaping out back door] but he refused. [Shot of earnest-looking executive of Channel 5] We could find no state or local office responsible for carnival safety. It is a miracle that the injury rate is low; we should regulate the carnival business before our miracles run out [switch to Channel 5 logo, announcer's voice]: Channel 5 welcomes replies to its editorials.

What is your reply? Is it a miracle?

→ **6. An economist working for the government shows you a statistical model he has developed of the housing market and makes the prediction with his model that the price of housing is going to climb 50% in the next year. Why do you not believe a word he says?**

True or false

- **7. Since holders of MBAs now make a lot of money, you should go to business school.**
- **8. Although the original owners of licenses to operate cabs in New York City have been enriched by the rise in value of a license to \$50,000 each, any subsequent purchaser of a license makes only normal profit on his or her purchase.**
- **9. A law preventing landlords from evicting nonpaying tenants will hurt the poor.**
- **10. Since silver has doubled in price recently, it is a good investment, now.**

CHAPTER

15

Taxes

What to Read For

How do taxes constrain exchange? How does the elasticity of demand and supply change the way a tax affects output? What happens to the revenue from the tax? What is the deadweight loss from a tax, and how is it calculated in a diagram of supply and demand? What is the “second-best” argument? What is a lump-sum tax? Does a tax stay where it is placed legally?

Taxes Add Constraints to Markets

The earlier chapters have described the workings of supply and demand in the absence of any constraint except for scarcity. Many markets do operate in this unconstrained way and when they do not, the constraints are for many questions unimportant. That the government by taxing cigarettes has added a constraint to the behavior of producers and consumers of cigarettes is no obstacle to predicting how a new report by the Surgeon General or a new cigarette-making machine will change the price and quantity of cigarettes.

Still, governmental constraints are common and they provide useful exercises in the theory of price. The economic choice facing governments is often whether or how to constrain exchange. Much economic thinking, therefore, has grown up around such constraints. The thinking, when fully grown, was found to apply to nongovernmental constraints as well.

Taxation is, aside from death, the most common constraint on exchange. Even in death the tax collector gets his due. It was said in 1820 and could be said today that “the dying Englishman, pouring his medicine, which has paid seven per cent, into a spoon that has paid fifteen percent, flings himself back upon his chintz bed, which has paid twenty-two percent, and expires in the arms of an apothecary who has paid a license of a hundred pounds for the privilege of putting him to death.”¹ The government intervenes between the

¹ Sydney Smith, in the *Edinburgh Review* (1820), review of Seybert's *Annals of the United States*.

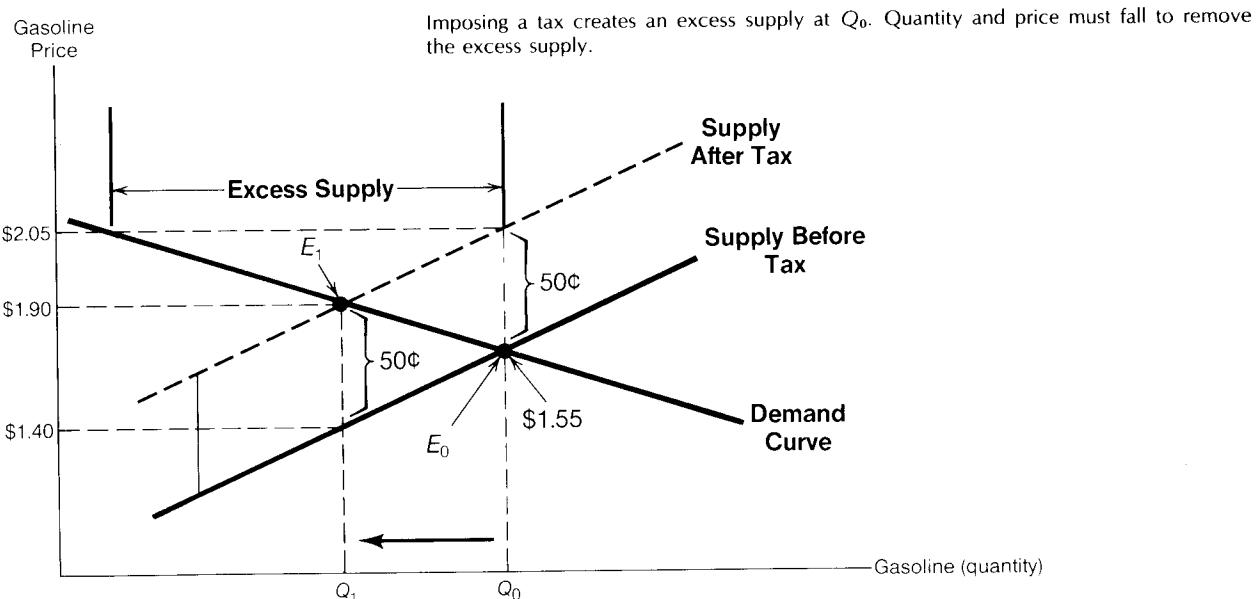
apothecary and his customer, demanding a cut in the gains from their exchange. In the United States at present all levels of government—federal, state, and local—take together a cut of some 40% of national income, and in some other countries, such as Britain and Denmark, their cut is higher. The effects of taxation are important.

A Tax Reduces the Amount of a Good

The first and obvious effect of a tax on spoons, chintz beds, or, to take a modern example, gasoline, is that it reduces the amounts of them consumed. A tax of \$0.50 on every gallon of gasoline consumed will reduce the consumption of gasoline below what it would have been. By the operation of the law of demand, a rise in the price will reduce consumption. The government imposing the tax can be viewed as providing an input, namely, the right to operate a gasoline station. In return for this valuable right, the government requires payment of \$0.50 on each gallon sold, increasing the cost at each quantity supplied by \$0.50. Were the tax imposed on a market for gasoline originally in equilibrium at E_0 in Figure 15.1, the immediate effect would be to raise the price to \$2.05.

At this high price, however, there would be excess supply—consumers do not wish to consume as much at \$2.05 per gallon as they did at \$1.55. The new price (including tax) would fall until, by increasing demand and reducing supply, point E_1 is reached. In this new equilibrium output is lower. Demanders pay more and suppliers receive less than they did before the tax put a wedge (of \$0.50 in width) between the demand price and the supply price.

Figure 15.1
Effect of a Tax on Gasoline



The First Fundamental Theorem of Taxation: A Tax Has Little Effect on Inelastic Goods

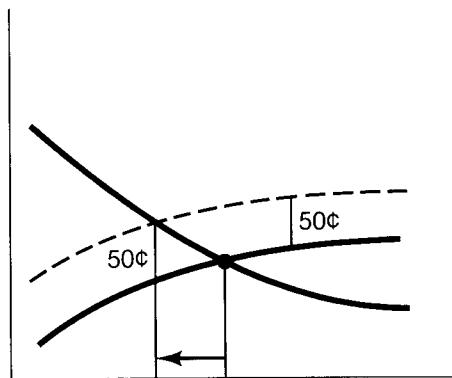
T or F: The less elastic is either the demand or the supply of gasoline, the less will a given tax reduce output.

The amount of the fall in output caused by a tax depends on the elasticity of the demand and supply curves. The First Fundamental Theorem of Taxation is this: A tax has little effect on *inelastically* supplied or demanded goods.

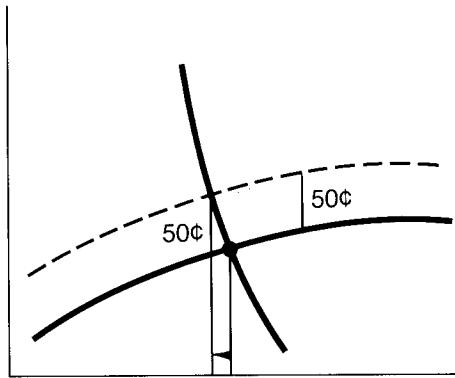
A: True, as is apparent in Figure 15.2. In each graph the initial, pretax equilibrium is at the same point, and

**Figure 15.2
The Less Elastic Demand or Supply, the Less a Tax Affects Output**

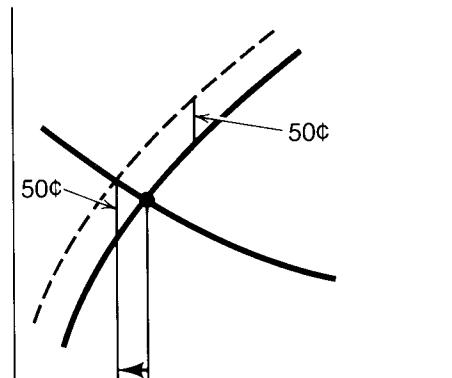
The price paid by demanders must rise and the price received by suppliers must fall, so that the tax can be paid. The less elastic the demand or the supply or both, the smaller the fall in quantity caused by these price changes.



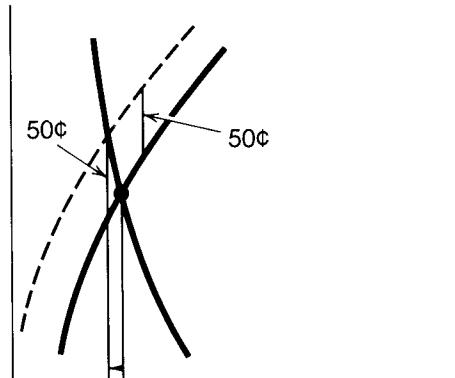
(a) Both D and S Elastic



(b) D Inelastic, S Elastic



(c) D Elastic, S Inelastic



(d) Both D and S Inelastic

in each \$0.50 has been added to the height of the supply curve. Look down a column for the effect of a lower

elasticity of supply and across a row for the effect of a lower elasticity of demand.

In truth, the result is obvious without the diagrams. Taxes increase prices to demanders and reduce them to suppliers. If neither suppliers nor demanders change the amounts they wish to supply or demand very much when the prices they face change (which is what "less elastic" means), then it is evident that the tax will have little effect on the amounts.

You need information about both elasticities to answer such questions as the following.

T or F: If the elasticity of demand for automobiles is -2.0 , an antipollution tax on their sale of 5% of their price will reduce the number of such automobiles consumed by 10% . (*Hint:* Look at a diagram.)

A: Only if the elasticity of supply of the automobiles were infinite (or, practically speaking, large) would the

full 5% tax appear as a 5% rise in the price consumers face, reducing the quantity demanded by $(5)(2) = 10\%$. Otherwise some of the tax will stick to the auto companies or to owners of secondhand autos. Therefore, false. Inelastic suppliers or demanders pay taxes. The increase or decrease of the price from its pretax equilibrium measures on whom the burden falls.

But think before you jump.

T or F: A tax on the crops of blue-eyed soybean farmers over 47 years of age in northwest Iowa will fall chiefly on consumers of soybeans if the world demand curve for soybeans is highly inelastic.

A: False. It will fall chiefly on the blue-eyed farmers over 47 years of age in northwest Iowa. The point is that the demand curve facing them is very elastic indeed,

regardless of the elasticity of the world's demand curve. This is because it is a residual from the demand of the world and the supply of farmers more popular with the taxing authorities. Not particular types of farmers but all farmers—blue-eyed or otherwise—face the world demand. A tax on the whole soybean crop would have the effect described, not a tax on this one small part of the crop.

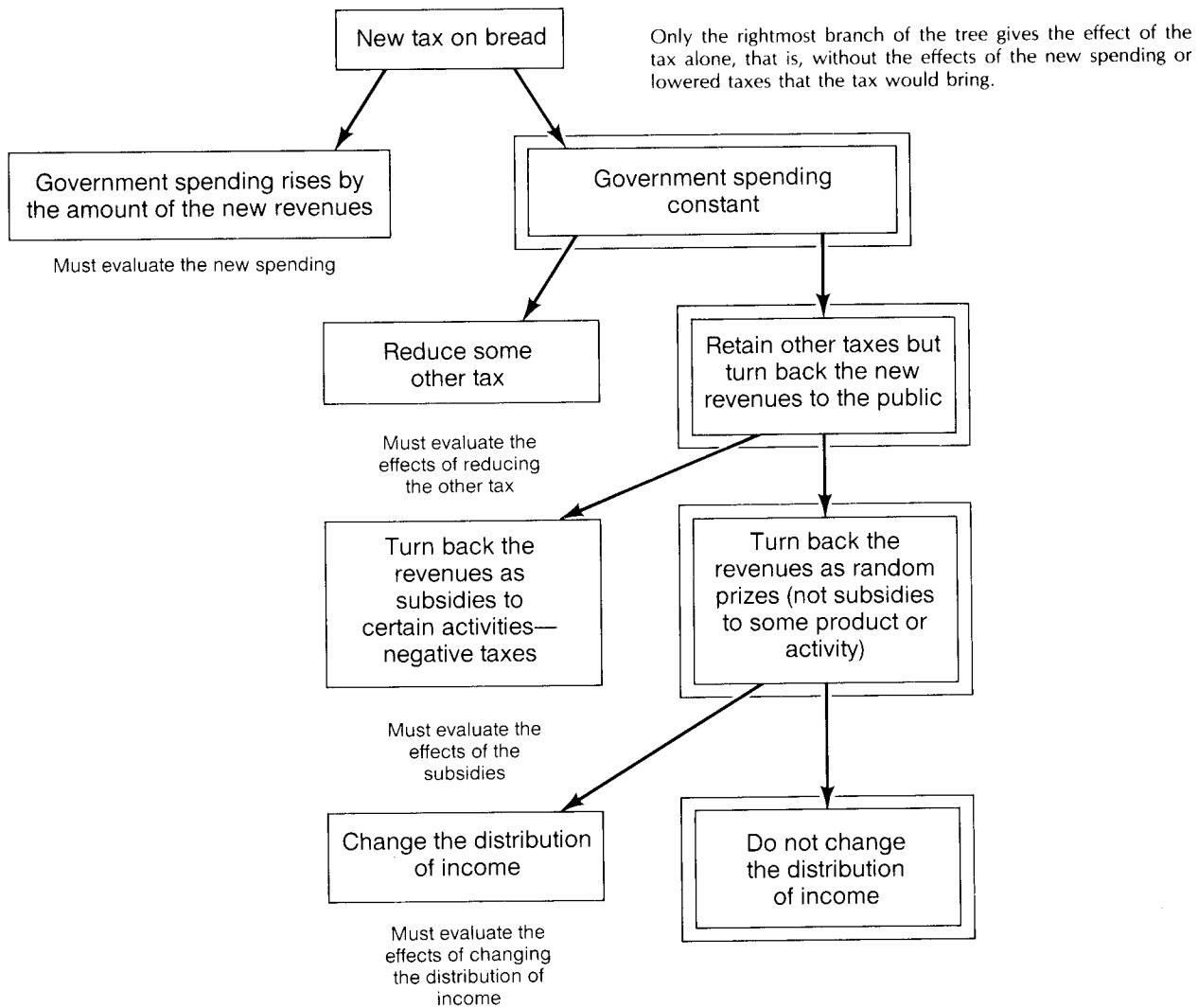
How to Sidestep the Question of How the Money from Taxes Is Used

The collecting of taxes, then, hurts somebody. But the collections from a new tax are not usually dumped into the sea. The government uses the command over resources acquired by taxation to redirect the resources toward palace guards, highways, diplomatic receptions, criminal courts, and other useful things. In other words, there is a spending as well as a taxing side to the interventions of governments in the market. And even on the taxing side by itself there are other taxes, which may or may not be changed when a new tax is introduced. These matters, however, are not the subject. For clarity they must be separated from the subject of a new tax. The effect of the tax alone must be separated from the new spending or lowered taxes that the tax would bring.

The way to accomplish the separation is to assume, solely for purposes of clear thinking, that the collections of a new tax on, say, bread are turned back to the public in a *neutral* way, that is, in a way that has itself no repercussions on the expenditures of government, the level of alternative taxes, or the size and distribution of income. It is not difficult to imagine *nonneutral* methods of turning back the collections of the tax on bread. They could be used to buy ammunition used to shoot half the population; or to eliminate a tax of \$10 per gallon on gasoline that had stopped all travel by automobile; or to subsidize the output of hallucinogens, distributed free to soldiers at guided mis-

sile installations or to the president and his cabinet, or to increase the income of people with a very high income elasticity of demand for Latin poetry. The decision to do such things with the collections, desirable or undesirable as they may be, is distinct from the decision to impose the bread tax. The tree of possible treatments of a new tax is as shown in Figure 15.3. The rightmost branch of the tree isolates the effect of the tax itself, leaving the examination of combinations of the new tax and changes in total expenditure, in other taxes or subsidies, and in the distribution of income for another day.

Figure 15.3
Possible Repercussions of a New Tax on Bread



The Second Fundamental Theorem of Taxation: A Tax Causes Deadweight Loss

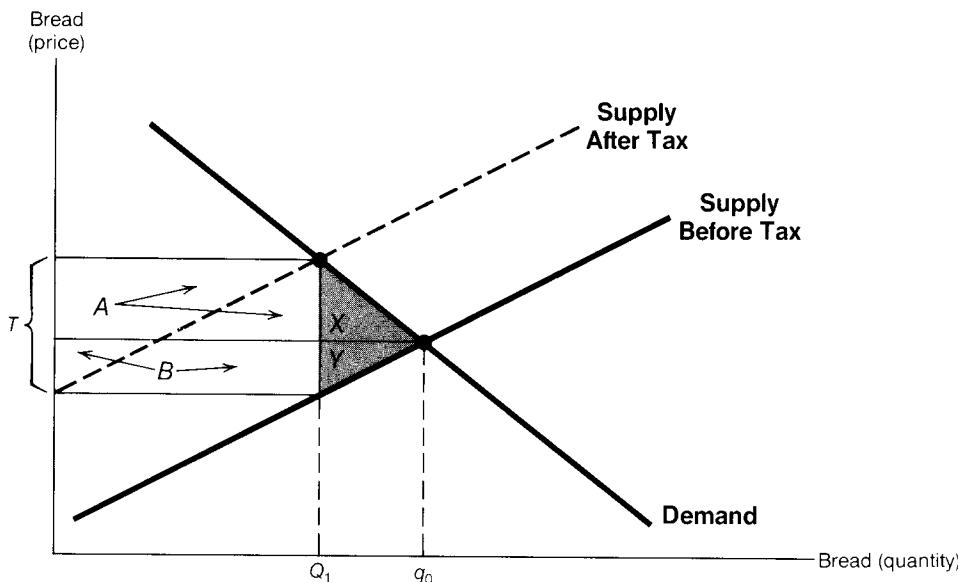
Looking at a tax on bread this way might seem to assume away all the effects of the tax. What goes out of one pocket of the public in the form of a tax on bread goes back into another pocket in the form of a neutral turning back of the revenues from the tax. This is almost true, but not entirely. The Second Fundamental Theorem of Taxation is that taxes cause additional, net, "dead-weight" loss. The fundamental truth about the effect of taxes can be demonstrated with consumers' and producers' surplus:

T or F: The sum of what suppliers and demanders lose from a tax is always larger than the revenue from the tax.

A: True. Consider the tax on bread, taking it to be a specific tax imposed on bakers as illustrated in Figure 15.4. The revenue from the tax is the area $A + B$, that is, the number of loaves of bread consumed after the tax, Q_1 , multiplied by the rate of tax per loaf. Recall the notions of consumers' and producers' surplus. As a result of the tax and the reductions in quantities supplied and demanded that it induces, consumers lose consumers' surplus in the amount $A + X$. Suppliers lose producers' surplus in the amount $B + Y$. The sum of the two losses to the participants in the market for

bread is therefore $(A + B) + (X + Y)$. The revenue portion of the loss, $A + B$, can be viewed as a transfer that makes someone else better off (whoever it is who gets the proceeds of the tax) even as it makes bread eaters and bakers worse off. Such reshuffling of income results in no net social loss. Alternatively, one could reach down into the lower right-hand corner of the tree of possible repercussions of a tax and suppose that the revenues were in fact turned back to bread eaters and bakers each year, on April 15, say. In either case the portion $A + B$ is not socially burdensome. But no matter what the government did with the revenue, there would remain the shaded triangle, $X + Y$, in the diagram, a loss to bread eaters and bakers that cannot be made up by turning back to them the revenues from

Figure 15.4
The Social Loss from a Tax



the tax. The triangle is the social hurt, the fall in national income, the *deadweight loss* from putting a wedge of taxation between the marginal cost of bread

(the supply price), and its marginal valuation (the demand price).

The General Theory of Second Best

The conclusion is that any disturbance of the competitive output reduces income. Any disturbance? Not quite. The assertion is true if the only disturbance to competitive equilibrium in the economy is a tax on, say, food. But if there *already exist taxes on other goods as well*, a tax on food can raise national income. This startling proposition is an example of the logic of *second best*, so named by the British economist James Meade and brought to the wide attention of the profession by Kelvin Lancaster and Richard Lipsey.² Suppose that there is initially a tax on Other Goods but not on Food. The tax on Other Goods reduces national income by lowering the relative price that suppliers perceive and raising it to demanders, persuading them to produce and consume less than the optimal amount. The point to grasp is that the tax raises the price of Other Goods relative to Food. *A tax on Other Goods amounts to a subsidy on Food*, by virtue of nothing more profound than the arithmetic of relative prices. To write out the arithmetic, if Other Goods sell at 2 tons of Food per dozen of Other Goods before the tax, then Food sells at 0.50 dozen of Other Goods per ton of Food. If the tax increases the price of Other Goods in equilibrium to 3 tons of Food per dozen, then by that very fact Food will sell for 0.33 dozen of Other Goods per ton. Food has essentially been subsidized to the extent of $0.50 - 0.33 = 0.17$ dozen Other Goods per ton.

Now suppose that Congress contemplates imposing a tax on Food as well. Will its economic advisors berate it for imposing still further losses of welfare on society? No, they will not, if the tax on Food is enough to offset the tax on Other Goods, namely, a tax of 0.17 dozen Other Goods per ton of Food. This tax *will just offset the implicit subsidy on Food provided by the previous tax on Other Goods*. It will reestablish the relative price of Food and other goods (that is, 0.50 dozen Other Goods per ton of food) that obtained before any taxes were imposed. The upshot of imposing a tax on Food equal in percentage terms to the existing tax on Other Goods is to eliminate the distorting effects of all taxes. In short, the correct strategy for a society trying to do as well as it can given that it cannot attain the first-best position of no taxes at all is to achieve the second-best position attainable. This may well involve imposing taxes, not eliminating them.

It might seem that we have here a formula for taxes with no costs in efficiency: Impose an equal percentage tax of everything, for this will leave the true marginal costs and valuations of things undisturbed. If "everything" means just that, taxes on commodities that pass through markets and also those that do not (such as home handiwork, sleep, and the contemplation of sunsets), then the formula is correct. If "everything" means, however, only marketed commodities, then it is false. As a practical matter the tax collector cannot reach everything. If he taxes only marketed commodities, then the taxes on vacuum cleaners, supermarket food, and psychoanalysis would amount to subsidies on unpaid housekeeping labor, backyard gardening, and transcendental meditation, and

² Kelvin Lancaster and R. G. Lipsey, "The General Theory of Second Best," *Review of Economic Studies* 24 (1956): 11-32.

society would be pushed away from its optimal output of marketed and nonmarketed commodities.

The Ideal Tax on Efficiency Grounds Is the Poll Tax

Aside from taxes on literally everything or taxes on goods inelastically supplied or demanded, the only tax without costs in efficiency is one that alters no incentive to do anything. No margin of substitution in production or consumption is to be disturbed. The only way in which to accomplish this is to make the amount collected by the tax independent of decisions to produce or consume a little more of something. The tax, that is, must be a fixed cost. It's sometimes called a *lump-sum tax*, which is to say that the tax falls in a lump, a fixed cost.

Q: In the 1370s a poll tax (having nothing to do with elections; *pol* means "head" in Middle English) of so many shillings per head was imposed by the English king. So irritating did Englishmen find this tax that it contributed to the Peasant Revolt of 1381. *True or false:* Instead of rioting in the streets, pillaging manor houses, and presenting petitions to the king, Englishmen should

have congratulated the king on hitting on a tax with no costs in efficiency.

A: True. Short of emigration or suicide, a head tax cannot be avoided by altering one's behavior. Therefore, it does not change behavior, in particular the behavior of supplying or demanding goods.

The Third Fundamental Theorem of Taxation: A Tax Falls Where It Will, Not Where It Is Put

The effect of taxation of gasoline on efficiency (the size of output, because the tax reduces it) and on equity (the distribution of the income between suppliers and demanders, because the tax changes it) depends, in short, on the size of the tax and the elasticities of supply and demand. Further, the effects depend *only* on these things. In particular, the efficiency and equity of a tax imposed on the exchange of gasoline for other goods does *not* depend on whether the tax is legally placed on suppliers or demanders. This is the Third (and final) Fundamental Theorem: Taxes fall where they fall, not where lawyers plan to put them.

T or F: A \$0.50 tax per gallon of gasoline imposed on consumers at the pump will hurt consumers more than will a \$0.50 tax imposed on producers at the refinery, because consumers pay the one, producers the other.

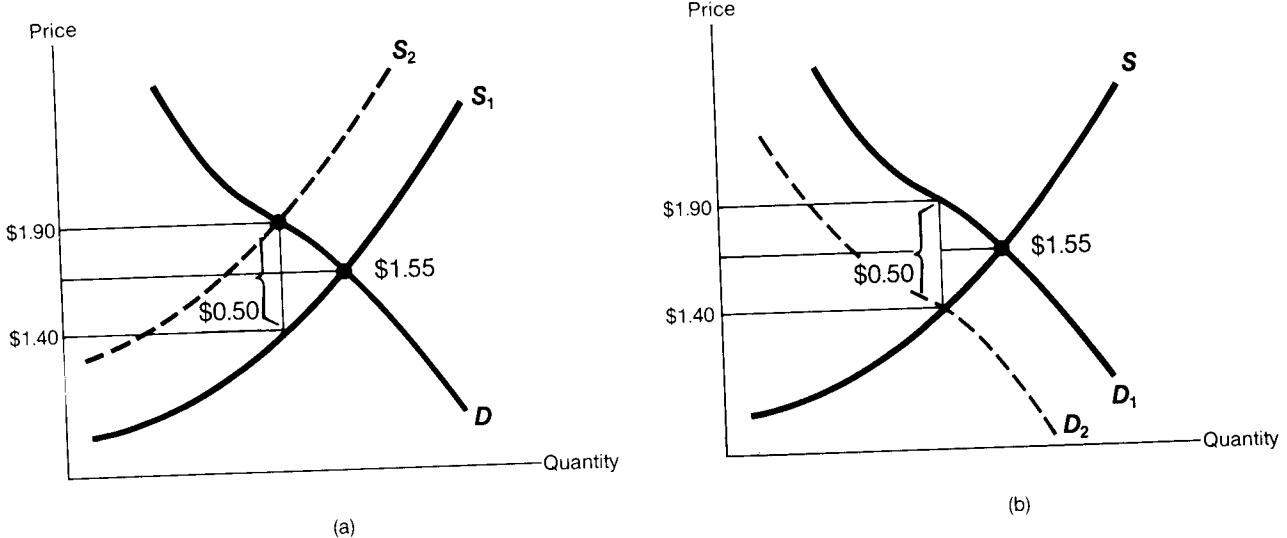
A: The consumer is hurt by a higher price. It does not matter to him whether the gasoline station owner supplies the stuff at \$1.90 a gallon (having added the tax of \$0.50 to its \$1.40 cost) or supplies it at \$1.40 cents, inviting the tax agent to swoop down to collect \$0.50 more. \$1.90 is \$1.90 and is higher than the price to the consumer before the tax was imposed, say, \$1.55 (which is higher than \$1.40 because supply equals demand at a higher price). One can think of the tax as

raising the supply price at every quantity by \$0.50, because suppliers will want their normal return in addition to the tax. Alternatively, one can think of it as lowering the demand price at every quantity of \$0.50, because demanders will value a gallon with a \$0.50 tax \$0.50 less than one without such a tax. It does not matter. Therefore, false. To put it another way, Figure 15.5 shows two equivalent ways of looking at the question of how a tax affects the market for gasoline. To put it still another way, in the final position the new supply price plus the tax must equal the new demand price. This is the same as saying that the new supply price must equal the new demand price minus the tax.

The proposition that the legal location of a tax does not matter for its economic effect is remarkably powerful. Or perhaps one should say that the proposition is not understood remarkably often.

**Figure 15.5
It Does Not Matter Whether a Tax Is Placed on Suppliers (a) or Demanders (b)**

A tax can be viewed as shifting supply upward, demand downward, or as placing a wedge between demand price and supply price. In each case the result is the same: Demanders pay \$1.90, suppliers get \$1.40, the tax collector gets \$0.50. Whether the supply or demand curve is shifted does not matter because the economic effect is the same whether the tax collector takes the money from the demander's hand or from the supplier's.



Q: The City of Chicago imposed some time ago an employment tax of so many dollars per month per employee in businesses over a certain size. Alderman Thomas Keane, defending the tax, asserted that the city's lawyers drafting the law had made very sure that it was a tax on employers, not employees. "The City of Chicago will never tax the working man," said Keane. Comment.

A: Well, no matter how skilled the city's lawyers, the tax does tax the working man. The tax increases the

cost of employing labor, which one can view indifferently, as a higher supply price for labor or a lower demand price. Like a tax on gasoline or food, the tax on employment will raise the cost of labor, hurting employers (and, by raising costs, hurt their customers). But it will also lower the wage paid, hurting the working man. The legal incidence of the tax—the people presented with the bill for the tax—is not the same as the economic incidence—the people made worse off by the tax.

A somewhat more subtle example of the same point arises in discussions of the corporate income tax. When state governments wish to increase taxes, the choice they face is often described as that between taxing people (with a sales or personal income tax) and taxing corporations. The contrast is imaginary. For one thing, corporations are people (if perhaps out-of-state people), not disembodied entities that taxation does not hurt. For another, the tax on corporate profits, which in the first instance might seem to fall on corporate stockholders,

need not stay where it is put. The tax initially drives down the normal return to corporate ownership. This will induce some capital to leave the corporate sector, or at any rate the corporate sector of the state imposing the tax. This in turn raises the relative price of goods produced in the corporate sector and reduces the wage (or employment) of resources employed in the corporate sector. The consumers and employees of General Motors Corporation and General Mills, Inc., pay some of the tax.

The Uses of Entry, Revisited

This third fundamental principle of taxation, like the first two, is a consequence of the principle of entry. A tax or subsidy does not stay where it is put if putting it there creates profits above or below normal. The shifting of the incidence of the tax or subsidy does not stop until all profits are normal, which is to say that it does not stop until the marginal suppliers and demanders are back on their supply and demand curves. The argument applies in fact to much more than taxes.

T or F: Renters of apartments close to the University of Iowa get a valuable benefit from performances at the university by champion musicians or by champion basketball players and wrestlers.

A: If a renter *did* get the benefit the renter would be earning, so to speak, supernormal profits. That is, by comparison with some renter far from the university paying the same price the renter nearby would get extra

benefits, free. Such a situation is not an equilibrium. If it existed, people would shift from the sticks to Iowa City (where the university is located), driving up rentals there and lowering them in the places they deserted. In other words, in equilibrium the renter near the university would pay more for the amenity of having the university close by. The amenity looks at first like a subsidy to renters in Iowa City, but in fact they pay for it in higher prices for apartments.

The beneficiaries of any new amenities in Iowa City are in fact the old owners of property, not the people who rent from them. The old property, especially the land itself close to the university, is supplied inelastically. Just as it would bear a tax on property (even if the tax were put legally on the renters instead of the landowners), so too the old property collects whatever element of subsidy to the neighborhood there is in the opening of a new theater or sports arena at the university. Reasoning of this sort is extremely important for all manner of questions of policy.

Q: The city fathers and mothers of Iowa City propose to run a highway through the west side of the city to make it easier for west-siders to get to the downtown. The downtown property owners claim that commuters into town will benefit. The near-west-side property owners over whom the highway will be built claim on the contrary that the only beneficiaries will be property owners, downtown and elsewhere. Who is right?

A: If the commuters into town in fact get any benefit in ease of travel, they will be willing to pay more for

their houses and more for the goods they buy in the downtown. The resource that is inelastically supplied—land—benefits. Only to the degree that commuters also happen to be landowners will they benefit. The near-west-siders are right. On such logic are city governments everywhere founded: Landowners are in the long run the only people interested in effective schools, easy roads, and other amenities paid for by the city.

Summary Taxes clog markets. A tax puts a wedge, to use another metaphor, between the buying and the selling price, raising one and lowering the other. If the taxed good is elastically demanded or supplied, the resulting change in price

has a large effect on quantities. The inelastic supply or demand bears the burden of the tax disproportionately, being unable to move out from under it (to alter the metaphor once again): That is what “inelasticity” means—inflexibility. Society as a whole, however, cannot move out from beneath the burden of a lower-than-optimal output of the taxed good. The area of unsatisfied willingness to pay in excess of willingness to supply is the “deadweight loss.” The deadweight loss is indeed the fall in the nation’s income caused by the imposition of a tax. But taxes do not always cause a net fall in income. If the tax offsets some implicit subsidy, for example, or if it offsets some other tendency to produce too much of a good, then, by the “theory of second best,” the tax may be desirable. The only truly neutral tax is one that affects no incentive to do anything, whether the thing is marketed food or nonmarketed creative leisure. It is nearly impossible to imagine such a tax. No existing income or excise tax approaches it.

The clogging effects of taxes are not felt necessarily where they are placed by law. The notion that taxes and subsidies end up affecting people whom the law did not believe it was affecting is a good example of characteristically economic reasoning. And it is a good example of the principle of entry. Until all incentive to enter or exit is gone, the analysis of a tax or subsidy is not finished. When it is finished, it often turns out that benefit and burden have shifted in the night.

EXERCISES FOR CHAPTER 15

1. Suppose that the world demand curve for automobiles is Q (in million of automobiles) = 20 million automobiles $-2P$ (where P is in thousands of dollars). Suppose that the supply curve is Q (in millions) = $2P$ (in thousands of dollars). (The formulas that are given are as simple as possible, so be suspicious if you get complex answers).

a. To make sure your high school algebra is in reasonable shape, find the equilibrium price and quantity. (*Hint:* That is, solve for the P and Q that satisfy both equations: Eliminate Q by setting the demand and supply equations equal to each other, then use the value of P you get from this to find out what value Q must take.) Draw the diagram to scale and show that one can find the equilibrium without the algebra, too.

b. Suppose that a tax of \$2 thousand per car is imposed on all automobile makers. Describe verbally what happens to the supply curve. (*Hint:* Make sure that it happens in the same \$2 thousand amount to each point along the curve.) Draw the supply curve plus tax on your scale diagram and identify the new equilibrium point. Without using algebra, what new values of price (paid by consumers and received by producers after paying the tax) and quantity does the tax cause? [*Hint:* Notice that the intercept of the demand curve along the price axis is \$10 thousand and the intercept of the supply curve is zero. The supply curve plus tax therefore starts out at \$2 thousand (the amount of the tax, costs being zero). But \$2 is one-fifth of \$10. So, looking at your diagram, the new equilibrium is one-fifth of the distance from the old equilibrium to the intercept of the demand curve.]

c. What was done geometrically in (b) can also be done algebraically. Since the tax is a dollar amount, though, it will be better to express the demand and supply curves as expressions for price given quantity instead of, as above, quantity given price. Do so. Show that (of course) the initial equilibrium is the same as in (a).

d. Continuing the algebra, take P to mean “demand price, including tax” (with the tax there are two prices, so a choice must be made). The demanders face this price, so their equation is unchanged. But the suppliers face a price \$2 thousand below the demand price.

To their previous supply price (which was an algebraic expression in A) must be added the \$2 thousand tax. The \$2 thousand must therefore be added to the supply expression. (Equivalently, think of the suppliers getting \$2 thousand less than the demand price, P , in order to pay the tax. So their *after-tax* price is $P - \$2$ thousand, and this is equal to the expression in Q that gives their supply price). Solve for the equilibrium after tax. Compare the algebraic result with the geometrical result.

e. Going back to the diagram and its geometry, identify the rectangle of tax revenue. What are its dimensions? What is its dollar value?

f. What are the dimensions of the triangle of deadweight *social* cost? What is its dollar value?

g. What is the ratio of social cost to the total expenditure after tax on automobiles? What do you conclude about how large relative to national income the triangles of social loss are likely to be?

2. Discuss in each case why the tax does not stay entirely where it is put:

a. A tax on beer, meant to put the burden on beer drinkers. [*Hint:* Who else is involved in the exchange?]

b. A tax on the income from corporate bonds (that is, on the interest that corporations pay on their IOUs), meant to burden the rich.

c. A *subsidy* to American shipbuilders, meant to help the workers in Bridgeport.

PROBLEMS FOR CHAPTER 15

- **1.** a. Suppose that the United States consumes a small enough portion of the rest of the world's supply of shoes that changes in the demand of American consumers have little effect on the world price. Draw the supply curve of shoes facing the United States as it imports shoes from foreigners.
 b. Suppose now that the United States produces some shoes. Draw a domestic supply curve on the diagram of (a). Draw the demand curve of Americans for shoes if the United States imports shoes as well as producing them. What determines the price of shoes in the United States? What is the quantity of imports?
 c. Suppose that the American government, responding to the wishes of members of Congress from eastern Massachusetts, imposes a tariff of \$7 per shoe imported. What happens to the foreign supply curve perceived by Americans after the tariff? What happens to the American price? To the American quantity supplied? To imports?
 d. Identify the increase in producers' surplus to American producers of shoes and the decrease in consumers' surplus to American consumers of shoes. Identify the government's revenue from the tariff. Is this revenue a net loss to the United States?
 d'. What, therefore, is the net loss to the United States of imposing the tariff?
 2. In terms of the diagram in Problem 1, what is the maximum amount that the beneficiaries of the tariff (the shoe producers and the taxpayers whose taxes are reduced by the revenue from the tariff) would be willing to pay to convince Congress to vote for the tariff? What is the maximum amount that consumers of shoes would be willing to pay for a vote against it? What do you conclude from the difference between these two amounts about the desirability of the tariff?
 3. The United States is a major importer of cameras (chiefly from Japan and Germany). It consumes, in other words, a large share of the rest of the world's supply.
- a. Draw the excess supply curve for cameras of the rest of the world and the excess demand curve of the United States in a diagram and identify the equilibrium point.
 b. What happens to the equilibrium point when a tariff of \$7 per camera is imposed?
 c. What is the loss to American consumers of cameras? What is the gain to the American

Treasury and thence to American taxpayers? Is the tariff ever desirable from the American point of view?

4. Suppose that Argentina, a small country in international trade, imposes a 10% tax on its exports, such as beef and wheat.

a. What will this export tax do to the Argentinian price of beef and wheat relative to the price of Argentinian imports, such as televisions and automobiles?

b. What will a 10% import tax do to the price of beef and wheat relative to televisions and automobiles?

c. If the value of imports is equal to the value of exports, what general assertion can you make about the effect of export taxes compared with import taxes?

- o 5. Britain both produces and imports automobiles, taking a small percentage of foreign supply as its imports.

a. Show in a diagram the effects of a £100 import tariff on autos on government revenue, on imports, on rents to British producers, and on British consumers.

b. Show that a £100 tax on all autos consumed in Britain, whether foreign or British made, combined with a £100 subsidy to British manufacturers for every auto they make will have effects equivalent in every detail to the tariff.

c. Reflect briefly but cynically on the following facts. Britain and other Common Market countries are committed to free trade in autos among themselves. Yet Britain and other Common Market countries have large taxes on automobiles and large government subsidies to national auto manufacturers.

True or False

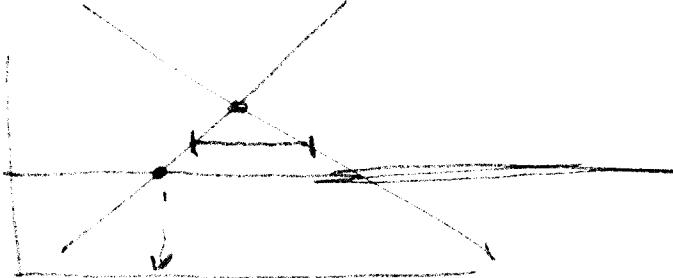
6. The social security plan is an excellent bargain for employees because their employer is required to contribute a dollar for every dollar the employees themselves pay into their retirement fund.

- o 7. If bread is supplied perfectly inelastically, a tax on bread has no social cost.

8. If bread is demanded inelastically, a tax on bread has no social cost.

9. A once-for-all tax on stocks of paintings by Picasso has no effects on efficiency.

10. Subsidies are negative taxes. Since taxes reduce efficiency, a subsidy to shipbuilding will raise efficiency.



Shirts Bought in US

Competition for Property Rights

16.1 Competition for Supernormal Profits

What to Read For How do quotas mimic and contrast with tariffs? Why are price controls bad? How do price controls, licenses, quotas, and the like produce corruption? How do they produce waste in competition? Does a person standing in a queue pay only the money price? How do ration coupons prevent the waste that is a result of queuing?

The ingenuity of governments in constraining exchange is astonishing. In addition to ordinary taxes collected in money, the government, usually with the enthusiastic approval of its citizens, imposes taxes and subsidies collected in goods, compulsory services, prohibitions, regulations, price controls, rationing, quotas, and constraints on entry.

Quotas Foreign trade is a good place to begin in exploring these fruits of the human imagination. An alternative to a tax on imports is a quota, a restriction on the physical amount per year of shoes from Italy or shirts from Hong Kong or autos from Japan permitted to enter the United States. In many ways the effects of a quota are the same as the effects of a tariff. If a quota on, say, shirts is to have any effect, the amount allowed to be imported must be less than the amount that would have been imported without the quota. In Figure 16.1, which shows the American excess demand for shirts, the constraint on the number of shirts imported must be to the left of Equilibrium. To the delight of stockholders and workers at Hathaway and Arrow, the American price of shirts is raised from the World Price to the Domestic Price, as though a tariff of \$7 had been imposed on the importation of shirts.

If the importation of shirts is not to become a mad scramble to import before the yearly quota is used up, however, the government must somehow allocate

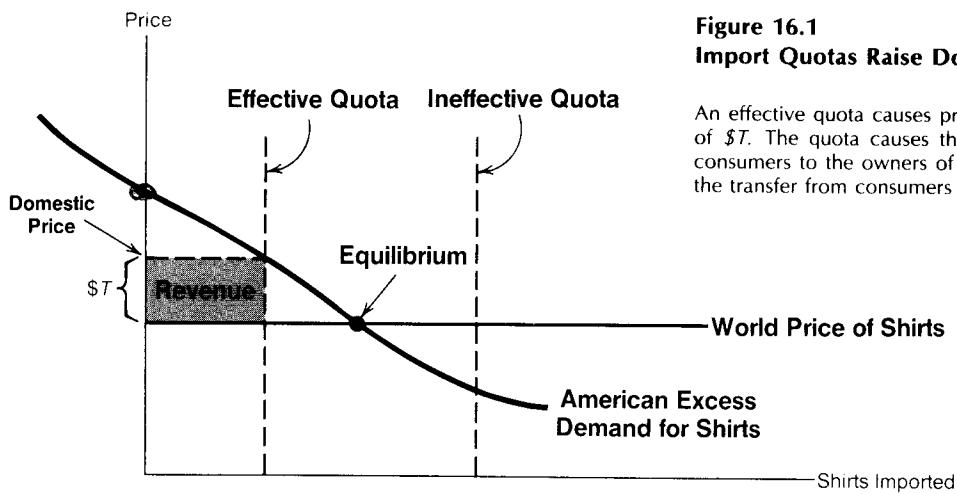


Figure 16.1
Import Quotas Raise Domestic Prices

An effective quota causes price to rise by $\$T$, as does a tariff of $\$T$. The quota causes the transfer of area Revenue from consumers to the owners of the quota rights; the tariff causes the transfer from consumers to the government.

the right to import shirts. It could sell the licenses to import some number of shirts to the highest bidder. In this case the quota would have effects identical in all respects to an import tariff of height T in the diagram (namely, the import tariff that would restrain imports to the amount specified in the quota: 100,000 shirts per year, say). More commonly it bestows on certain lucky people the license to import, with the result that the revenues from the implicit "tariff" accrue to these people. This is the main difference between a tariff and a quota. Under a tariff the revenue goes to the government. Under a quota the revenue (from being able to buy at the World Price and sell at the higher Domestic Price) goes to the fortunate holders of the right to import so many shirts, tons of butter, or barrels of oil.

Licenses A quota on foreign trade is merely a special case of restrictions on entry. Whenever governments—or, for that matter, gangster chieftains, local temperance societies, unions, professional societies, holders of patents, or hostile public opinion—restrict entry to particular businesses, they bestow on the lucky few who remain a prize. The prize is equivalent to the revenue from a tax, and the value of the license (which becomes apparent if it can be sold) is a measure of the size of the tax. Professional sports, for example, have long been exempted from the antitrust laws in the United States, with the result that owners of, say, baseball teams can enforce restrictions on entry. The right to operate a major league baseball team changes hands at prices of millions of dollars.

Price Controls The enthusiasm with which the government imposes price controls is usually a reflection of the enthusiasm with which they are called for by the people, or at any rate by most of the people most of the time. Judging from their prevalence and persistence, few laws are more popular, for example, than the ones restricting interest rates to some figure below 8% or 18% or 80%—whatever figure is currently considered "usury." Eight centuries ago Christians in Europe were not permitted to loan money at all (needless to say, many Christians nonetheless did), and Jews and Moslems in Europe were permitted to loan it (to Christians) only at rates less than 42% per year. Rent controls, likewise, are

perennially popular. Rents on some buildings in Paris have been frozen since World War I. Minimum wages are still another case in point, in this case a minimum rather than a maximum price. The periodic deliberations on its level find few members of Congress willing to vote against a rise. Price controls are a search for a just price lower or higher than the one that demanders are willing to pay in view of suppliers' willingness to accept.

The question that arises naturally is, if a ceiling on interest rates of 8% on mortgages in Illinois is better than the "usurious" rate of 10% at which the market would arrive without constraint, is not 4% or 0% still better? If rents in Paris at their level in 1915 are better than the "exorbitant" level they would reach without constraint, is not the level of 1789 or 1315 still better? If Congress can in fact improve the lot of workers by raising the minimum wage from \$3 to \$4 an hour, is not \$10 or \$100 an hour still better? The answer is made obvious by the extremes. The \$100 an hour would prevent a great many mutually advantageous exchanges between the suppliers and the demanders of labor, namely, those numerous exchanges for which the labor was not worth \$100 an hour. The loss from price controls, as from licenses or quotas or taxes, is a triangle of forsaken opportunities for exchange. Some particular person may be helped, but people as a whole are hurt.

Completing the Analysis

But the argument that quotas and price controls stay where they are put and result merely in triangles of social loss (and rectangles of private gain) is incomplete. It is incomplete because it assumes that the citizens do whatever the government tells them to do. This is questionable.

People in fact try to avoid the government's imposition, or to shift it to someone else, or to get for themselves the benefit it creates. An analysis of quotas or price controls that stops at full compliance is as incomplete as an analysis of the incidence of a tax that stops at its legal incidence, and for the same reason. A quota on auto imports or a license to operate a taxicab creates so to speak a social vacuum into which the air of self-interest rushes.

The First Result: A Market for Rents Develops

The vacuum is the difference between the demand price and the supply price that a perfectly law-abiding and passive acceptance of the quota, control, license, or tax would bring. There arises a market in the very right to buy low and sell high that a license to operate a cab, say, creates. Most cities restrict the number of cabs allowed to ply the streets for hire. The enforcement of the restriction has sometimes been given unwelcomed teeth by the owners and operators of licensed cabs. In London the attempt by "minicabs" to compete with the familiar black cabs was met with beatings, firebombings, and murder. The restrictions on entry reduce the supply of taxicab services, raise the price above the level to which open competition would drive it (cities also regulate the price of rides, but that is another story), and make the right to own and operate a taxicab valuable. In many American cities the price of a cab license purchased from its previous owner is many thousands of dollars.

Q: Illinois restricts the number of liquor stores to a small number well below the number that would exist without the restriction. Describe how to derive the demand curve for licenses from the supply and demand curves of the industry.

A: The simplest assumption is that all liquor stores have identical cost curves, in which case the long-run supply curve of the industry (in the absence of restrictions) is flat at the point of minimum average cost. With the number of liquor stores limited by law to a small num-

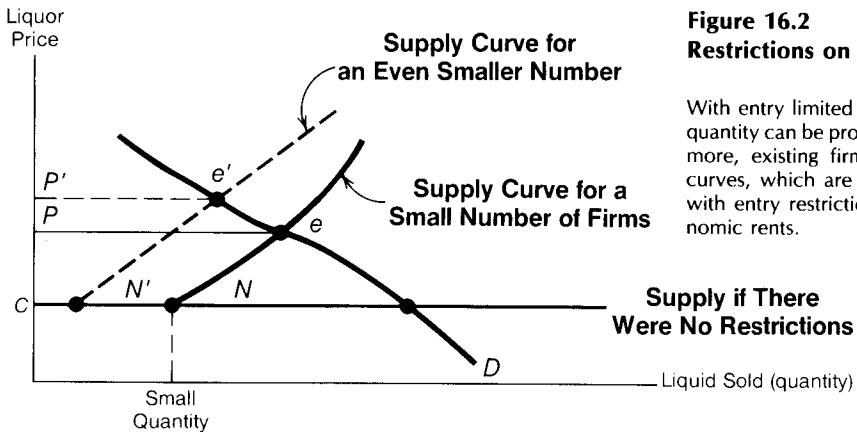


Figure 16.2
Restrictions on Entry Result in Supernormal Profits

With entry limited to a small number of firms, at most a small quantity can be produced at minimum average cost. To produce more, existing firms must expand along their marginal cost curves, which are upward sloping. Therefore, industry supply with entry restrictions is upward sloping. Firms can earn economic rents.

ber, however, the supply curve of the industry beyond the point at which the small number of liquor stores are producing at minimum average cost is the horizontal sum of the liquor stores' long-run marginal cost curves, the solid upward-sloping curve in Figure 16.2.

At an even smaller number of stores the supply curve is the dashed curve in the diagram. The net revenues accruing to the owners of liquor licenses (who may or may not be the same people who operate the liquor

stores) is different, area $P'e'N'c$ instead of $PeNc$. The various areas divided by the corresponding number of firms (and to be precise, converted into a dollar sum that one would be willing to pay for such a stream of dollars) are the demand prices per license (see Figure 16.3). At the point Competitive or beyond, of course, no one is willing to pay anything for a license, for only the competitive rate of return can be earned by using it.

The institution of licenses, in other words, creates a valuable right, the right to collect rents. People seek rents. The demand price for liquor licenses can be viewed as a demand price for restriction on entry, payable as an illegal bribe to whatever public servant has liquor licenses within his or her gift.

T or F: The lower the salary of building inspectors in Chicago and the tighter the restrictions they are asked

to enforce, the more common will be incidents of bribery of inspectors.

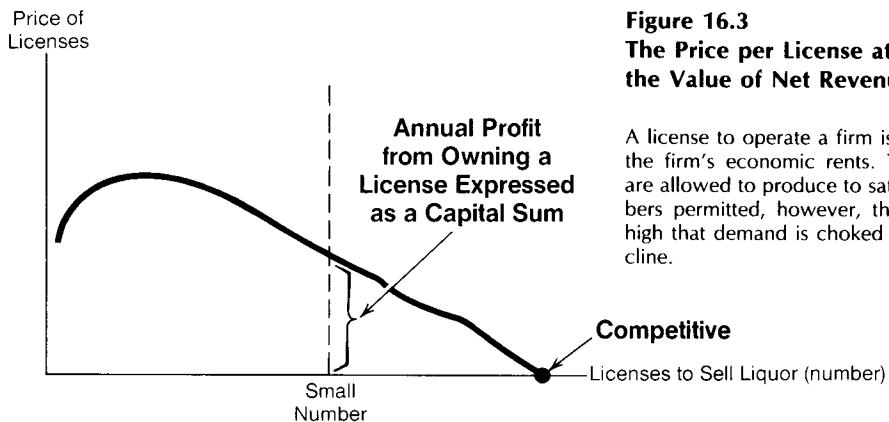


Figure 16.3
The Price per License at a Small Number of Licenses Is the Value of Net Revenues at a Small Number

A license to operate a firm is worth an amount equal to the value of the firm's economic rents. This amount is smaller, the more firms are allowed to produce to satisfy a given demand. At very small numbers permitted, however, the price of the product is driven up so high that demand is choked off significantly, and profits per firm decline.

A: Other things must be assumed equal, such as the law of bribery, the wages of inspectors in alternative employment, and the prevailing standards of morality. The

lower the salary, the lower the supply price of incidents; the tighter the restrictions, the higher the demand. In short, true.

**The Market's
Evasions of
Restrictions
Encourage
Efficiency**

From the point of view of efficiency alone, evasion of a restrictive law and the associated bribery is not all bad. The restriction on unsafe electrical work, for example, might be argued to result in too few buildings. It may be better on balance to build many unsafe buildings than too few safe ones. Any evasion of the law would therefore result in an increase of efficiency, an increase in the number of mutually advantageous exchanges between builders and buyers of houses. That the building inspectors pocket part of the difference between the selling price and the buying price of housing created by the building code is a matter of the distribution of income, not its size. Petty and not so petty graft in government is in fact often defended on such grounds. It is said in Chicago and in many other places that no building could be built that adhered literally to every item in the building code. Graft is necessary.

And in less controversial cases the point is still plainer. Few economists think highly of quotas on imports, for example; they view their evasion with indifference bordering on pleasure. That there is a market in the right to import under a quota, whether or not the market is fully legal, strikes the economist as a good thing, bringing the right to import into the hands that value it most.

**The Second Result:
The Wastefulness of
Nonmarket
Competition**

The second result of self-interest rushing into the vacuum created by restrictions, however, is less pleasing. Prohibited by law from buying a Japanese car at the low world price or even from offering a bribe in the secondary market to the customs officer in charge of enforcing the quota on cars, the self-interested person turns to other forms of competition. To the regular astonishment of legislators, lawyers, and journalists, if people are prevented from paying for objects of desire in money, they will attempt to pay in other coin—in theft, bribery, friendly persuasion, or queuing. A series of acts of Parliament limiting rent increases in Britain has encouraged “squatting,” that is, the stealing of housing by occupying it when it is temporarily empty and trusting to the law’s leniency and the law’s delay to keep it for a while. Payments under the table (called “key money”) to induce tenants of rent-controlled apartments in New York City to surrender them to a new one are commonplace. Rent control makes it costless for landlords to indulge their preferences for white, middle-class Protestants with no dogs or children. Competing for housing with a pleasant, educated manner and the right color skin is the natural alternative when competing with money is outlawed and when the queue outside each apartment is, at the controlled price, long.

These adjustments to price control reduce national income, not only by the amount of the triangle of lost exchange but *also by the amount wasted in competing*. The point is familiar from other contexts.

T or F: The value of the goods stolen plus the cost of police and locks to prevent stealing is a good estimate of the amount by which stealing reduces national income.

A: The question repeats an analysis given earlier. The cost of the police and locks is the obvious portion of the sum. The value of the goods stolen might seem at first irrelevant to the size (as distinct from the distri-

bution) of national income: It is merely the amount transferred from legal owners to the thieves. But thieves will steal things to the point at which their reward in this profession (namely, the value of the goods stolen) is equal at the margin to their reward in an alternative profession, such as running a flower shop or lecturing

on the economics of crime. In other words, the value of the goods stolen will equal the value of the flowers and lectures *not* provided to society when thieves embark on a life of crime. Therefore, true. National income is lower by this amount.

Even when applied to literal stealing, the point is widely useful. The Scottish highlander spends his days raising beef in order to transfer it to lowlanders in exchange for bread. An alternative way in which he can compete for the bread is to spend his days training for war in order to steal the bread from the lowlanders. The highlander by himself may well be indifferent between these two uses of his time, since both get him bread. In the one case, however, the time used results in a benefit to someone else (namely, beef to the lowlanders), while in the other, it does not. Scottish society as a whole is not indifferent between the two methods of establishing who eats the bread, the method using up resources to make beef to entice lowlanders into a voluntary exchange or the method using up the same resources to seize the bread. Scottish income will be higher (adding beef to the goods produced, for example) if competition is limited to voluntary exchange. One can in fact shed light on the economic development of Scotland in just these terms.

**Waiting Time Is the
Most Important
Example of the
Waste**

Spending time waiting in a queue to buy bread is wasteful in the same way as is stealing and is an even more common method of competition alternative to money payment.

T or F: Time is money, it is said. Therefore, it is equally efficient to allocate groceries by time (keeping the price of groceries low and letting people compete by offering more time in waiting lines for the underpriced groceries) as by money (raising the price and letting people compete by offering more money).

A: Waiting in line uses up resources (time in this case) in competing for the groceries without a corresponding benefit to someone (namely, owners of grocery stores). The resources are merely thrown away, with the most profligate line-stander getting the most goods. Therefore, false. Hours are used up *in allocation itself*, as distinct from making goods.

The cause of queuing is an inability to bid for the item with money. A selfish owner of a parking place, a loaf of bread, or a tennis court would always wish money to be given to him rather than time in a queue to be given to no one. That queues do form on roads, in supermarkets, or at tennis courts is testimony to the difficulty, natural or artificial, of raising the price on short notice to the extent necessary to eliminate the queue.

Queuing is usually defended on grounds of equity. People are more equally endowed with time than with money, it is argued, and therefore permitting a queue to form will be more equitable than allowing a price to rise. This is sometimes true (although even if true not necessarily decisive), but it is also sometimes false. If one of the members of a wealthy household stays at home, then that household has more time to shop for groceries than does a poorer household in which both partners work. If companies provide chauffeured cars to their top executives (as British companies do), then the top executives (that

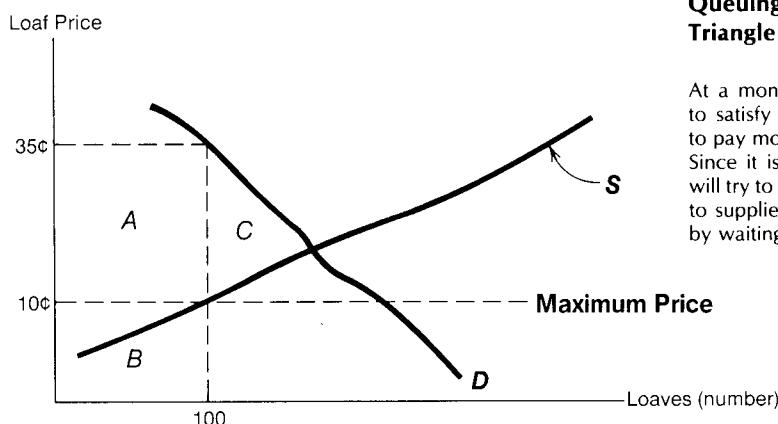


Figure 16.4
Queuing Wastes Resources in Addition to the
Triangle of Foregone Exchange

At a money price of 10 cents, not enough bread is supplied to satisfy demand. As a result, buyers who would be willing to pay more than 10 cents will try to bid up the price of bread. Since it is illegal to bid higher than 10 cents in money, they will try to find other means that do not involve money payments to suppliers. One is queuing. No one collects payments made by waiting in line. These payments (area A) are lost to society.

is, their chauffeurs) will have more time to search for parking spaces than will less wealthy executives who must drive themselves. The point is that money can sometimes buy waiting time.

Here is another example of the waste from queuing, with a more quantitative flavor.

Q: A maximum price of 10 cents per loaf is imposed on bread. At 10 cents the bakers will supply 100 loaves. At 100 loaves the demand price is 35 cents: 35 cents is what a consumer would pay for an additional loaf if necessary.

1. If consumers compete for the cheap bread by queuing, what is the *full* amount they pay for the bread (including the value they put on their queuing time)? (*Hint:* What are they *willing* to pay for the last loaf?)
2. What is the full social loss from the price control on bread?

A: 1. The situation is as shown in Figure 16.4. Bakers produce the 100 loaves sold at 9 A.M. when the shops open each morning, for which consumers pay 10 cents per loaf, or area B (\$10) in total. But they would be willing to pay 35 cents per loaf, and will. They will rise at dawn or before to join the queue outside the bakers' shops each day, paying in the end 10 cents per loaf in money *and 25 cents (by their valuation) in time wasted waiting for 9 A.M.* The consumer who

does not value a loaf of bread as much as 35 cents will not join the queue—and will not get the bread. The customer who valued it more than 35 cents would be willing to join it still earlier. The queue will lengthen until the last person to bother joining it pays exactly 35 cents in total, *the equilibrium price for a constrained supply of 100 loaves*. The full amount that consumers pay is $A + B (= \$25 + \$10 = \$35)$. Notice that if the demand curve has an elasticity of less than 1.0, they will pay *more* than they would pay at the unconstrained equilibrium.

2. The social loss contains two parts, the usual triangular area C, but also A (\$25). Consumers have been induced by price controls to throw away resources valued at A (that they might throw away time for sleep or leisure rather than time at the factory is irrelevant: The hours are nonetheless scarce and desirable, a part of income correctly measured). The area A is the new element. For many analyses of the inefficiencies springing from constraints it is much the largest.

Ration Coupons Can Eliminate Queues

The only way to avoid the loss of queuing or stealing or resource-using bribes or the like is to make them pointless by defining clearly who has the right to exploit the price differential. Ration coupons do so by requiring a consumer to give up along with the money a coupon entitling the purchaser to a certain number of loaves, the total number allotted to all consumers being the number

of loaves supplied at the controlled price. Rationing eliminates queuing (setting aside the possibility that people will devote resources such as queuing time to competing for the ration coupons themselves: as, of course, they will).

If the rationing device is to be effective, it must make it pointless to join a queue.

Q: During the Arab oil embargo of the 1970s, a scheme known as the Oregon plan was widely adopted as a method of reducing queuing for price-controlled gasoline. Under the plan automobiles with odd-numbered license plates were allowed to buy gasoline only on odd-numbered days of the month and those with even-numbered plates only on even-numbered days. *True or false:* Aside from a small psychological effect, which might stop the hitching-up time of the nervous few who wanted to top up their tanks every day (after the plan they would top them up every other day), the plan would have no effect on the length of queues.

A: Consumers would still value the gasoline supplied at more than its controlled price and would still be able to express this valuation by joining queues using up their time. If the equilibrium price for the restricted quantity was \$1.80 a gallon and its controlled price \$1.50 a gallon, \$0.30 worth of time would have to be spent somehow—on odd-numbered days if the sages of the community so decree; or before 5 P.M. except on Sundays; or while dancing a jig—in order to reduce the quantity demanded to the lower quantity supplied. That is, true. The Oregon plan was ineffective.

Ration coupons can eliminate the waste of queuing, but they must be freely exchangeable for money if they are not to result in a pointless social loss by allocating to the wrong hands the available bread or gasoline. The reasoning is familiar. Efficiency requires that all consumers have the same marginal rates of substitution for all pairs of goods. This condition will be violated if coupons are not exchangeable. If the rich (or the red-haired or the fat) value a little of the rationed good more than the poor (or the black-haired or the thin), then the rich will be willing to buy a little of the coupons from the poor at a price at least as high as the poor are willing to accept. Both will be better off.

Q: Suppose that gasoline, fixed at a price of \$1.50 a gallon, is rationed by coupon (number of coupons = number of gallons forthcoming from suppliers at \$1.50 a gallon) and that the marginal valuation of the quantity supplied at \$1.50 is \$1.80. If a ration coupon for a gallon of gasoline can be bought and sold, what will be its price? Does the initial distribution of coupons between sales representatives who drive 500 miles a week and old people who drive 1 mile a week matter for the outcome?

A: The coupons will sell for \$0.30 each. That is, it must be true in equilibrium that $\$1.80 = \$1.50 + \text{the price}$

of a coupon. The right-hand side of this equality is the opportunity cost of buying a gallon of gasoline, namely, its money price plus the money that can be gotten for a coupon.

To a first approximation the initial distribution of coupons does not matter. The old people will be enriched if all the coupons are given to them, as will sales representatives if all are given to them. Aside from the income effects on the price, therefore, the coupons—rights to buy a gallon at \$1.50—are simply a commodity to be allocated to those who value it most.

Summary Competing in ways other than offering money, then, can be wasteful. Giving bribes to police officers, expensive entertainment to members of regulatory commissions, key money to tenants in rent-controlled apartments, and money to people nearer to the front of a queue are not wasteful in this sense, for these gifts merely transfer money or goods from one person to another without using up resources to no one's benefit. They are the market: money competition thinly

disguised. But the time spent in queues, the bullets used to hold up a bank, and the accountants employed to evade income taxes are wasteful. The excess waste arises when people want to enter an exchange but are unable to do so, that is, are unable to compete with money. The resulting gap between what people will pay and what the imported oil or illegal whiskey or rent-controlled apartment cost is supernormal profit to whoever can claim it. Competing for the privilege of supernormal profits—achieved by owning a patent, a loaf of bread priced below equilibrium, an import license, or untaxed moonshine liquor—can result in social waste well in excess of the usual triangular areas.

The waste from attempting to circumvent the market is not always on balance bad. Most would agree that "wasting" resources in campaigning for the presidency is better than simply putting the presidency on the auction block. Using up resources in advertising Anacin or the Bible, it is said, transmits information about healing headaches or sin.

People will compete regardless of whether the government has outlawed particular sorts of competition. Competing by exchange is in most cases the most effective sort and will therefore pop up. Political offices not for sale will be competed for by market transactions that closely approximate sale, building inspectors will be bribed, professional linestanders will offer their places for a price. The market is irrepressible, in two senses: First the problem of allocation it solves must somehow be accomplished, by nonexchange competition if not by exchange; and second, that exchange, even if outlawed, has charms.

EXERCISES FOR SECTION 16.1

1. Suppose that the demand curve for imports of steel into the United States is, expressed as depending on quantity, $\$P = \$100 - 2Q$ millions of tons (it will be, of course, $Q = 50 - \frac{1}{2}P$ expressed as depending on price). Notice that you can do all the problems by drawing the situation to scale; or you can use algebra (always, though, guide your algebra with a rough sketch, as a check for common sense).

- a. If the world price of steel is \$50 a ton, how much is imported?
- b. If a quota of 12.5 million tons is imposed, what will be the market price? What level of tariff on imports would achieve the same result?
- c. What is the profit accruing to holders of import licenses? What would the holders be willing to pay to bribe the government to leave them in possession of the licenses?

2. Hotel owners raise their prices for the Olympic Games in a well-known western city. The mayor of the well-known city introduces legislation to stop the "price gouging." If the gouging is not permitted, will the hotel guests be better off? Answer by drawing a supply and demand curve for hotel rooms before the Olympics. Now move the demand curve out (as people flock to the well-known western city to see athletes huff and puff in the afternoon smog). How many hotel rooms will be provided if the price is not allowed to rise? How much will the marginal guest be willing to pay for a room if this old quantity is provided? If he can't pay for it in money, how will he end up paying for it?

PROBLEMS FOR SECTION 16.1

1. The local government of London, in common with many other governments, wishes to reduce the volume of traffic in the center and does this by restricting parking. Instead of raising the metered price of parking spaces on the streets it has decided to eliminate a good many of the parking spaces. Comment on the social cost of this decision.

2. The Polaroid Corporation owns over 1000 patents on various features of the Polaroid Land and similar cameras that take instantly developed pictures.
- Thinking of the marginal cost of Polaroid Land cameras and the demand curve for them, describe what price Polaroid would charge to maximize its net revenues. Contrast this price and the resulting quantity sold to the equilibrium that would occur if Polaroid's patented knowledge were free for anyone to use.
 - Kodak spent millions of dollars devising a substitute technology for instant pictures. Why? What is the social waste in this?
- 3. An alternative to allocating gasoline by queue during the oil embargo of 1974 was dropping price controls on gasoline. Most public figures were opposed to this alternative, and their opposition was given weight by the results of man-in-the-street surveys by journalists asking people what they would do if the price of gasoline rose. All answered, "Nothing. I *need* the amount of gasoline I'm consuming now." Suppose that the reduction in the supply of gasoline during the embargo was 25%, that the equilibrium price before the embargo was \$0.50 a gallon, that the government did not allow price to rise at all, that consumers spent on average one hour waiting to buy 6 gallons at the controlled price, and that they valued their time on average at \$1.80 an hour. Was the elasticity zero, as the man-in-the-street interviews implied? What *was* the elasticity?
 - 4. In the dear dead days of free international competition in oil, the United States had a system of import quotas for oil as a result of which oil inside the United States was a dollar a barrel more expensive than oil outside.
 - Supposing that the allocations of the right to import were distributed at random and that the recipients were able to buy and sell the rights after they were allocated, who were the gainers from the system and how much did they gain? Who were the losers?
 - What would happen if the total quota were the same as in (a) but were allocated to oil companies in proportion to the amount of domestic oil they raised and refined? In particular, what would happen to the domestic price of oil?
 - 5. When the King Tutankhamen exhibit came to Chicago, the Field Museum charged only \$3 for admission. Enormous lines developed at that price. The museum then offered queue tickets, like tickets at a bakery: If you got a ticket at 8 A.M. you could go about your business, such as touring elsewhere in Chicago, and come back to claim your place at, say, 3:00 P.M., when your place in the line was expected to be almost at the door. *True or false:* The giving of such tickets would not reduce the total value of the inconvenience caused by having to wait to get into the exhibit.
 - 6. During much of the 1970s the U.S. government imposed a great many regulations on the oil industry at various points from well to filling station. By the late 1970s foreign oil was selling for much more than "old" domestic oil (that is, oil from wells drilled long ago). There was a controversy about whether or not to remove the regulations, the assumption being that the price of oil products would rise to the foreign level if "old" wells were permitted to charge what the market would bear. *True or false:* In view of the absence of long lines at gasoline stations or other evidence of too low a price, the removal of the regulations could well be expected to have no effect at all on prices and might well be expected to reduce, not raise, them.
 - 7. The price of cabs in New York is set by the Hack Bureau to be the same no matter what the time of day. During the day cab drivers in New York are courteous and accommodating, at least relative to what they become at night: spurning short rides, unaccommodating in manner, and unwilling to take people to Harlem. Why?
 - 8. a. A government imposes a maximum price on gasoline and issues (marketable) coupons, each one of which bestows the right to buy 1 gallon. *True or false:* If the government issues coupons in a number exactly equal to the 100 gallons of gasoline forthcoming from

DO
THIS
IN
CLASS

DO
IN
CLASS

suppliers at the maximum price, then the price of coupons will be equal to the demand price of 100 gallons minus the maximum price, and queues will vanish.

b. *True or false:* If the number of coupons is less than the 100 gallons, the money price and the quantity of gasoline sold will fall (coupons are necessary to buy gasoline) and the price of coupons will rise to the difference between the demand price and the supply price.

c. *True or false:* If the government in its generosity issues 1,000,000 one-gallon coupons, queues equal in length to those without coupons (but still a maximum price) will form.

True or False

9. Since an effective maximum price results in a queue, so will an effective minimum price.

16.2 Unassigned Property Rights and External Effects

What to Read For

What is the simplest way of eliminating the waste from price controls? What is Adam Smith's generalization? Is it a good or bad idea to extend private property to whales, lakes, fish, freeways, and buffalo? What are nonpecuniary externalities? How would tolls, if set optimally, cause highways to be used efficiently? What are external economies? Is it always best to tax polluters in the amount of the harm from their pollution? What is Coase's theorem? Does it hold in a situation in which deals and bribes are cheap to make? Should the person morally responsible for an accident bear the costs, on grounds of efficiency? Why does private property become private?

Private Property with Zero Transactions Costs Eliminates Inefficiency

It will have occurred to you that a simpler way of curing wasteful competition brought on by price controls is to abandon the price controls. There is a more general way to look at this. When rights to use a resource are assigned unambiguously to someone, to sell the rights for what he or she can get for them, the resource will be used efficiently. This is *Adam Smith's generalization*.

The generalization is in some cases wrong, doubtless. But an amusing sidelight on two centuries of attempts to dispute it is that a good many of the counterexamples put forward are in fact fine examples of the generalization in action. Consider, for example, the extreme case of controlled prices, that is, prices set by custom or convenience at zero.

T or F: The optimistic view that self-interest operating in an unconstrained market will lead a competitive industry to conserve its raw materials is plainly mistaken in the case of whaling. Whales are hunted and killed in larger numbers than the interests of the whaling industry as a whole would dictate. Indeed, they are in danger of extinction, taking with them into extinction the business of hunting them.

A: The facts are true enough, but the interpretation in the first sentence does not follow from them. The whales are overhunted because they are not owned by anyone and because the opportunity cost in terms of yet-to-be-born whales of hunting them now, therefore, is zero. For the same reason alligators in Florida and buffalos on the Great Plains were overhunted before they were protected by law: No one owned them. There-

fore, false. If whales were owned, the killing of them would have a price (collected by the owner), and the prospect of selling the right to kill them and their offspring in the future would lead their owners to conserve

them. Conservationist organizations could if they wished bid for the whales in order to prevent them from being hunted at all.

The way in which whales and other denizens of the deep might come to be owned is indicated in the following question.

T or F: The selfish behavior of Peru and Iceland in extending their national fishing rights 200 miles out from their coasts is a great tragedy for a hungry world, for it will reduce the future supply of fish protein.

A: Before the extensions no one owned the fish, and the fish therefore were overfished. Now two govern-

ments own the fish. They have an incentive to conserve the breeding capacity of the fish by restricting fishing: by selling licenses to catch N tons, for example. In short, false. If it is owned, the resource will be better used, producing more not less fish.

In similar fashion the Santa Monica Freeway may be misused (that is, made crowded, with queuing costs to drivers), Lake Erie may be misused (that is, made dirty, with costs to swimmers), the peace of neighborhoods close to Kennedy Airport may be misused (that is, made noisy, with costs to residents), and the air close to U.S. Steel's South Chicago Works may be misused (that is, made polluted, with costs to breathers) if they are not owned. If a Mr. J. P. Morgan, for example, could charge admission to Lake Erie, swimmers and fishermen could bid for the use of the lake against the operators of chemical plants and ore ships. If it is true that the lake is more valuable as a recreation spot than as a dump for chemicals and a fluid medium for ships, Lake Erie will be used for recreation. Lake Erie is in fact owned by the federal governments of Canada and the United States, by the Province of Ontario and the states of New York, Pennsylvania, Ohio, and Michigan. This has meant in effect that it has been owned by no one, that is, by anyone who wished to pollute it.

**An Externality Is
One of the
Inefficiencies
Eliminated by
Property**

As was explained in Chapter 13, economists call such misuse of resources "neighborhood effects," "external effects," "nonpecuniary external diseconomies," or, for short, "externalities," the notion being that using Lake Erie for a chemical dump affects people *external* to the business of making and using chemicals, namely, swimmers, drinkers of water, and conservationists.

Jones imposes a *pecuniary* externality on Smith merely by bidding more for a gallon of the lake than does Smith. No man is an island in an exchange economy. Each affects others, and should if the gallon of lake water is to be placed in the hands that value it most. It is the *nonpecuniary* nature of polluting Lake Erie that leads to misallocation, because swimmers and the rest have not been permitted to bid for the use of the water. They can "bid" only through political agitation and the government's compulsion. Another way of putting the matter is to say that because of a failure to define property rights in the lake, the privately perceived opportunity cost of using it for a dump (namely, zero) is not equal to the social cost (namely, whatever swimmers and other potential users would be willing to pay).

**If Roads Are Owned,
the Externalities of
Congestion Are
Eliminated**

Congestion on freeways is a case in point.¹ When you join the Dan Ryan freeway at the Loop in Chicago already clogged with traffic at 5:00 P.M., you add another obstacle to every other car on the freeway. You join the queue, so to speak, in the middle, adding to the waiting time of all people behind you. Your external effect is small on each car, but large when summed over all cars. Each trip from the Loop to 55th Street that took exactly 30 minutes without you, for example, might take 30 minutes and 0.10 second with you. But the additional 0.10 second affects, say, 6000 cars, making 600 seconds—or 10 minutes—of additional travel time in total. When you decide to enter the freeway, you judge the cost to be roughly 30 minutes, as it in fact is by a selfish calculation. The social cost, however, is not 30 minutes but 40 minutes. That is, it is 30 minutes of private cost plus 10 minutes of externality—other peoples' costs. *But you do not face other peoples' costs.* You and each other individual person, therefore, will not make socially correct decisions.

In particular, if you could get home by side streets that were not subject to congestion but were poorer and slower roads than the mighty Dan Ryan (by Cottage Grove Avenue, say) you would choose the Dan Ryan when your *private* costs were equal by both routes. Too many people (including you) would take the Dan Ryan. It would be too many people, that is, in view of the high marginal *social* costs of another car joining the freeway.

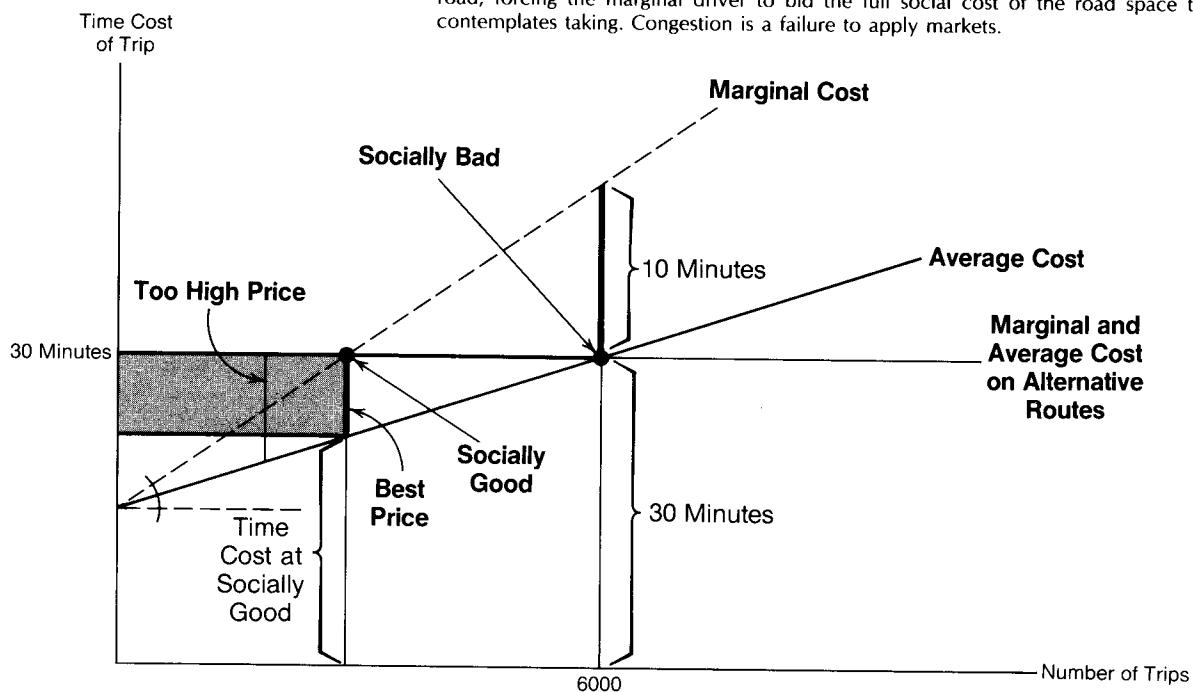
The situation can be made clear in a diagram of time cost measured against the number of trips on the Dan Ryan (see Figure 16.5). The average time cost of a trip rises with more trips: At 6 trips from the Loop, each of the 6 cars could zoom home at 80 miles an hour without fear of collision; at 6000 trips each car would have to drive slower and would take longer. Suppose that the time cost on the poor but numerous and uncongested alternatives to the Dan Ryan is 30 minutes for the standard trip from the Loop to 55th Street, no matter how many trips there were (since this is what it means to say that they are "uncongested"). Rational drivers would join the Dan Ryan until point Socially Bad in the diagram was reached. At this point the rational driver's privately perceived cost of travel on the freeway is equal to the marginal cost of an alternative route.

By contrast, the correct allocation of traffic between the Dan Ryan and the uncongested alternatives is at Socially Good, where the marginal costs are equal. The reason this point is the correct one is that the curve of Marginal Cost on the Dan Ryan does indeed reflect all the costs of an additional car joining the freeway: It is marginal social cost. It is in fact the curve marginal to the Average Cost curve. That is, the horizontal shaded sliver of one additional second of average cost at 6000 trips is the same as the vertical shaded sliver of other peoples' total cost of an additional car (which is the marginal cost, as required). And if the marginal cost on the Dan Ryan were not set equal to 30 minutes, which is the Marginal and Average Cost on Alternative Routes, then by the usual argument something could be gained socially by shifting cars to different routes.

¹ The argument that follows is a translation of Frank H. Knight, "Some Fallacies in the Interpretation of Social Cost," *Quarterly Journal of Economics* 38 (August 1924): 582–606, reprinted in George Stigler and Kenneth Boulding, eds., *Readings in Price Theory* (Homewood, Ill.: Irwin, 1952), pp. 160–179, especially pp. 162–166.

Figure 16.5
The Ownership of a Road Eliminates the Externalities of Congestion on It

If the road is owned by no one, a driver contemplating entering it thinks only of the average cost she alone faces and does not take account of the increase in average cost the trip would impose on other drivers. In other words, drivers will enter the road until Average Cost equals the Marginal and Average Cost on Alternative Routes, which is Socially Bad. If the road is owned by a profit-maximizing firm, or by a state authority that maximizes revenue, is the one that presents a driver contemplating entering the road with the full social marginal cost of her decision, arriving at Socially Good. The firm acts so to speak as a middleman between the marginal driver and the drivers already on the road, forcing the marginal driver to bid the full social cost of the road space that she contemplates taking. Congestion is a failure to apply markets.



So, externalities of congestion lead to too much traffic on a freeway. The word "freeway," however, is the key. No one owns the space on the Dan Ryan. It is "free," in money if not in time. It is the failure to define property rights in the road that leads to the congestion, not some feature of the technology of roads.

It is intuitively plausible that bidding for road space among drivers would solve the problem of too much congestion. Bidding for road space is obviously impossible in a simple way. At least in an age before cheap computers, the transaction costs associated with such a property right are too high. But what of turnpikes, that is, of someone's literally owning the entire road and charging for admission to it?

Q: In Figure 16.5, which expresses price in terms of minutes, what price at most could a turnpike authority

charge for the Dan Ryan at various different volumes of traffic? What is the profit-maximizing price? What

are its welfare characteristics? (*Hint:* Guess the socially optimal price and then show that it is in fact the profit-maximizing one.)

A: The most that the turnpike authority could charge per trip would be the difference between the time cost on the Dan Ryan and the 30-minute time cost on the alternatives, a difference that varies down from a large number at zero trips to zero at 6000 trips. The vertical line marked Too High Price is one such price, one that does not lead to the maximum profit to the turnpike authority. The best price is the one that maximizes the shaded rectangle of profit because only at Best Price

is what the authority loses on its previous sales by lowering the price a little (the short horizontal sliver) just equal to what it gains on new sales (the short vertical sliver). Such a point of balance in marginal gains and losses is clearly a maximum of profit. But it is also the best point from the social point of view. To be sure, the turnpike authority pockets the cash. But the authority is part of the society whose point of view is being taken. In the style of the last chapter, the authority could be defined simply as a device for putting the cash back into the pockets of the drivers through lump-sum transfers.

In short, selfish maximization of the profit from a property right in a highway leads to efficiency (yes, as though by an invisible hand). The negative externality is eliminated by bringing the market to life. Because you do not recognize the full marginal cost of setting out on a congested freeway, you enter it up to the Socially Bad point where your privately perceived cost of travel on the freeway is equal to the marginal cost on an alternative, uncongested route. A turnpike authority is led to charge the Best Price as the one that maximizes its net revenue. Happily, the Best Price is also the one that induces you to recognize to exactly the correct degree the costs you impose on others by adding to the congestion. Ownership of the road, in other words, leads to its socially correct use.

External Economies

External effects need not be hurtful to cause misallocations of resources. If Lebergott gets pleasure from Mendels's consumption of housing but is unable for some reason to express his pleasure by providing Mendels with a housing subsidy, the amount of housing Mendels consumes will be suboptimal. You can see the force of this by imagining that Mendels is Lebergott's son-in-law. Another popular example of an *external economy* (as distinct from a diseconomy, which as you recall is a bad thing) is the relationship between the owner of an orchard and a nearby keeper of bees. The bees pollinate the orchard and the orchard provides the nectar for the honey. It is said, however, that the beneficial effect of more bees on the fruitfulness of the orchard is not captured in a payment to the beekeeper, with the result that he will keep too few bees; likewise, the orchard will be too small. The inability to specify, enforce, and sell property rights in the pollinating activities of bees and the nectar-producing activities of an orchard leads to suboptimal output of both. As it happens, this classic case of the inability of a market to operate was poorly chosen. In the state of Washington, for example, owners of orchards and bees *do* in fact pay each other for their services, exchanging them in elaborately worded and enforced contracts.²

Taxes as Substitutes for Property

Still, markets do sometimes fail. That is, sometimes it is too expensive to enforce and sell a property right. The Dan Ryan may be a case in point, although tolls could be imposed. A clearer case is noise pollution close to airports. What is

² See Steven N. S. Cheung, "The Fable of the Bees: An Economic Investigation," *Journal of Law and Economics* 16 (April 1973): 11–34.

to be done? Two extreme solutions are either to prohibit airplanes from flying anywhere or to permit them to fly anywhere. An extreme solution is not always inefficient. It is doubtless wise to prohibit active cases of cholera or smallpox from wandering where they will, imposing externalities on those with whom they come in contact.

An alternative solution in the case of airplanes is to introduce quotas on the cause of the noise. Notice that the location of ears close to airports is as much a "cause" of the noise as are the engines of the airplanes. On these grounds one could justify ordinances (which in fact exist) to limit the building of houses close to airports. It is more usual to restrict airplanes to certain approach paths and times of day (as in Canada, where all airports close at 10 P.M.). These are quotas on the output of airlines, similar to those that ban automobiles from the centers of many British cities or that restrict the amounts of untreated sewage that may be dumped in many American waterways.

Most economists believe, however, that the better alternative is not to enforce quotas but to impose a tax on the airplanes in an amount that brings the marginal private cost of a landing into equality with the marginal social costs. A landing that takes place despite the tax, assuming that the tax reflects correctly the amounts residents would be willing to pay to avoid the noise, is evidently worth more to the airline (that is, its passengers) than it is worth to the people on whom the noise falls. By contrast, quotas would achieve this equality of marginal private and social cost crudely if at all—a party of economists willing to pay a large amount to escape from Toronto after 10 P.M. would not be able to do so. A tax on noise, as on air pollution by a steel mill or water pollution by a municipal sewerage department, would appear to result in the optimal amount of pollution.

Flaws in the Solution by Tax: The Coase Theorem

The argument seems attractive, but in its usual expression has a flaw. The tax usually contemplated is one equal to the damages to the ears of residents from the landing of the airplane. After all, this will present to the airline the full social cost of a landing. Suppose that the value of a landing at Kennedy Airport is \$500 to the passengers, that a landing by an airplane without equipment to abate its noise causes \$450 worth of damage to the residents (in other words, had they no other options they would pay \$450 rather than tolerate the noise), and that the installation of abatement equipment on an airplane costs \$400 per landing. If the state imposes a pollution tax of \$450 per noisy landing (equal to the hurt from a noisy landing), the airline will install the abatement equipment, spending the \$400 rather than sacrifice the \$500 it can earn from a landing or pay the tax of \$450 for the privilege of landing noisily. But suppose—this is the nub of the issue—that residents can avoid the \$450 worth of hurt at a cost of only \$200 per landing to themselves, the \$200 being the cost of insulating their houses, say, or moving away. *This* would be the socially desirable event, at least in terms of efficiency. National income will be highest when airplanes land without abatement (saving the airlines and society the \$400 opportunity cost of such equipment) and residents devote resources worth \$200 in some other occupation to avoiding the noise. This is event 3 in Table 16.1. It is assumed that the airline pays the tax only if noise is heard (that is, only in event 4 in the table).

There are two possible outcomes. If, on the one hand, the costs of arranging deals between the airline and the residents are very low, then the airline faced

Table 16.1
The Opportunity Costs of Landings and Nonlandings

Events	Output Forgone:	
	By Airline	By Residents
1. Do not land	\$500	0
2. Land, but airplane noise abated	400	0
3. Land, but residents' houses insulated	0	\$200
4. Land, with no control of perceived noise by either airlines or residents	0	450

with a \$450 tax when noise is heard will be able to bribe the residents to insulate their houses (paying them, say, \$201 to do so). The noise will not be heard and the airline will not have to pay the tax. In other words, if transactions costs are low, the exchange of rights will lead to event 3, the social optimum: Smith's generalization revisited. If, on the other hand, the costs of arranging deals are high, then the airline will not be able to bribe the residents and will be driven to install the abatement equipment at a social cost of \$400—event 2. Setting the pollution tax at the value of the hurt from the pollution does *not* necessarily achieve the social optimum. This last proposition is known as the *Coase theorem*: In the presence of transaction costs the location of a pollution tax or of other liability for damages *does* matter for efficiency.³

Setting the Tax Correctly in View of Coase's Theorem

A number of insights follow from this analysis. Pollution taxes are usually not in fact arranged to cease when the nuisance ceases to be felt. In the present case, the airline might be required to pay \$450 whether or not residents insulate their houses, giving the airline no incentive to bribe the residents into the social optimum. Further, were the tax set at the true opportunity cost of avoiding the noise—namely, \$200—the social optimum would be reached, because the airline would land the planes without abatement equipment and the residents, facing \$450 worth of peace and quiet sacrificed, would insulate their houses. But economists in pursuit of an equalization of marginal social costs and benefits have been prone to set the tax at the damage done.

T or F: In performing a cost-benefit analysis of a new steel mill, one must count as a social cost the damage to the health of the neighborhood from smoke pollution.

A: One must determine the least-cost way of having steel and health (with due regard to the marginal valuation

of each), which may well imply, for example, letting the smoke drive residents away from the mill. Therefore, false. This determination is a considerably more difficult one than simply measuring the damage from smoke (itself not simple).

Looking at the "Cause" of Pollution in the Light of Coase's Theorem

The most valuable insight of the Coase theorem, however, is the following: The technological, legal, or moral "cause" of some damaging externality is not necessarily the correct location for liability for the damages. The airline is usually viewed as causing the noise, but it has just been shown that placing the responsi-

³ Ronald H. Coase, "The Problem of Social Cost," *Journal of Law and Economics* 3 (October 1960): 1–44, especially the last few pages. Coase's theorem is often mixed up with Adam Smith's generalization. But we already knew the generalization.

bility for compensating for the noise on the airline (instead of on the residents) may lead to an inefficient result. Of course, it may not (if noise abatement equipment is cheaper than insulation). That jet engines produce noise, however, is irrelevant to determining on whom the liability for the noise should lie. That there are ears in the neighborhood of the engine is equally relevant to the issue, that is, not at all. A morally repulsive case will make the point clear.

T or F: On grounds of efficiency alone, it is unclear whether or not a drunken and reckless driver who injures a child should be held responsible for the crime.

A: Making the child responsible (that is, demanding that the child avoid the car rather than that the car avoid the child) may result in a combination of consumed

alcohol and surviving children superior to the one that would result from the opposite assignment of liability. Therefore, true. It might be desirable to excuse the person morally responsible. The judgment is of course appallingly amoral. Efficient arrangements are not necessarily moral, nor are moral arrangements necessarily efficient.

This is a case of high transaction costs. The child and the drunk cannot at the moment of the accident sit down to negotiate a price at which the child will agree not to run across the street in return for the drunk's undertaking not to drive the car on the sidewalk. Low transaction costs make the placement of liability inconsequential, for the same reasons that placing a tax on suppliers has the same result as placing the same tax on demanders.

What is critical is a cash connection. For example, high wages for dangerous work (such as coal mining) serve the same function as placing the liability for accidents in the mine on the mine owner.

Q: The liability for accidents in coal mines can be placed on either the miner or the owner of the mine. That is, either the miner or the mine owner can be required to pay the hospital bill. *True or false:* If the owner of the mine is made liable for accidents, the owner will definitely spend more preventing them than if the miner is made liable.

And, to reiterate, the moral or legal position is irrelevant to the efficiency of the solution.

T or F: On grounds of efficiency, mine owners should be made liable for all accidents in their mines.

A: If the cash connection operates cheaply, it may not matter one way or the other. If it is expensive, making

A: The owner may, but may not. If the owner spends less on safety, the miner will demand (and get) higher pay, higher by the amount that it costs the miner to insure. Therefore, false.

the owner liable may be inefficient. For example, miners may have more incentive to follow safety regulations if they themselves pay for accidents. In short, false.

Even in the presence of a cash connection, however, ambiguity in the placing of liability can itself lead to inefficiency.

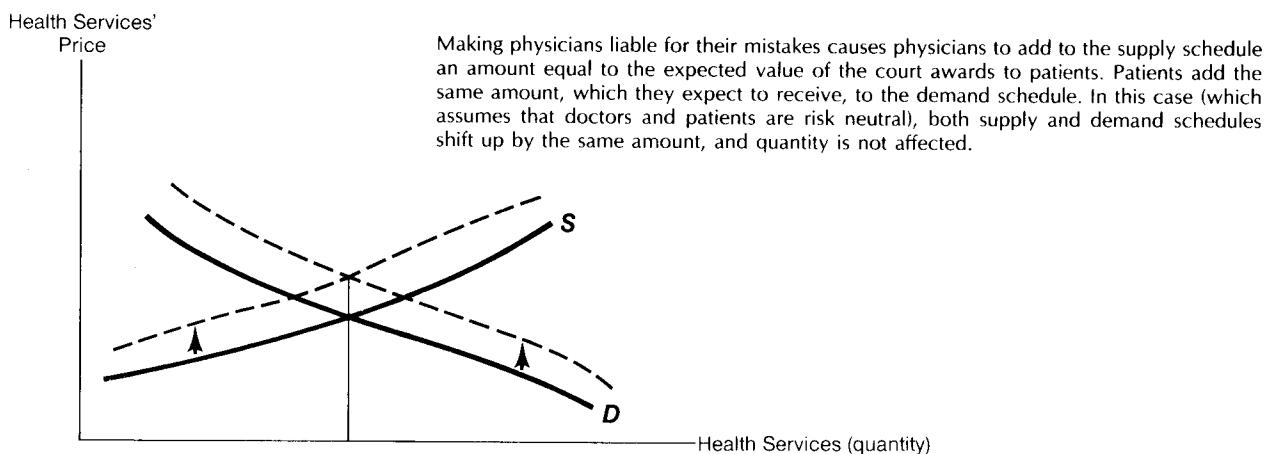
Q: Until recently it has been nearly impossible to sue one's doctor for negligence (cutting off the wrong leg).

1. What would be the effect on the supply and demand curves for medical services if it became very easy to sue doctors for negligence? The liability for mistakes, in other words, would be fixed firmly on doctors. What

would be the effect, if any, on the health of the nation?

2. Suppose instead that the legal rules of what constituted "negligence" were muddy, with liability not fixed firmly on either the patient or the doctor. What would be the effect on the nation's health and other income?

Figure 16.6
The Location of Liability May Not Matter If Transaction Costs Are Low



A: 1. The key point is that both the supply and the demand curve move. The supply curve moves backward if doctors are made liable because doctors must now be more careful (which is presumably more expensive than taking less care) or must buy insurance to cover their mistakes. At the same time the demand curve moves outward because the probability of mistakes becomes smaller or the desirability of carrying insurance against doctors' mistakes becomes smaller. The shift in liability amounts to a shift in the location of a tax, and it is therefore unclear as to whether or not there

will be any change in the quantity of health purchased (see Figure 16.6).

2. A muddy definition of liability may be worse than a clear definition. It is certainly worse for the income of the nation aside from health, because resources are devoted to adversary proceedings in the courts to establish liability in each case. And it may be worse for health, because the uncertainty of the result of these proceedings will lead both parties to insure, pushing back both the demand curve and the supply curve.

Why Property Rights Are as They Are

The argument so far has concerned the welfare economics of property rights, that is, the goodness or badness flowing from the creation of property in whales, roads, clean air, safety, and other things. In the spirit of the story of rent creation and dispersal in the last section, however, it is possible to give an account of why in fact property comes to be created in the first place. The gist of the account, as imagined by economists over the last 20 years or so, is simply that property will be created when it is in someone's self-interest to do so. In most places no one objects if you pick up a small rock and carry it away. Although the rocks on my land are formally my property, it is not worth my while to define and enforce my property so stringently that alarms ring and lawyers come running whenever someone lifts a rock from my front lawn. If the small rock is a piece of petrified wood in the Petrified Forest, however, it is worth the while of the park rangers to erect threatening signs in an attempt to keep people from taking the forest away piece by piece. If the small rock is a gold nugget in the California gold rush, it was worth the while of the miners to define and enforce property rights in mining claims with great vigor, even in the absence of an outside government to do the enforcing.⁴

⁴ John Umbeck, "The California Gold Rush: A Study of Emerging Property Rights," *Explorations in Economic History* 14 (July 1977): 197–226.

A piece of property, then, is not merely a thing but a social relation. Many things are not property, not "appropriated" (literally, "made one's own"). Air in Iowa City, for example, is not appropriated. It is in no one's self-interest to set up a stand selling bottled air (bottled pure oxygen is another matter, of course); nor is it in anyone's self-interest to engage in expensive legal suits over who is breathing whose air. In Los Angeles in the summertime, on the other hand, it may be worth the trouble to define and enforce rights to *clean* air.

Q: Most of the land in what is now the United States and Canada was not appropriated in the European sense before the Europeans came. The fact might be attributed to some peculiarity of native American culture, although the great variety in the culture makes the argument dubious. *True or false:* An economist, however, would find the low density of population of most places then a more persuasive explanation.

A: At low densities of population, the land is not worth much. It is not scarce. Therefore, it is not worth the trouble to define and enforce ownership in it. Therefore, true. Some of the social relations of the land *were* made explicit, namely, those that were valuable even with a low density of population, such as hunting rights over a certain range.

A particular physical thing, such as an acre of land, may have associated with it property rights at all stages of development, from full to none. And the rights in the same piece of land may be assigned to many different people. Land in so-called "open fields" in England eight centuries ago was owned for the most part by the peasant who worked it: The peasant had rights to the crops grown on it, for instance. But many rights to the piece of land were owned by others. The right to graze animals on the land after the crop was in, for example, was often assigned to the village herd, that is, to the rest of the village as a whole. Particular neighbors, likewise, had carefully defined rights to trespass at certain dates on the piece of land on their way to their own lands. The tangle of rights in the open field system was sorted out in the "enclosure movement," especially in the fifteenth, seventeenth, and late eighteenth centuries. Essentially, all the rights to one piece of land were assigned to one owner, creating a more modern-looking bundle of property rights.⁵

With such arguments and observations, the science of property (that is, modern economics) turns its techniques onto the study of the origins of property itself. The American institutionalists and the German historical school could argue truly in the early years of this century that modern economics had no theory of the origins of property and could therefore not claim to understand the long sweep of economic development. The study of property rights may some day answer the charge.

Summary Adam Smith's generalization asserts that, as a matter of logic, if everything is owned and if exchange is costless, then efficiency obtains. It will not matter in such a world where some right—or liability, the obverse of a right—is placed at the beginning of exchange. By its end (setting aside any effects through the distribution of income), the right will be found in the hands of the person who values it the most. Most "market failures" that have agitated economists

⁵ D. N. McCloskey, "The Persistence of English Common Fields" and "The Economics of Enclosure: A Market Analysis," in William N. Parker and Eric L. Jones, eds., *European Peasants and Their Markets: Essays in Agrarian Economic History* (Princeton, N.J.: Princeton University Press, 1975).

are from this perspective not market failures at all, but failures to apply the market. If whales were owned, the argument says, they would be hunted at the optimal rate, which may be zero. The argument applies to the externalities of smoke pollution as well, except that in this and many other similar cases the solution of creating a property right in clean air, or in whatever is being overused, is difficult.⁶

The obvious solution is to have the government impose a tax that will simulate the result of a properly placed rights to air. Likewise, the government could bring charges against polluters or encourage citizens to bring suit, placing the liability where it appears to belong—on the cause of the pollution, technologically speaking. But technologically speaking is not economically speaking. The obvious “cause” of pollution is not always the best place to put the liability or the tax burden. If the costs of making deals were low, the party that could most cheaply avoid the hurt from the pollution could be induced to do so. In such a case it would not matter for efficiency where the liability was placed. Even if it were placed incorrectly, Adam Smith’s generalization would assure that it would find its way to the correct party. But if the costs of making deals are high—as they are, say, between a refinery and the neighborhood’s being polluted by its smells—then Adam Smith’s generalization does not hold and Coase’s theorem comes into play: If costs of making deals are high, then it does matter where a liability is placed.

All this is in aid of recognizing that clearly defined property rights underlie well-functioning markets. How the rights came to be defined clearly is an historical question, only recently taken up seriously by economists. The answer to the question that modern economics would like to give is that rights became clear when it was in someone’s interest to make them clear. In particular, very scarce things, such as land in a heavily populated area, would give sufficient incentives to preserve it by creating and enforcing property rights. Historical research will determine whether the economists’ theory is correct. It can only be remarked now that it fits with the rest of economic theory, such as the theory of rent seeking outlined in the previous section and the theory of the firm outlined in previous chapters.

EXERCISE FOR SECTION 16.2

1. For each of the following externalities describe the missing private property that is causing the externality. For instance, the hunting of whales to extinction is caused by no one owning the whales:

- a. Racket from late-night fraternity parties.
- b. The near extermination of African elephants for ivory.
- c. The trampling down of new grass in a common pasture overburdened with cattle. This is known as the Tragedy of the Commons, and is supposed to have occurred in English villages before the eighteenth century.
- d. Dangerous crowding in the shipping traffic of the English Channel.
- e. Offensively designed office buildings.
- f. Congestion costs caused by too many people joining the line in front of Easton’s Eatery at lunch.

⁶ Nonetheless, the Environmental Protection Agency has moved toward the goal of allowing free bidding for the air by proposing to create “pollution rights,” which a polluting factory must buy from another, earlier factory if it is to set up its factory in the airshed.

PROBLEMS FOR SECTION 16.2

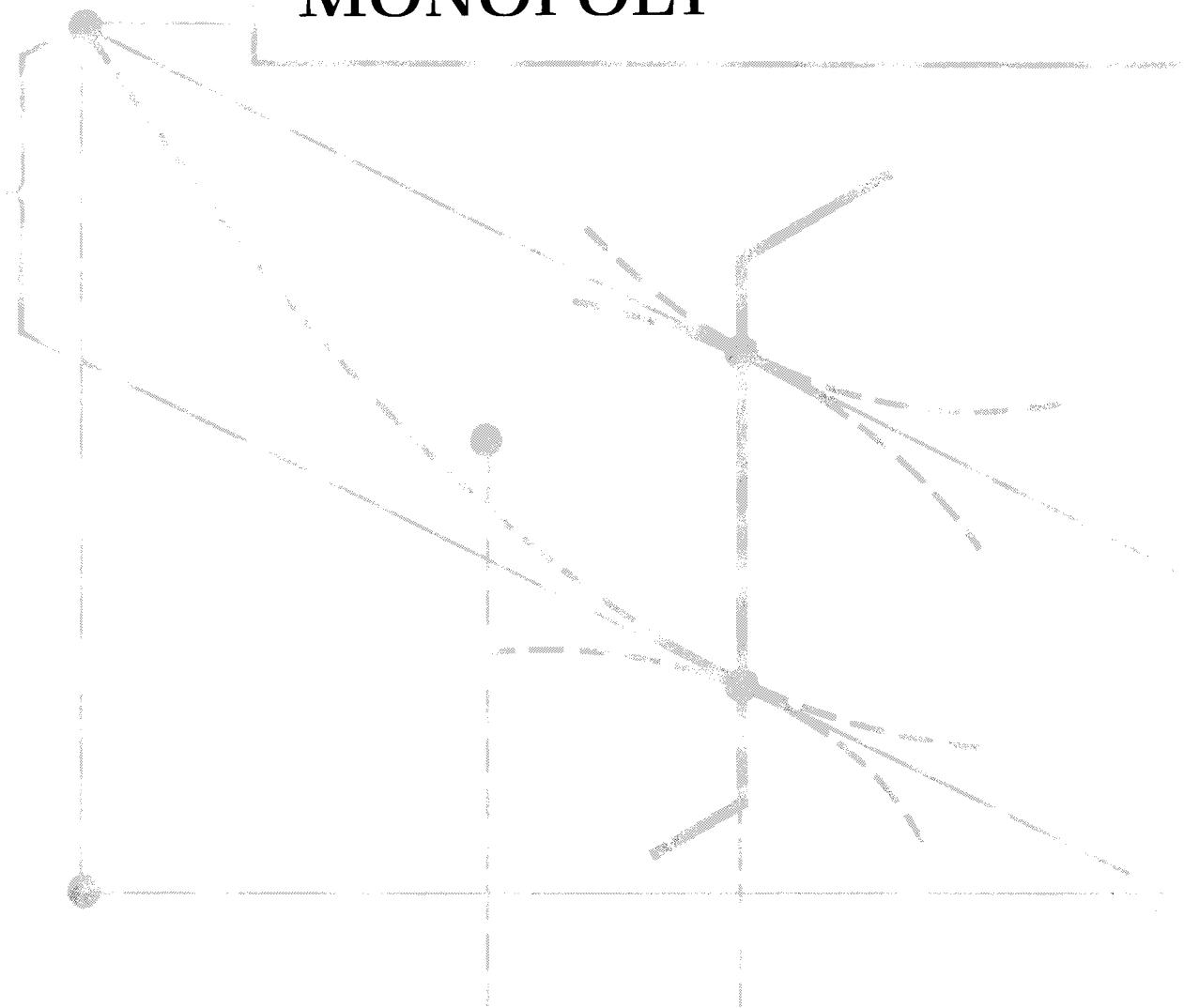
1. It is observed that the citizens of large industrial cities complain about air pollution more than do the citizens of isolated company towns, with a single mill. *True or false:* An explanation is that in the small company towns, facing a fixed price for their product and a fixed price of labor (laborers being mobile in and out of the town), the company has an incentive to adopt air pollution devices to the socially optimal degree.
- o 2. National income would be higher if the iron ore of Minnesota were located next door to the coal of Pennsylvania or of southern Illinois; or if oranges grew in Times Square instead of Miami; or if, in general, production happened to be located close to consumption (or, equivalently, if transport costs were zero).
- Exhibit the truth of this proposition. (*Hint:* Think of transport costs as a tax on supply.)
 - Explain why “transport” costs of moving the right to use a cubic yard of clean air from one person to another, or moving capital from low- to high-valued uses, or moving risk-bearing from very risk-averse to less risk-averse people are all analogous to ordinary transport costs.
 - Each of the “transport” costs in (b) is called a “market failure”: in order, an externality, an imperfection in the capital market, and a failure of the insurance market. The notion is that a market that “fails” warrants governmental intervention. In view of the analogy with transport costs, what do you think of the terminology of “market failure”? What is being assumed about the costs of governmental decision making and allocation?

True or False

- o 3. California is beautiful and has a large number of magnificent public parks. Therefore, California is likely, from the strictly economic point of view, to be overpopulated.
4. If the people already on the Dan Ryan could bid against newcomers for the property right in the road, then the externality would be eliminated.

V

MONOPOLY



3/4 of price theory comes from anti-trust

CHAPTER

17

The Behavior of Monopoly

17.1 Monopoly: The Elements

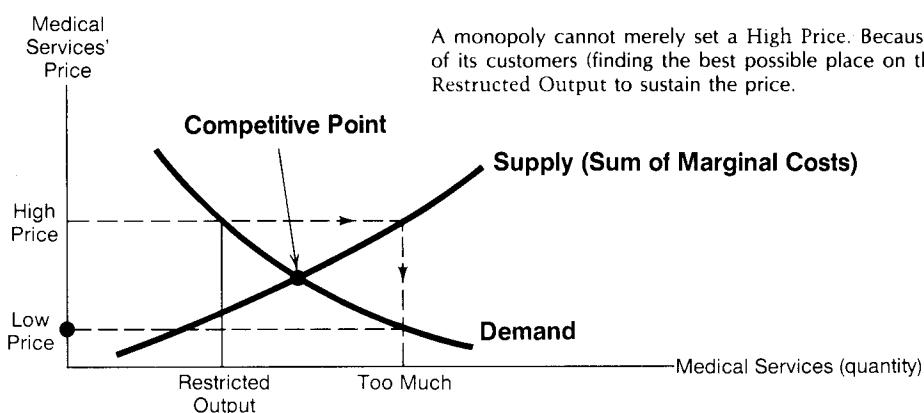
What to Read For Can a monopolist just announce a higher price? How do unions, professional associations, and governmental agencies act as monopolies? What is monopsony? What is price taking and price searching? How does a higher output spoil the market of a monopolist? What is the algebraic relation between marginal revenue, price, and spoilage? How do you construct a marginal revenue curve for a straight-line demand curve? What areas represent total revenue? How does a monopolist with no costs follow the Rule of Rational Life? Does a monopolist produce the socially optimal output? Does a monopolist ever produce along an inelastic segment of his demand curve?

Monopolies Must Restrict Output

Adam Smith observed that "people of the same trade seldom meet together, even for merriment and diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices." His observation is not to be wondered at, because, as he observed elsewhere, "it is 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." Their own interest is to raise the prices, and if their mutual competition drives the price down to marginal cost, it is no wonder that their thoughts turn to raising the price by conspiracy or, a still better contrivance, by law.

A conspiracy, cartel, central selling agency, combination, corner, exclusive franchise, marketing board, pool, professional society, public utility, regulated industry, syndicate, trade association, trade union, trust, or monopoly can raise its price only by restricting the output supplied to demanders. If at a new High Price the members of the European Coal and Steel Community or the American Medical Association (known to its friends as the AMA) attempt to sell more than their customers wish to buy at that price, the price will fall (see Figure

Figure 17.1
The AMA Must Restrict Output to Raise the Price



17.1). Each member of the AMA, facing the High Price for medical services and knowing that his or her individual decision will not affect it, has an incentive to cut the price a little or, equivalently, to offer a larger quantity than the proper share of the Restricted Output. To be precise, summing each of these decisions, the members have an incentive (namely, the incentive of more profit for each cheater on the agreement) to offer the quantity Too Much. But if they do so, the price will fall, initially all the way to Low Price and eventually back to the Competitive Point.

To prevent such a distasteful result, the AMA as an organization must find some effective way to prevent individual doctors from cutting prices. One way is to restrict the number of doctors to a number that will wish to offer only Restricted Output, less than the competitive amount. This the AMA, with the help of state legislatures, has been able to do. Since the Flexner Report of 1910 (which recommended that the number of medical schools be cut) and subsequent enactments (which recommended that the AMA have charge of the cutting), the number of doctors relative to the population has fallen. The demand curve for medical care on the other hand has moved out: Real income per head has more than doubled and medical care has a high income elasticity; subsidized medicine (Medicare) and medical insurance (Blue Cross) have moved the demand curve out still farther and made it less price elastic. The result is apparent in the incomes of doctors. From 1939 to 1959, for example, the average doctor's income grew two-and-a-third times faster than did the average manager's income and three-and-a-third times faster than did the average industrial employee's.¹

Entry Must Be Blocked

Such are the rewards of monopoly. The key to the rewards is a limitation on entry. Without a way of punishing interlopers (such as shooting them, the method favored by monopolies of gangsters), or a patent (such as Polaroid had on the taking of instant photos), or a crushing natural advantage (such as an expensive railway line is said to have once it has been built), or a law (such

¹ L. E. Davis and D. C. North, *Institutional Change and American Economic Growth* (Cambridge: Cambridge University Press, 1971), pp. 204–208.

as those that enrich undertakers by requiring embalming of bodies even if they are to be cremated), monopolies cannot survive.

Q: Chicago Local 546 of the Amalgamated Meat Cutters & Butcher Workmen of North America limits membership in its union and compels supermarkets in Chicago to employ only its members. Until recently, furthermore, it prohibited its members from working after 6 P.M. (one could not buy freshly cut meat from union supermarkets in Chicago after 6 P.M.). *True or false:* The 6 P.M. restriction can be interpreted as an additional limitation on entry, raising the income of union members.

A: Without the 6 P.M. restriction, the union would not have full control over the number of hours its members supplied, for some members would be willing to work overtime or at unusual hours (for instance, at 6:01 P.M.). By restricting the total number of hours supplied to supermarkets, the union raises the wage that supermarkets are willing to pay for each hour. That is, true.

The key to monopoly, in other words, is stopping a buyer from buying elsewhere. If you buy first-class mail, you must buy it from the U.S. Postal Service. Each year postal inspectors in tan trench coats come around to investigate little children selling you delivery of Christmas cards in the neighborhood. The Service thus prevents you from buying service where you please. The AMA prevents you from buying doctoring where you please. The Amalgamated Meat Cutters union prevents you from buying meat cutting (and therefore meat) where you please.

Monopoly is, then, a restriction on the relation between a buyer and a seller. The analysis is similar and in some points identical to the analysis of taxes and other restrictions. The only difference is that the analysis of monopoly provides a theory of the origin of the restrictions and of the way in which they are exploited.

The theory starts by contrasting monopoly with competition: If buyers and sellers can pair up in any way they wish, and if there are many buyers and sellers, there is no monopoly. *Monopoly* means one seller (from the Greek, *monos*, meaning "single," and *polein*, meaning "to sell"). The seller is the one—one company, one union, one trade association, one licensing body—from whom the buyer must buy. Similarly, *duopoly* means two sellers, *oligopoly* means few sellers, and if you like this sort of thing, *polypoly* (rhymes with Tripoli) means many sellers (that is, competition). Likewise on the buying side, a single buyer is a monopsony: The federal government is a monopsony in the buying of atomic bombs and the single coal mine in a remote village in West Virginia is a monopsony in the buying of labor.

T or F: The "reserve clause" in major league baseball before 1975, which required that the players bargain for their yearly contracts with their present owner alone (and not with other clubs), conferred monopsony power on the owner.

A: The owner was the sole buyer of their services as ball players. Other clubs could not bid for Carl Yastremski or George Scott if the 1967 Red Sox gave them low pay. The rules of the leagues restricted the pairing

of buyer and seller, making the owner of the Red Sox the one buyer of the playing services of the Red Sox. The restriction was not slavery, because players had the option of leaving baseball entirely to become journalists or singers. Similarly, in general, monopoly or monopsony is not utter dependence, it is merely an advantage that one person has over another because luck, circumstance, or contrivance makes that person the exclusive buyer or seller. So, true.

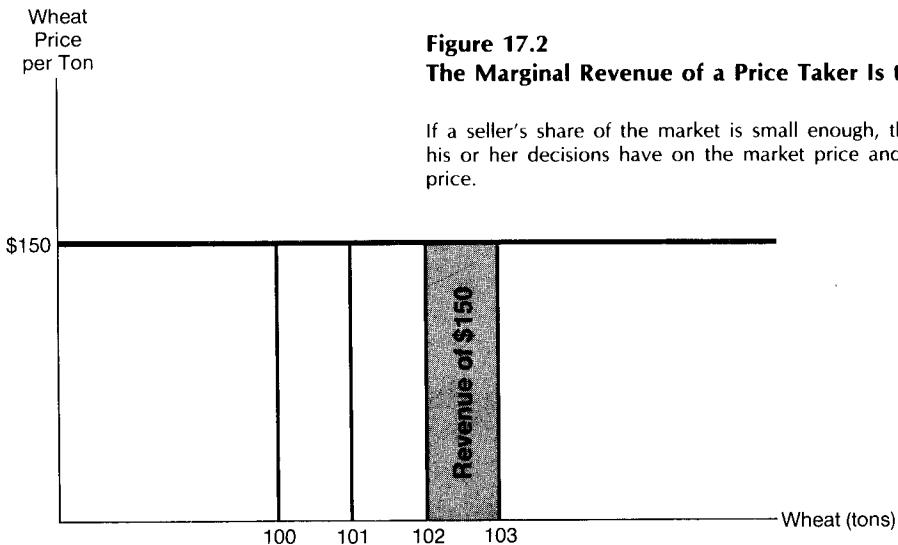
**The Result Is Price
Searching by the
Monopolist**

The exclusion of competitors creates a relationship between monopolist and victims quite different from the casual and anonymous relationship between buyer and seller in the grain pit of the Chicago Board of Trade or in the central food market of Hong Kong. No longer will competing suppliers rush in to fill the victim's demand at a price 1 cent above the competitive price. The monopolist is left alone with the victim. In this delicate situation, the question is, exactly how does the monopolist behave?

The monopolist's behavior is summarized in the phrase *price searching*. A single competitive firm among many is a price *taker*—taking its price as given by the market and marketing what it can of it. A monopolist, by contrast, *faces not a price but an entire demand curve* and is therefore able to search about for the best price to charge. The price that can be charged is fixed by the quantity that the firm offers for sale, so the analysis of a price searcher, like that of a price taker, can focus on the firm's choice of a quantity to offer. The Rule of Rational Life, of course, is to bring the marginal cost of an activity into equality with its marginal benefit, which maximizes the difference between benefit and cost. The monopolist, like the competitor, follows the rule.

Consider the benefit half. The marginal benefit (that is, the marginal revenue) to a competitive, *price-taking* seller of wheat of selling one more metric ton of wheat is of course the current market price of \$150 a ton. Thus, at 100 tons the marginal revenue of 1 more ton is the price; at 101 tons it is the price; at 102 it is the price (see Figure 17.2). The price does not fall when farmer Shlomowitz increases his output (since Shlomowitz is one of millions of farmers in the market), and each additional ton produces additional revenue to him that is equal to the market price.

But the marginal revenue to a *monopolistic* seller of telephone calls, postal services, photoreproduced copies, or exhibitions of the movie *Gone with the Wind* of selling one more call, stamp, copy, or exhibition is *lower* than the



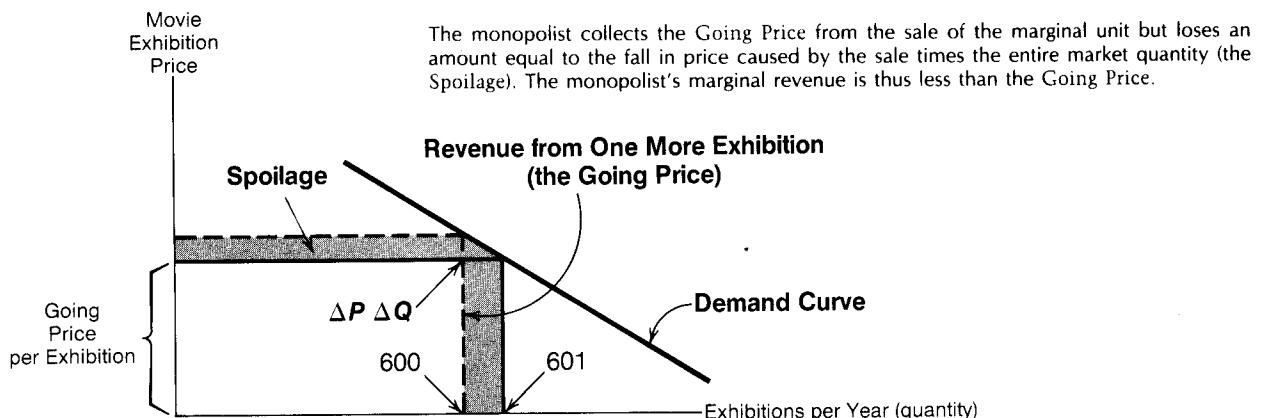
current price. Because the monopolist faces a downward-sloping demand curve rather than a given price, the one additional exhibition of *Gone with the Wind* (on national television, say) reduces what people are willing to pay the owner for "earlier" exhibitions (not earlier in time but earlier along the quantity axis and with higher marginal valuation). As the phrase goes, allowing another exhibition to some degree *spoils the market*. The loss of revenue on earlier exhibitions must be subtracted from the price to get the net *marginal revenue*. The marginal revenue of one more exhibition is the going price minus the spoilage. The algebraic way of saying this, which will work wonders later, is that marginal revenue equals $P - Q \Delta P$. In words, the equation is price minus spoilage (as in Figure 17.3).

Suppose for instance that exhibitors were willing to pay \$1000 for each of 600 exhibits of *Gone with the Wind*, and that their willingness to pay—their demand curve—varied as follows:

Exhibits per Year	Price Paid per Exhibit	Total Revenue (exhibits times price)	Marginal Revenue (change in total revenue)
598	\$1002	\$599,196	
599	1001	599,599	\$403
600	1000	600,000	401
601	999	600,399	399
602	998	600,796	397

The price falls as more exhibitions per year are put on the market. As it happens (it can happen differently), the price doesn't fall enough to entirely offset the extra revenue from more exhibitions. That is, total revenue rises. So marginal

Figure 17.3
The Marginal Revenue of a Price Searcher Is Less Than the Price



revenue, which is the change in revenue caused by a one-unit rise in quantity, is positive. But it's less than price: \$401, for instance, is less than \$1000. The move from 600 exhibitions to 601, imitating Figure 17.3, can be broken down as follows. One exhibition at the new going price is worth \$999. The Spoilage is the old quantity (600) multiplied by how much the price fell when moving from 600 to 601 exhibitions (namely, by \$1). So marginal revenue is \$999 - (\$1)(600) = \$399, as in the table.

The Algebra of Marginal Revenue

If the words and diagrams and arithmetic do not convince you that the equation is true, then you deserve the following algebra. The meaning of marginal revenue is the new revenue minus the old revenue, these being the prices (new and old) multiplied by the quantities (new and old): $MR \equiv P_1 Q_1 - P_0 Q_0$. Another way of expressing this comes from replacing the new price and quantity with the old plus the change from old to new (or in the case of price, minus the change, because price falls): $MR = (P_0 - \Delta P)(Q_0 + \Delta Q) - P_0 Q_0$. Multiplying out this expression leads to a simpler expression, by cancellation of the term $P_0 Q_0$: $P_0 Q_0 - Q_0 \Delta P + P_0 \Delta Q - \Delta P \Delta Q - P_0 Q_0 = -Q_0 \Delta P + P_0 \Delta Q - \Delta P \Delta Q$. The last term, $\Delta P \Delta Q$, is very small when ΔP and ΔQ are small (it is the small triangle in the diagram). So call it zero, approximately. The change in quantity, ΔQ , is one unit, 1.0, leaving the expression as $P_0 - Q_0 \Delta P$. For small changes, it is not going to matter whether P_0 or P_1 or some intermediate going price appears in the equation, since all will be approximately the same. Likewise for Q_0 . Therefore drop the subscripts, leaving $MR = P - Q \Delta P$. Q.E.D.

Now ask yourself the following, by way of review.

Q: Is marginal revenue above or below the going price?

A: The equation of marginal revenue can be rearranged to say that marginal revenue plus spoilage is the demand

price: $MR + Q \Delta P = P$. Marginal revenue falls short of price (by the amount of the spoilage). In other words, the marginal revenue curve is always somewhere below the demand curve.

A Special Case: Straight Lines

With a straight-line demand curve the location of "somewhere" is simply described and makes straight-line demand curves especially simple for thinking about monopolies. Marginal revenue is the straight line that runs from the point where the demand curve cuts the price axis to (and through) the point midway out to where the demand curve cuts the quantity axis (see Figure 17.4).² That the marginal revenue of a straight-line demand curve intersects the quantity axis—that is, that marginal revenue is zero—at the midpoint is no surprise. At that quantity of exhibitions, recalling Chapter 7, the elasticity of the demand curve is 1.0. That is, total revenue remains constant as the quantity offered is changed a little: The elasticity of 1.0 means that price falls by the same percentage

² It is straightforward to show this with calculus. If the demand price is a linear function of the quantity—say, $P = A - BQ$ —then total revenue at any Q is $PQ = (A - BQ)Q = AQ - BQ^2$. Marginal revenue, then, is $dPQ/dQ = A - 2BQ$. The equation reveals that the marginal revenue curve is also linear, that the demand and marginal revenue curve have the same intercept on the price axis (that is, A), and that the intercept of marginal revenue on the quantity axis $\frac{1}{2}(A/B)$ is half the intercept of the demand curve A/B .

Price of an
Exhibition

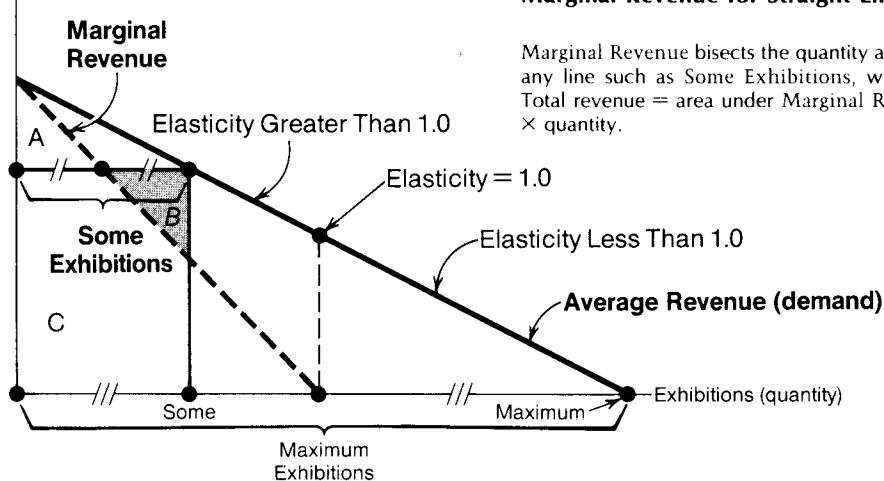


Figure 17.4
Marginal Revenue for Straight-Line Demand Curves

Marginal Revenue bisects the quantity axis out to Maximum and also bisects any line such as Some Exhibitions, which is parallel to the quantity axis. Total revenue = area under Marginal Revenue = $A + C = C + B$ = price \times quantity.

as quantity rises. And to say that total revenue is constant is to say that marginal revenue is zero.

If the marginal revenue curve is a straight line that bisects the distance Maximum, it evidently also bisects the line Some Exhibitions (and similar horizontal lines), which implies that area A is equal to the shaded area, B , since A and B are then identical triangles. This is a general feature of a curve marginal to another (as marginal revenue is marginal to average revenue, the demand). Another way to see this is as follows. Because the marginal revenue is the change in revenue, the sum of all the changes from an output of zero to an output of Some Exhibitions in the diagram will give the total revenue. Diagrammatically, the summing measures the area $A + C$ under the marginal revenue curve. The area under the marginal revenue curve, in other words, is total revenue. But so is the average revenue (the demand price) multiplied by the quantity (that is, area $C + B$). Since these have area C in common and both add to total revenue, area A must equal area B , as asserted.

Straight Lines in the Case of One Buyer

Similar reasoning applies to a monopsonist (single buyer) facing a straight-line supply curve, such as U.S. Steel might face for skilled blast furnace operators in Gary, Indiana.

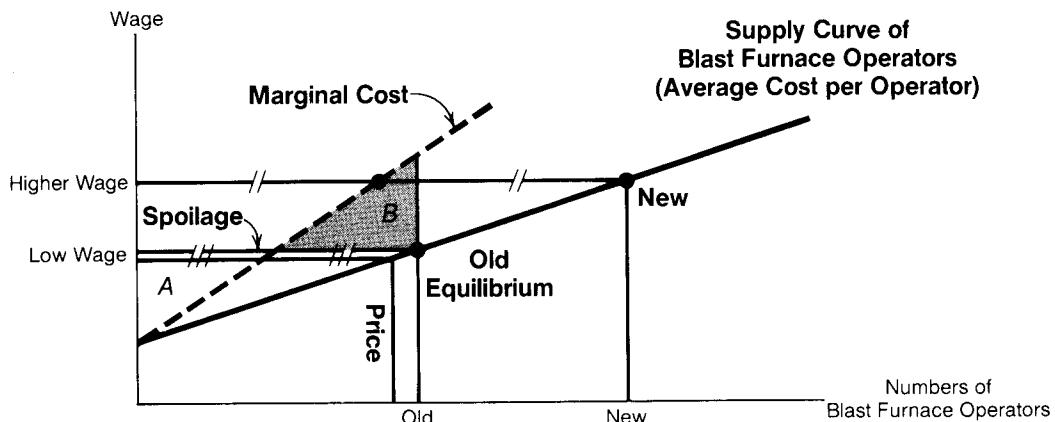
T or F: The marginal cost of labor of U.S. Steel's blast furnaces is higher than its average cost.

A: Look at Figure 17.5. If U.S. Steel decides to increase its work force from Old to New, it may pay a Higher Wage to get the increase and therefore spoils the Low Wage it was able to pay before (unless it can discriminate, all workers—old as well as new—must receive the higher wage). The marginal cost is the price (average

cost) plus the spoilage. U.S. Steel's marginal cost of labor, therefore, exceeds its average cost of labor. Therefore, true, the marginal cost is higher. In the same spirit as the construction of a marginal revenue curve, the marginal cost curve is constructed by bisecting a line such as that out to Old Equilibrium or out to New. Likewise, area A equals the shaded area B . For the one buyer (monopsonist) as for the one seller (monopolist), then, the marginal and average price are not equal.

Figure 17.5
The Marginal and Average Wage for a Monopsonist

If the Higher Wage paid to attract an additional worker must be paid to each worker, the monopsonist's marginal cost is greater than the average wage.



The Point the Monopolist Chooses Without Costs

The marginal benefit of a monopolist selling movie exhibitions or bridge crossings is half of the Rule of Rational Life. The other half, the marginal cost, involves no new principles (a monopolist is just a firm). Bringing the two together yields the equilibrium of the monopolist, which of course is verbally the same as the equilibrium for the competitor: Choose the output that equalizes marginal cost and marginal benefit. The only difference is that for a monopolist marginal benefit is no longer equal to price.

The simplest case of monopoly is one with no marginal costs, such as a monopolist's owning all existing Impressionist paintings or a patent on the air brake or a bridge from Brooklyn to Manhattan. None of these has, at present, costs of producing one more unit.

Q: The New York Port Authority owns the Brooklyn Bridge. If it wishes to maximize revenue from the bridge and believes that the demand curve for crossings on the bridge is linear, what price will it charge?

A: Following the Rule of Rational Life, it will charge the price that will cause the quantity of crossings to

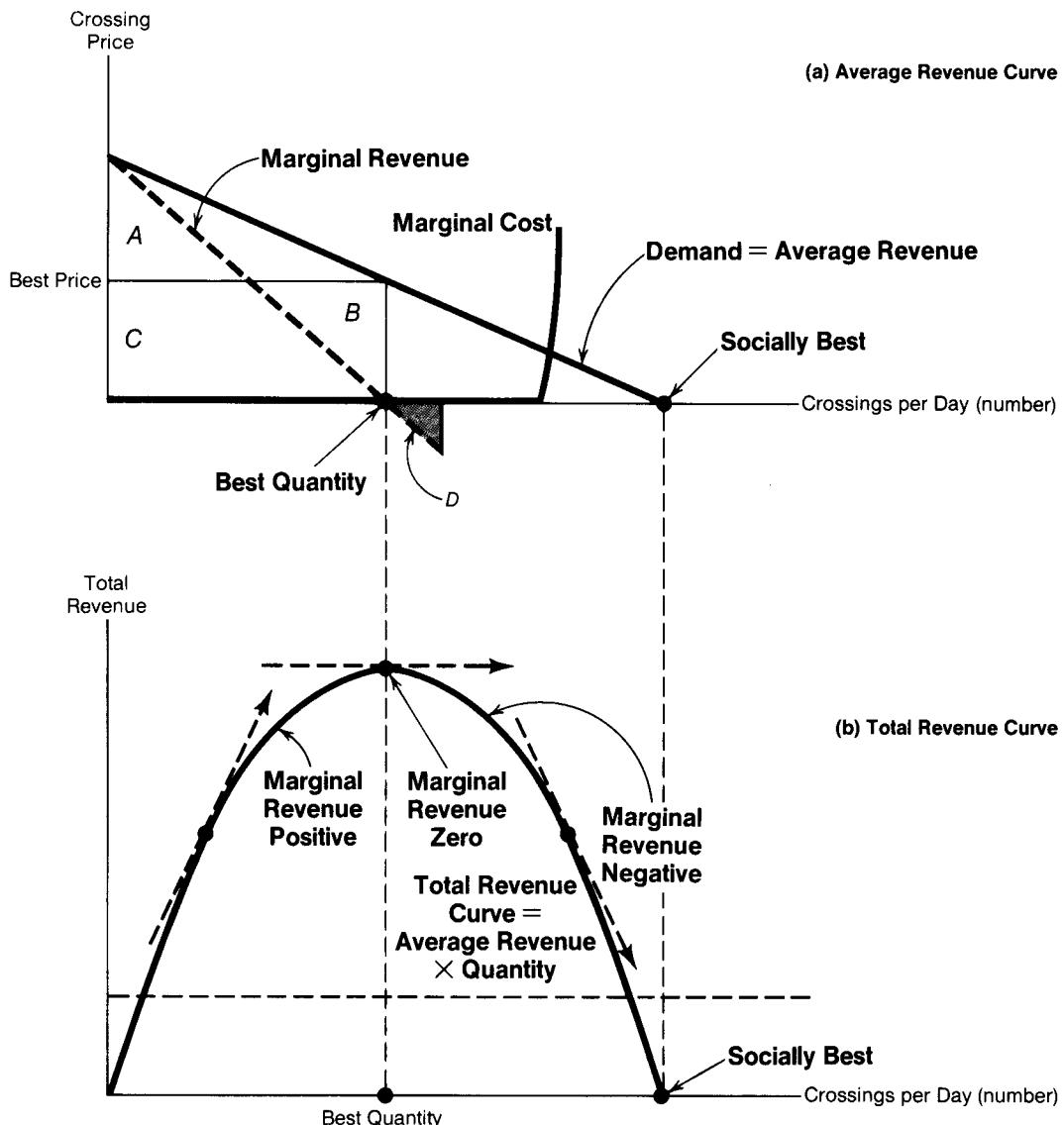
be such that marginal revenue equals marginal cost. Since marginal cost is zero, this quantity will be where marginal revenue is also zero, that is, where the marginal revenue line crosses the quantity axis, at the point Best Quantity in the top panel of Figure 17.6.

Notice that the revenue can be expressed either as the rectangle $C + B$ (the price charged multiplied by the quantity demanded at that price) or as the triangle $A + C$ (the sum of all marginal revenues out to Best Quantity, the area under the marginal revenue curve). The two are equal. Beyond Best Quantity, area $A + C$ is diminished by areas such as the shaded area D because marginal

Figure 17.6

A Monopolist with No Costs Arranges Marginal Revenue to Be Zero: Average Revenue Curve (a) and Corresponding Total Revenue Curve (b)

If costs are zero, maximizing net revenue is equivalent to maximizing total revenue. The monopolist chooses a quantity such that marginal revenue at the quantity equals zero. The same argument can be represented in (b). The point of maximum revenue is evidently the same as the point of zero marginal revenue.



revenue is there negative. Total revenue is maximized. The condition that marginal cost equal marginal revenue has evidently had the desired result, namely, maximum revenue. A price that encouraged people to make more than Best Quantity of crossings would reduce total revenue, for this is what negative marginal revenue means.

In other words, a monopolist with no costs maximizes total revenue. By the usual reasoning, if the monopolist had only fixed costs (such as interest payments on the cost of constructing the bridge), the result would have been the same. The argument is plainer in the bottom panel of the diagram, which gives the total revenue corresponding to each quantity. The point at which marginal revenue is zero is of course the peak of the hill. The presence of a Fixed Cost such as the dashed horizontal line would clearly not change the desirability of being at the peak. The peak would still represent maximum profit.

**Monopoly Is Bad
Because the Point
Chosen Is Inefficient**

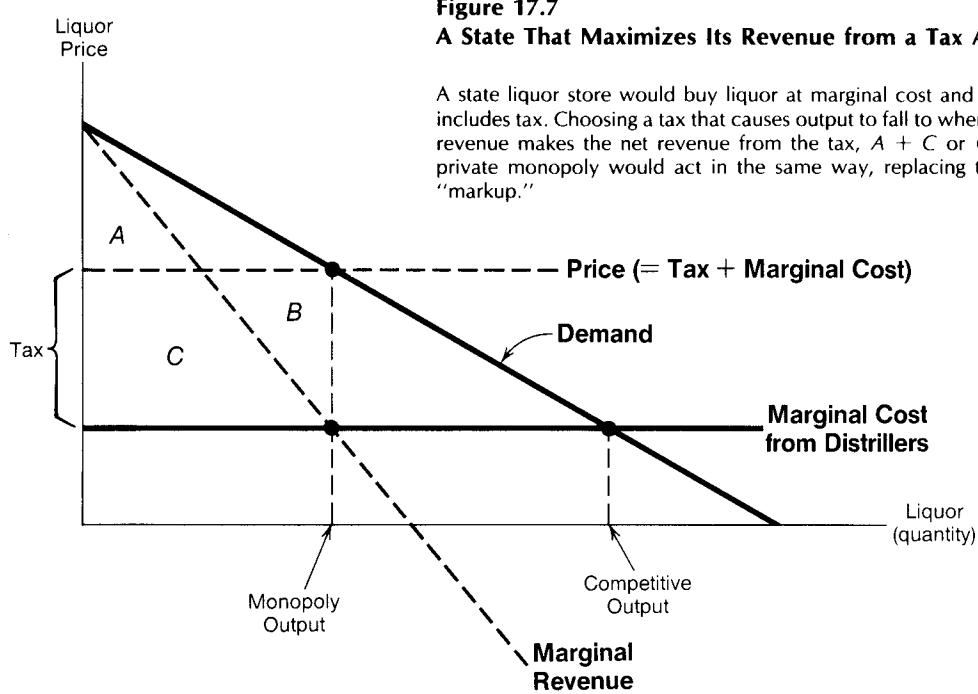
Figure 17.6 shows that the pursuit of profit by a monopolized industry, unlike a competitive industry, leads away from social happiness. The point of maximum profit for the bridge owner is Best Quantity. The point that corresponds to the socially best use of a bridge costing nothing to use a little more is Socially Best, where the marginal valuation of a crossing is equal to its true marginal cost, namely, zero. This argument in welfare economics is pursued at length in the chapters following, after the behavior of monopolies is explored more fully. In any event, it is not the profit or the arrogance of a monopolist that makes him socially obnoxious but *the monopolist's desire to sell a smaller than socially best amount*.

**The Point the
Monopolist Chooses
(with Costs, but
Constant Costs)**

The generalizations to a monopoly with costs are straightforward. For example, a government that taxes liquor can be viewed as a monopolist "buying" liquor from distillers at, say, a constant marginal cost and "selling" it to the drinking public at a higher price (in states such as Iowa or Vermont with state monopolies of the retailing of liquor the buying and selling is literal, not figurative). This is one of many examples of the monopoly model applied to subjects other than private monopoly. Governments are public monopolies, and when they pursue the maximum advantage from their position, the model applies. If the state of California agrees with John Stuart Mill that "taxation . . . of stimulants, up to the point which produces the largest amount of revenue . . . is not only admissible, but to be approved of,"³ it will set the tax on liquor at Tax (see Figure 17.7). The price will be the tax plus marginal cost, and at that price only the Monopoly Output will be demanded.

By reducing the quantity demanded to Monopoly Output, the state has brought its marginal cost and marginal revenue into equality and has maximized the net revenue from the tax. The net revenue can be viewed either as the rectangle $B + C$ (the tax per bottle multiplied by the number of bottles at Monopoly Output) or the triangle $A + C$ (the excess of marginal revenue over marginal cost on each successive bottle out to Monopoly Output, where the excess has fallen to zero).

³ John Stuart Mill, *On Liberty*, first published in 1859 (New York: Dutton, 1910), p. 156.



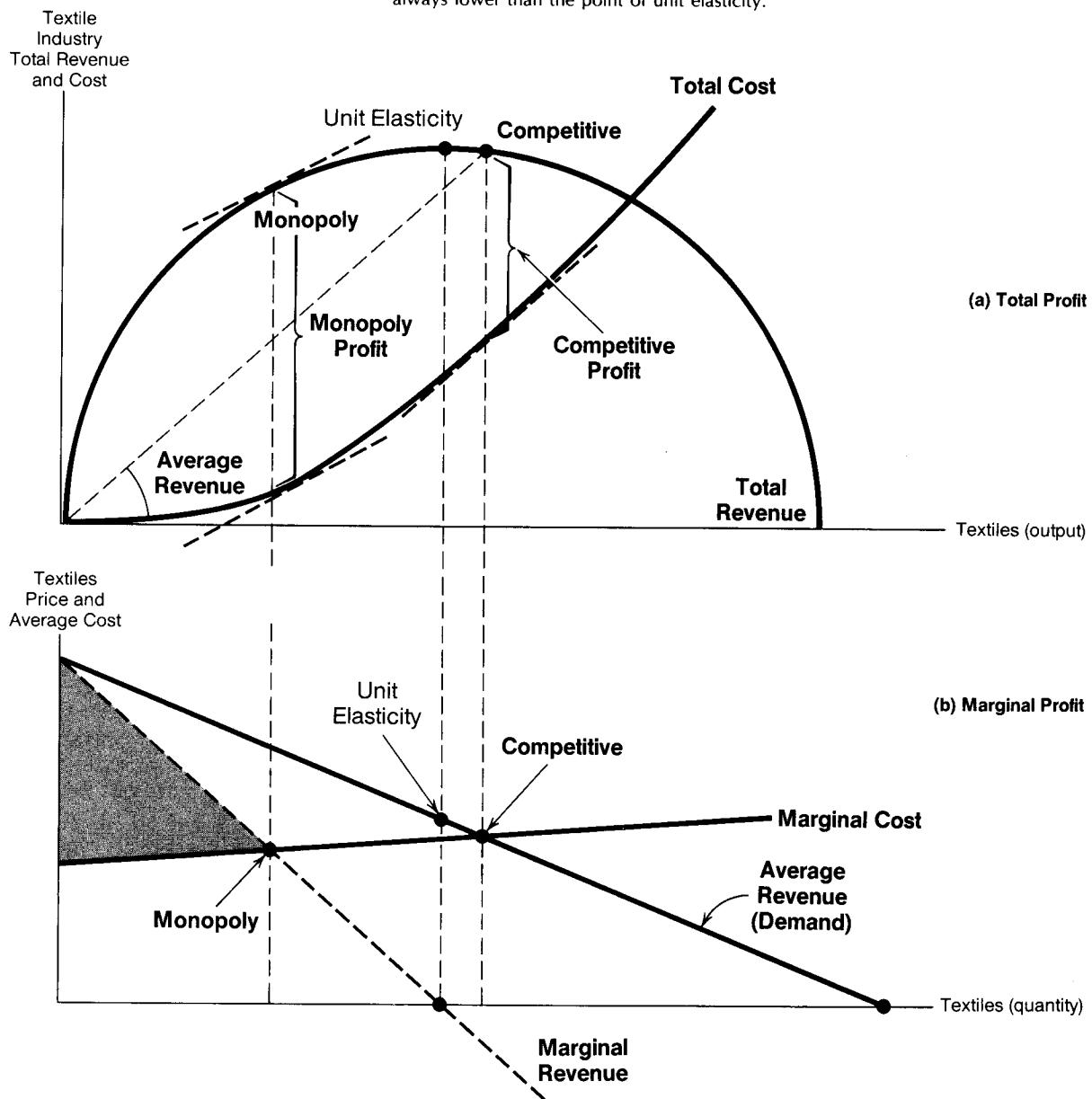
The Case of Varying Costs The still more general case is a monopoly with varying (instead of constant) marginal costs. Under the National Recovery Administration (NRA) of the 1930s, for example, the federal government sought to aid recovery from the Great Depression by setting up monopolies. Through its trade association, an industry would submit a code of fair practices, and the approved code signed by the president had the force of law. Hundreds of such codes were approved in 1933 and 1934, covering industries from burlesque theatricals (Code 348) and dog food (Code 450) to cotton textiles and soft coal.⁴ A code authority in, say, cotton textiles would increase or at least change cost by changing wages, hours, and working conditions, and the increase in cost would by itself reduce output. But the reduction came directly as well. By a 1933 order the textile machinery in any factory was limited to two 40-hour shifts per week (and in 1934 to 30-hour shifts). To represent the situation, Figure 17.8 flips the usual order of panels, putting total revenue on top. Taking costs and the number of firms in the industry as unchanged, and supposing that the demand curve is linear, the goal of the industry acting as a monopoly can be viewed as moving from the point Competitive to the point Monopoly.

Monopoly, you see, gives higher profit to the cotton textile industry as a whole (the total profit gap is larger at Monopoly than at Competitive). Competitors sell too much for their own collective good. At Competitive the marginal cost equals *average* revenue; at Monopoly the marginal cost equals *marginal*

⁴ William E. Leuchtenburg, *Franklin D. Roosevelt and the New Deal, 1932–1940* (New York: Harper & Row, 1963), p. 68.

Figure 17.8
Monopolies with Costs Choose a Quantity Smaller than Either Maximum Revenue or Competition: Total Profit (a) and Marginal (b)

The monopoly's task in (a) is to find the line that makes revenue minus cost as large as possible. This it does where marginal revenue (the slope of total revenue) equals marginal cost (the slope of total cost). The equality is duplicated in (b). The geometry implies that the best output for the monopolist is always lower than the competitive output and is always lower than the point of unit elasticity.



revenue. And an equality of marginals is better for the industry. The equality is clearest in the bottom panel, but it holds true in the top as well. A move to lower output brings more profit to the industry than does the Competitive point.

A Monopoly Operates Only in the Elastic Portion of Its Demand

The top panel, incidentally, makes unforgettable a technical condition on the monopoly: A monopoly will never choose a price at which the demand curve is inelastic (elasticity less than 1.0). The demand curve is inelastic (recall: steep in slope, roughly) when a rise in quantity along it causes such a large fall in price that total revenue goes down instead of up. You can remember this by reflecting that a negative condition such as *inelasticity* is naturally associated with *falling* revenues. The neutral point is Unit Elasticity (elasticity equals 1.0), the peak of the hill in the top panel and the midpoint of the straight-line demand curve in the bottom. Since total cost is rising, it is obviously silly for the monopoly to choose a point to the right, that is, on the downward slope of the hill, because it is impossible for something with an upward slope (cost) to have the same slope as something with a downward slope (the total revenue curve to the right of the point Unit Elasticity). So it chooses a point to the left. The demand curve at the point of equilibrium for a monopolist must be elastic (elasticity of demand greater than 1.0), as asserted.

Summary

Monopoly begins with a natural desire to earn more profit than is allowed under "cutthroat" or "ruinous" competition (as ersatz economics has it). More profit requires a higher price and a higher price requires a lower quantity. A successful monopolist restricts quantity, preventing others from supplying the customers. Because the monopolist prevents the customers from going to alternative suppliers, the monopolist faces the customers' demand curves, not a given market price. The monopolist's marginal revenue, therefore, is below the price, because additional output spoils the price on earlier output. To induce customers to purchase another unit of medical care or postal service, the price must fall. This applies to competition as well to monopoly. The difference is that the monopolist is one seller facing the demand instead of one among many, *and therefore recognizes the spoilage*.

The analysis of a monopolist's behavior is especially simple with straight-line demand curves. A similar analysis applies to a monopsonist, a single buyer facing the supply curves of the victims. A monopolist with no costs, such as a bridge authority, will if rational set price (and therefore quantity) at the point of maximum revenue. A monopolist with costs, such as a state taxing liquor or an NRA code authority in cotton textiles, sets price or quantity at the point that maximizes net revenue, that is, at the point at which marginal revenue and marginal cost are equal. These arguments hold in fact for all shapes of the curves involved. They lead to the central conclusions that monopoly output is lower than competitive output and that the profits from the monopoly output are higher.

EXERCISES FOR SECTION 17.1

1. For each of these points on demand curves facing a single seller calculate the total revenue and the marginal revenue. (Take care; it's easy to make a mistake in arithmetic). Say whether marginal revenue is positive, negative, or zero.

a.	Exhibits of <i>Gone with the Wind</i> per Year	Price Paid per Exhibit
----	---	------------------------------

598	\$1003
599	1001.5
600	1000
601	998.5
602	997

b.	Exhibits	Price Paid
----	----------	------------

598	1004
599	1002
600	1000
601	998
602	996

c.	Bridge Crossings per Day	Toll Charged per Crossing
----	--------------------------------	---------------------------------

98,039	\$1.02
99,010	1.01
100,000	1.00
101,010	0.99
102,041	0.98

d.	Bushels of Wheat Sold by Ralph Austen in a Year	Price Received by Austen at These Amounts Sold
----	---	--

49,800	\$3.00
49,900	3.00
50,000	3.00
50,100	3.00
50,200	3.00

e. Total Bushels of Wheat Sold by American Farmers (millions)	Prices per Bushel That Would Be Received at Such Amounts Sold
49,800	\$3.010
49,900	3.005
50,000	3.000
50,100	2.995
50,200	2.990

(Hint: Make sure that you express marginal revenue in dollars per bushel; otherwise it's not in the same units as ordinary price, and will not fit on the same diagram, or be useful in the same analysis.)

2. Would the owner of the rights to *Gone with the Wind* in Exercise 1(a) want to increase his exhibitions beyond 602 per year if he reckoned his marginal costs per exhibition (wear and tear on films, mailing) at \$70 per exhibition?
3. Would the owner in the case of Exercise 1(b) want to stay at around 600 exhibitions per year? What can you say about the number of exhibitions he would allow relative to 600? Larger or smaller?
4. Does it matter where the bridge owner of Exercise 1(c) sets the price? (Suppose she has no costs.) What is the elasticity of her demand curve?
5. Does Austen in Exercise 1(d) ever have any incentive to cut back his output to get a higher price?
6. If the marginal cost of wheat is \$2 a bushel, do American farmers in total [Exercise 1(e)] have an incentive to cut back output to get a higher price?

PROBLEMS FOR SECTION 17.1

- 1. In 1784 the British import tariff on tea was 119% and 5 million pounds of it were imported. The accountant of the East India Company reckoned that no more than a third of the whole British consumption was imported legally, the rest being smuggled. He was apparently correct, for tea was inelastically demanded, and in 1785 after the reduction of the tariff to a mere 12.5%, the amount of legal tea imported increased to 16 million pounds. In light of the implied elasticity of the demand for *legal* tea, and using a straight-line "demand curve" (after smuggling), what was the revenue-maximizing tariff rate? (Hint: Use the information to run a demand curve between the 119 point and the 12.5 point. Then determine where the tax take will be highest.)
- 2. The telephone company is a monopoly. Suppose that an empirical study shows that the elasticity of demand it faces is -0.7 . True or false: Something is wrong.
- 3. Professor Harberger calculates the welfare loss of monopoly in the American economy by assuming unit elasticities of demand in the monopolized industries. Professor Stigler argues that Professor Harberger's calculation is inconsistent because literal monopolists would not operate at such a point.
 - a. Briefly explain Professor Stigler's argument. Professor Harberger replies that the "monopolies" in question are not literally single sellers but, rather, a few. He points out that the elasticity of a whole industry's demand is therefore no measure of the firm's elasticity.

The behavioral economics and the welfare economics are in this case quite consistent, says Professor Harberger.

b. Explain Professor Harberger's argument.

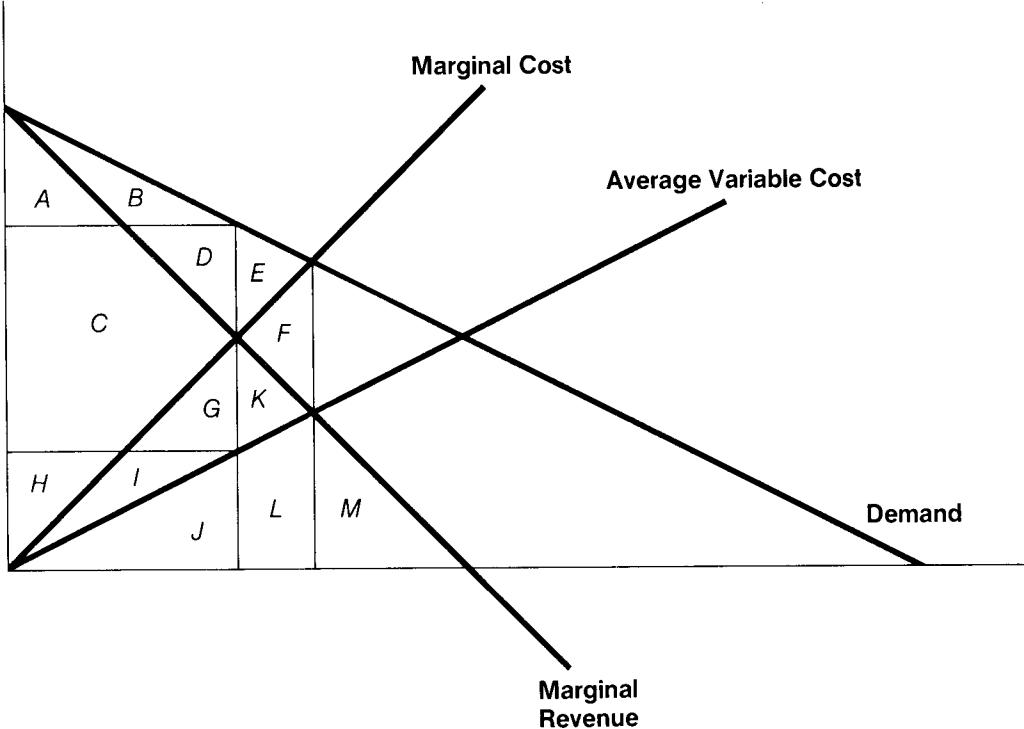
4. The racetracks in Illinois are scattered about, and therefore each has some monopoly power in its region. A state board assigns racing dates to each track each year, a process notorious for its corruption. *True or false:* No matter how many dates were already assigned to him, the owner of the track would always be willing to bribe the state officials to get additional dates.

5. Name all the areas in Figure 17.9 that equal the monopoly's net profits (putting the letters within each in alphabetical order).

True or False

6. If the British government in 1784 were attempting to maximize its own revenues, it would act in setting an import tariff on tea as though it were a monopolist selling to its citizens, having a marginal cost equal to the world price of the tea.
- 7. A tax of t cents on an item selling initially in a monopolized market at \$1 per unit will reduce the marginal revenue curve at the pretax output of the monopolist by more than will a tax of $t\%$. (*Hint:* Draw the diagram of average revenue and the corresponding marginal revenues.)

Figure 17.9
The Areas of Monopoly Profit



17.2 Advanced Applications of Simple Monopoly

What to Read For Does a monopolist always end up exactly where marginal cost equals marginal revenue? How does the output and price of a monopolist change if the monopolist has a price ceiling imposed on him? How can this idea be used to undermine cartels? What is a counterfactual? Can one compare monopoly and competition? Are monopolies always bad? Are patents, which are monopolies, always bad?

The Monopoly's Information Problem

The analysis of monopoly is an alternative to supply and demand. The alternative applies when one mind has lots of influence on either supply or demand. But the one mind doesn't know everything. Even if they pursued profit single mindedly, neither a monopolist nor a competitor would know price and cost perfectly, nor such subtleties as the slope of the demand curve. In an auction market such as that for hogs (competitive) or fine art (monopolized), the price can vary hourly. In a list price market such as that for shoes at retail or steel at wholesale, the actual price received will depend on spoilage on the shelf, discounts for special customers (credit card customers or long-standing customers), and other things. Costs of any enterprise are difficult to know with exactness. And even if the manager knew the prices and the costs exactly, she would have to search about for the point of maximum profit by trial and error. She would be lucky to find the point at which marginal cost exactly equals marginal revenue. And only monopolists who have had a course in microeconomics think of the problem in these terms anyway; without the course they just try to find the most profits. Like all theories of rationality the theory of monopoly is merely approximate.

A Monopoly Does Not Have a Supply Curve nor Does a Monopsony Have a Demand Curve

Having located in one's mind the point of most profit toward which a monopoly tends, one can examine how it changes. That is, one can perform "comparative statics" in the same spirit as watching in competitive markets the results of shifts in supply and demand or of impositions of price controls. The first lesson has been mentioned earlier, namely, that a monopoly faces a demand curve but does not itself have a supply curve. No single curve can tell how much quantity a monopolist will supply at various given prices, because the monopolist faces a whole given curve, not a given price. The same is true of a monopsonist, a single demander who faces a supply curve but does not have a demand curve.

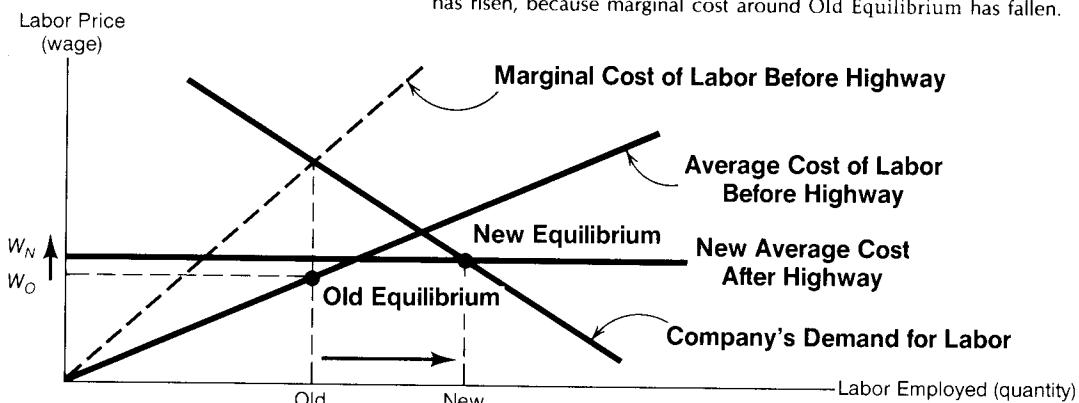
Q: A new interstate highway brings a formerly isolated Vermont town, Northfield, within easy commuting range of ten other towns, all of which, like this one, have one major employer. The major employer, in other words, is a monopsonist (a sole demander). Nothing happens to the price of the Northfield Woolen Mill's product at the factory gate, yet both the number employed and the average wage paid by the company rise.
True or false: Evidently the demand for labor by the

company is perversely shaped, upward sloping: An increase in the potential labor supplied to the company (coming by the new highway) has increased, not decreased, the wage it is willing to pay.

A: The company was before the highway a monopsony in the buying of labor, employing less the better to pay less. The highway breaks its monopsony power, flattening out the supply curve of labor that it faces

Figure 17.10
A Monopsony Does Not Have a Demand Curve

When the market for labor becomes competitive, the company's marginal cost of labor becomes the same as the wage rate, or New Average Cost. The company, originally at Old Equilibrium, increases employment, although the wage it must pay at Old Equilibrium has risen, because marginal cost around Old Equilibrium has fallen.



(shown in Figure 17.10). The demand curve (which is really the monopsonist's curve of marginal benefit from hiring labor) is unchanged and perfectly normal. Therefore, false. The demand curve is not perversely shaped. Before the highway the company, as a monopsonist, was not operating along its demand curve. So

it jumped from being off to being on the curve and in the process traced out two points, Old Equilibrium and New Equilibrium. A monopsonist does not have a demand curve, just as a monopolist does not have a supply curve.

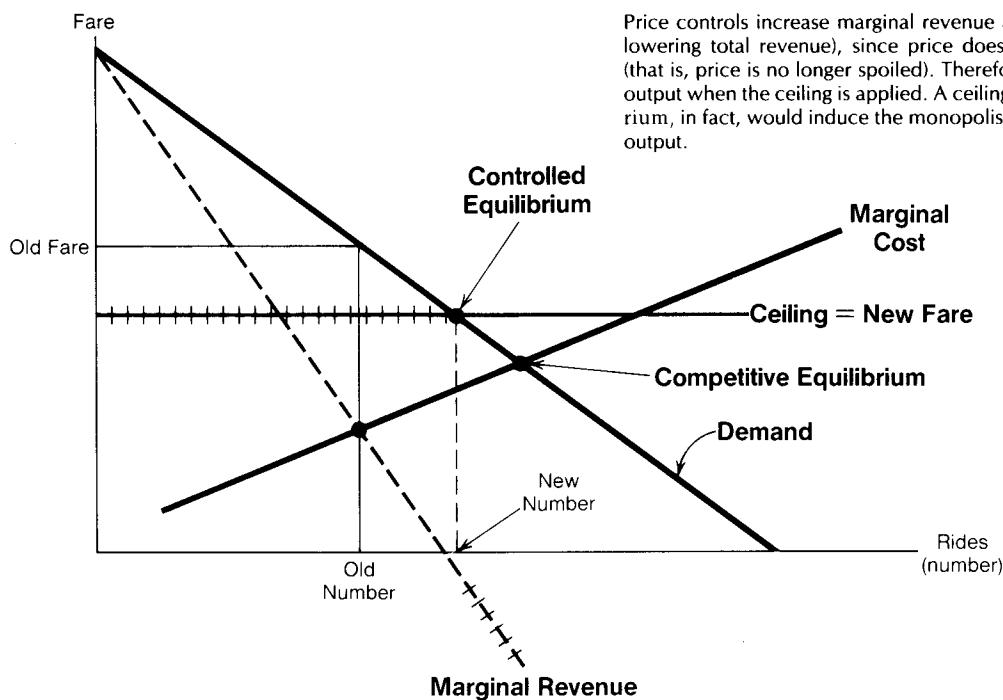
Monopolies with Price Limits Behave Like Competitors A similar case for a monopolist with the usual, upward-sloping marginal cost

T or F: Placing a ceiling on the price a monopolistic stage coach service between Abilene and Dodge City can charge will reduce the amount the monopoly will sell. On each fare the monopoly will make less money and will therefore run fewer coaches.

A: Before the price ceiling is imposed, the monopoly charges the Old Fare in Figure 17.11 and finds that riders demand the Old Number of rides (or alternatively and equivalently, it offers the Old Number of rides and finds that riders are willing to pay the Old Fare for each). After the Ceiling is imposed, however, its demand curve at a price higher than the Ceiling is no longer relevant. Up to the Ceiling along its demand curve, nothing is changed, but at lower outputs, the Ceiling is its demand curve. To put it the other way, out to the quantity of Controlled Equilibrium of rides, its demand curve is now flat and its marginal revenue, therefore, is not below the price it gets (namely, the Ceiling) but

equal to it. Beyond the point Controlled Equilibrium, after a leap downward, the marginal revenue is the portion of the old marginal revenue curve to the right of Controlled Equilibrium. In total, then, the monopoly's marginal revenue is the crossed line. There is nothing mysterious about the leap downward at Controlled Equilibrium. At that output an *increase* in output, for which consumers are only willing to pay something less than the former (ceiling) price, spoils the price, as before. But a *decrease* in output does not enhance the price (as it did before), because the price is limited by the Ceiling and cannot be enhanced. At such a point the marginal revenue for an increase in output is radically different from the marginal revenue for a decrease. To answer the question, then, the monopolist facing the price ceiling produces where marginal cost goes through the marginal revenue curve, in this case through the leap (or discontinuity). The monopolist produces the New Number of rides and sells them at the Ceiling Price.

Figure 17.11
Marginal Revenue for a Monopolist Facing Price Controls



Price controls increase marginal revenue at the Old Number (though lowering total revenue), since price does not fall as output expands (that is, price is no longer spoiled). Therefore the monopolist increases output when the ceiling is applied. A ceiling set at Competitive Equilibrium, in fact, would induce the monopolist to produce the competitive output.

Notice that the monopolist has been induced by the price ceiling to offer *more* rides. In general, then, the answer is false: A price ceiling on a monopolist can cause it to offer more, not less. If the price ceiling were

set at Competitive Equilibrium, in fact, the self-interest of the monopoly would drive the monopolist to offer exactly the competitive output, where price is equal to marginal cost.

Here is another question with the same point—namely, that monopolies with artificially flattened demand curves produce an output closer to the competitive output:

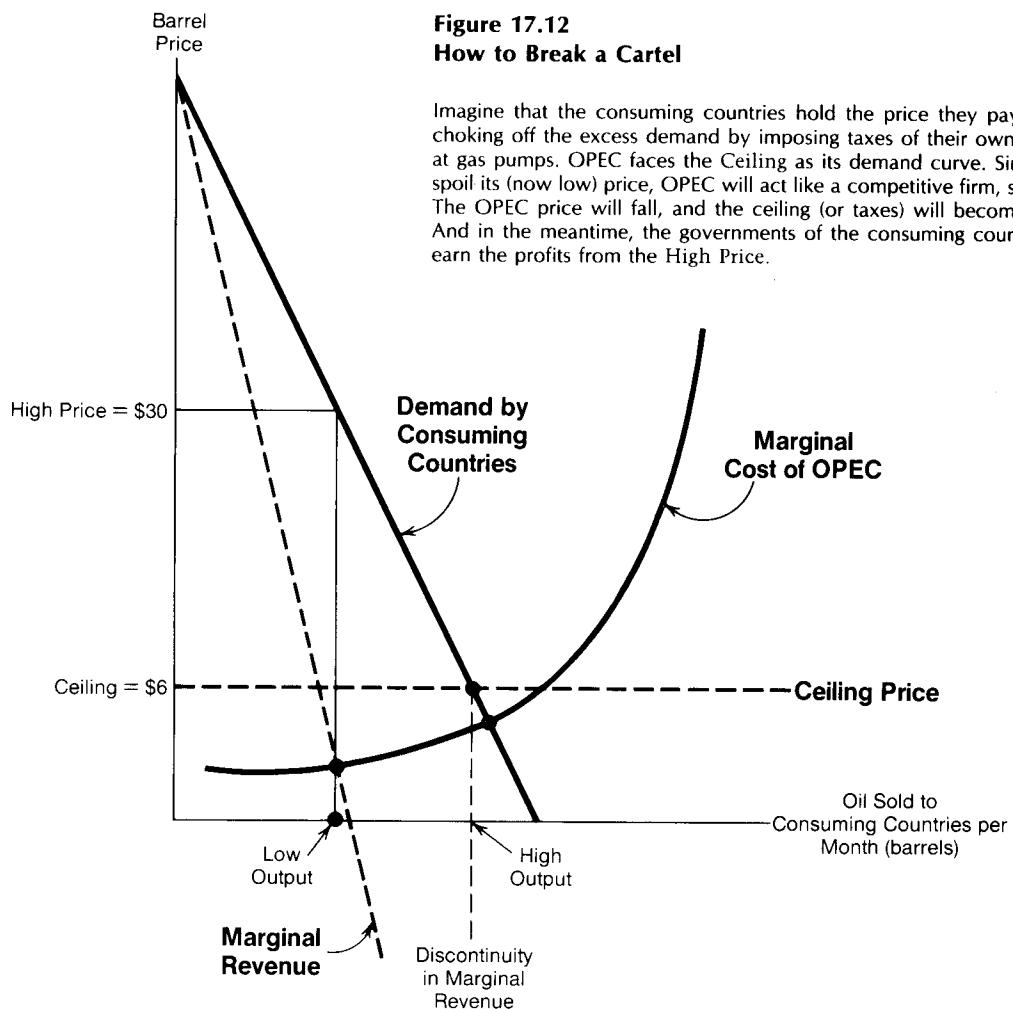
Q: The Organization of Petroleum Exporting Countries (OPEC) is a monopoly or more accurately a *cartel*, that is, a group of sellers acting together as a monopoly. Since 1974 it has been able to maintain a price of oil higher than the marginal cost of the least efficient producer, by restricting the supply. Suppose that the cartel price were \$30 per barrel (by now it is higher). Senator Henry Jackson proposed once that a ceiling of \$6 a barrel be placed on the price of oil in consuming countries (America in the first instance and also, he would hope, other countries). He suggested that the resulting excess demand from the cheapness of oil be eliminated

by rationing. Some economists, alternatively, have proposed that consuming countries impose import duties on OPEC oil, on a sliding scale: \$24 a barrel when the domestic American price is \$30, \$23 when it is \$29, and so on down to zero when it is \$6. Show that either of these schemes (assuming perfect enforcement) could induce OPEC to increase the output of oil and reduce the price to \$6.

A: Before the schemes are in place, OPEC, as a sensible monopolist, sells Low Output at the High Price in the diagram of \$30 a barrel. If the consuming countries

hold the price they pay to OPEC at the Ceiling, choking off the excess demand by imposing taxes of their own or by letting queues form at gas pumps, then OPEC faces the Ceiling as its demand curve. Since further sales no longer spoil its (now low) price, OPEC will act like a competitive firm, selling out to High Output (see Figure 17.12). The OPEC price will fall and the ceiling (or taxes) will become in the end unnecessary. And in the meantime the governments of the consuming countries instead of OPEC will earn the profits from the High Price. Either the ceiling price or the sliding scale of import duty has the same effect, namely, to flatten out the demand curve facing OPEC at a price

below the cartel price. No one is permitted to pay OPEC \$30 a barrel. Under Jackson's scheme no one pays \$30, except in resources devoted to queuing or stealing. Under the alternative scheme the government gets in tariff revenues the difference between OPEC's price and the domestic price that clears the market, a zero difference when OPEC increases output to High Output. Faced with a flat demand curve at \$6 a barrel, OPEC increases output to High Output, for that is now its point of maximum net revenue. OPEC has been induced to behave as a price taker, because it faces, and must take, the price of \$6 per barrel.



COMMENT

The analysis is incomplete in one important respect, to be treated in more detail in Chapter 22. OPEC may well adopt the countervailing strategy of persisting in withholding oil even though it reduces its present revenue by doing so, for it may believe that the consuming countries will not stick with their self-denying policy. In that case the outcome is a matter of pure bargaining, a matter, that is, of who is best able to convince the other party that he will stick with it.

The Rule of Rational Life Applies to Monopoly

The point is to keep firmly in mind what marginal revenue (or, more broadly, marginal benefit) a monopolist faces.

Q: A monopoly selling iron ore merges with a monopoly that makes steel from iron ore. The new, *vertically integrated* monopoly will price ore sold to its steel mill in a competitive fashion, that is, at marginal cost. *True or false:* The merger, therefore, even though monopoly has increased its reach, produces more output of steel.

A: The intuition is simple. It is obviously foolish to extract monopoly profits from oneself, which implies that the price of ore, formerly high (since the ore mine was

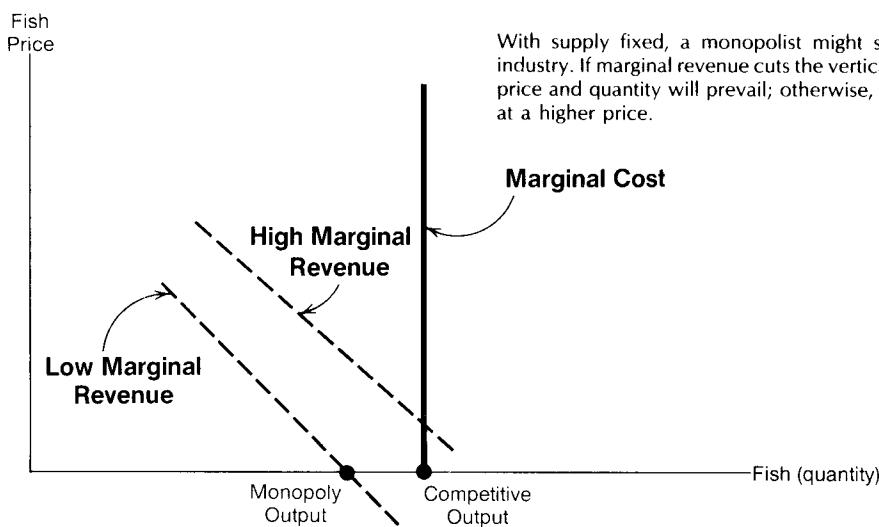
a monopoly) will fall. The old monopoly price of ore was the marginal benefit that the consumer (the steel mill) got from buying a ton of ore; this is what a demand curve means. The price was above the marginal cost of producing ore; this is what a monopoly does. Now that the mill and the mine are one, the integrated firm will equalize marginal cost and benefit by lowering the price to marginal cost (for itself, not, of course for outsiders). With lower costs of making steel, it will sell more. In short, true.

The marginal cost facing a monopoly deserves close attention as well.

Q: On any given day the supply of fish to the Billingsgate fish market is perfectly inelastic, because fish cannot be stored and within a day the daily supply of fish

cannot be altered. The fishmongers undertake to sell any quantity supplied. The market closes on Sundays, any fish left over from the Saturday night being left

Figure 17.13
A Monopoly May or May Not Let Fish Spoil Before Letting Them Spoil Their Price



With supply fixed, a monopolist might sell on the same terms as a competitive industry. If marginal revenue cuts the vertical portion of marginal cost, the competitive price and quantity will prevail; otherwise, the monopolist will sell a smaller quantity at a higher price.

to rot. On Saturday the fishmongers of Billingsgate form a cartel. *True or False:* On that Saturday the price of fish will always rise above the competitive level and the quantity sold will always fall (that is, some fish will be left over to rot).

- A:** Not necessarily. Out to Competitive Output in Figure 17.13 the marginal cost of fish is zero, because once the day's fish are at the market it costs no more to supply Competitive Output than to supply no fish. At Competitive Output the marginal cost is infinite (or very high), because once the day's fish are at the market, at no price (or only at a very high price) can more fish than Competitive Output be supplied.

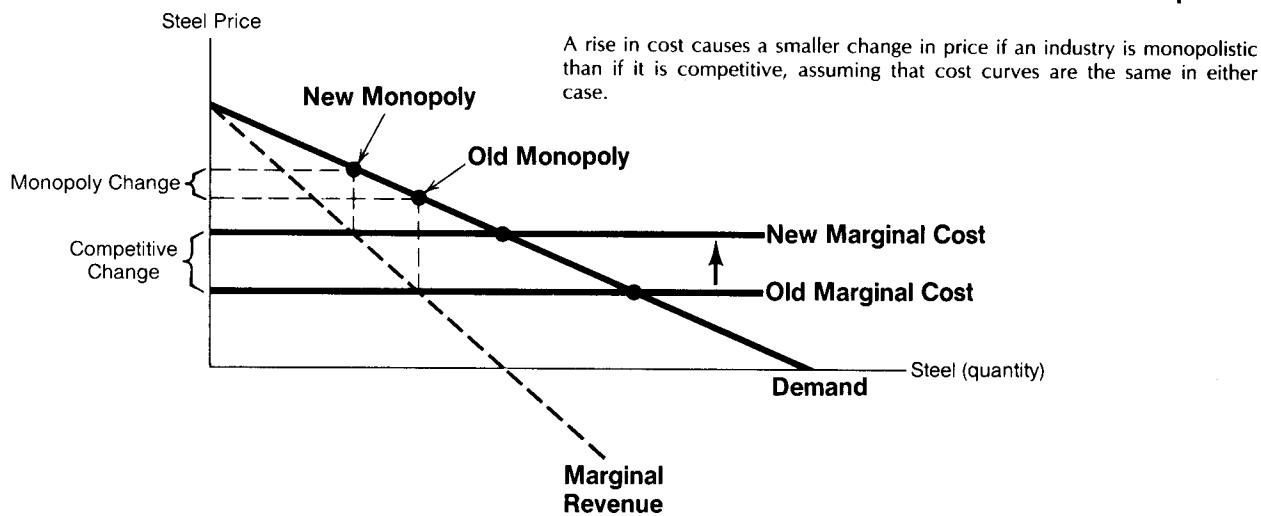
COMMENT

For one Saturday the analysis is complete. For many Saturdays it is not. The supply is not inelastic in the long run. And even if it is—even if fishermen continue to supply the Competitive Output each Saturday regardless of the price—the cartel can be expected to reduce the sale, for a cartel is not worth forming (supposing that there are costs of organizing and enforcing it) unless it raises the income of its members. In other words, if you observe a cartel, you can presume that marginal revenue cuts the marginal cost curve at a quantity *below* the Competitive Quantity. This is another way of saying that a monopoly operates only in the elastic portion of its demand curve.

As is evident by now, the simplest model of how monopolies behave is widely applicable. For example, it applies to a problem in the theory of inflation.

If the marginal revenue curve cuts the quantity axis at a point such as Monopoly Output, the price will rise and the quantity fall: The sale of fish that maximizes the revenue of the fishmongers is smaller than the amount of fish available. Fish will be left to rot. But if the marginal revenue curve cuts the vertical marginal cost curve, as does High Marginal Revenue, the price will not rise and the quantity will not fall. The sale of fish that maximizes revenue is in this case the competitive sale. A simpler way to make the argument is to imagine a diagram of *total* revenue and to note that only if supply is beyond the peak of revenue will it be desirable to cut supply back. So a cartel may wish to behave competitively.

Figure 17.14
Monopolies Can Pass on Lower Cost Rises Than Can Competitors



T or F: Monopolists contribute to cost-push inflation by always passing on a higher share of a rise in costs than would a competitive industry.

A: Many members of the Council of Economic Advisers have believed it. The simplest counterexample (which is all that is required, because the assertion uses “always”) is a flat, straight-line demand curve, for, say steel (see Figure 17.14). A competitive steel industry

would respond to a rise from Old Marginal Cost to New Marginal Cost by raising the price of steel by the same amount. A monopolized steel industry would respond by raising it from Old Monopoly to New Monopoly, which for straight-line demand and marginal cost curves (and marginal cost moving up in parallel) is always *less*, not more, than the competitive rise. Monopoly Change can be (as here) less than Competitive Change. Therefore, false.

**If My Grandmother
Had Wheels, She
Would Be a Tram:
Comparing
Monopoly and
Competition**

Comparisons between monopoly and competition, however, are to be made gingerly. One is liable to fall into nonsense of the sort “If my grandmother had wheels she would be a tram.” The steel industry either is or is not a monopoly. Suppose that it is a monopoly. To imagine that it was not and to compare the hypothetical results with the actual is to step into a *counterfactual* world. The monopoly presumably exists in the industry because of some cause, such as major advantages in marketing by one large company or a patent held by a few firms or a political atmosphere in which governments support cartels. One must assume away the causes to think about the effects of a régime of competition rather than monopoly. This is acceptable if the causes themselves have *only* the effect of supporting the monopoly. But they might not. The marketing advantages of the dominant firm, for example, might reduce its costs. If these advantages are eliminated, the industry’s costs may be higher. Therefore, by the nature of the question, one cannot hold the costs constant while imagining the industry moving from monopoly to competition.

There is nothing inherently objectionable about counterfactuals, despite their somewhat alarming name. The innocent assertion that the demand curve for housing, say, has an elasticity of -0.5 involves a counterfactual. If the price were not what it is, but 10% higher, the quantity demanded would be 5% lower. But the point is to avoid the sin of misusing the *ceteris paribus* clause, namely, holding constant what cannot, by the nature of the question asked, be held constant.

Another example of how the cost curves might be altered is the following.

Q: Suppose that the ironmaking industry, with a production function exhibiting constant returns to scale, is composed initially of 1000 identical firms. All inputs *except iron ore* are supplied to it perfectly elastically. The 1000 firms are now grouped into one cartel. What happens to the price of ore? What happens to the cost curve of the cartel by comparison with the cost curve of the competitive industry? In view of this, what do you think of the proposition “A cartel has lower output than does the corresponding competitive industry”?

A: The cartel faces the supply curves that the competitive industry as a whole faced. That is, the cartel faces an upward-sloping supply curve for ore. That is, the cartel

is *also a monopsonist*. So the cartel will cut back its purchases of ore so that the *marginal* (not average, as before the cartel) factor cost of ore equals the marginal valuation of ore. This reduces the price of ore (the average factor cost) that the industry pays (the other factor prices being constant by virtue of their elastic supply). Therefore, the cost of the industry falls. Consider an industry with constant factor shares for labor, capital, coal, ore, and so forth. The percentage change in factor prices is zero except for ore, and for ore it is negative. So the average of factor prices (weighted by their shares) falls and the cost curve falls. It is even possible for the cost curve to fall so much that the cartel produces more output.

In other words, "there are superior methods available to the monopolist which either are not available at all to a crowd of competitors or are not available to them so readily. . . . There are advantages . . . secured only on the monopoly level . . . because monopolization may increase the sphere of influence of the better, and decrease the sphere of influence of the inferior brains, or because the monopoly enjoys a disproportionately higher financial standing."⁵ This element of the case for competition may fail completely because the monopoly prices are not necessarily higher or monopoly output smaller.

Natural Monopoly and Unnatural Competition

The advantages of the monopoly might be technological, which is the case of *natural* monopoly. The one railroad between Lynchburg and Danville, for example, might leave no economic room for the 100 others necessary to make each a competitive price taker. The relevant comparison between monopoly and competition is obscure. And it's not clear what might be done to improve the situation. It would be silly to insist on replacing the monopoly by the 100 competitors, or even by 2. If 2 or many had to share a market best supplied by 1, the net social benefits might well be lower. The outcome depends on the counterfactual, that is, on what the alternative to the natural monopoly is imagined to be.

Q: Suppose that the alternative to a monopoly railroad between Lynchburg and Danville is transport by pack mule, which is supplied competitively both before and after the railroad, but at higher cost. *True or False:* If the advantage of cheap transport by rail is inseparable from the existence of a monopoly providing it, then the coming of the railroad, even though it acts like a monopolist, is good for the consumers of transportation.

A: The price of transport by pack mule is the worst that consumers can do, since the pack mules stand ready to go if the price gets above their entry price. The monopoly price, therefore, must be below the old competitive price (the pack mule price), and consumers are better off (see Figure 17.15). So, true; monopoly can be good for you.

Public Utilities as a Solution to Natural Monopoly

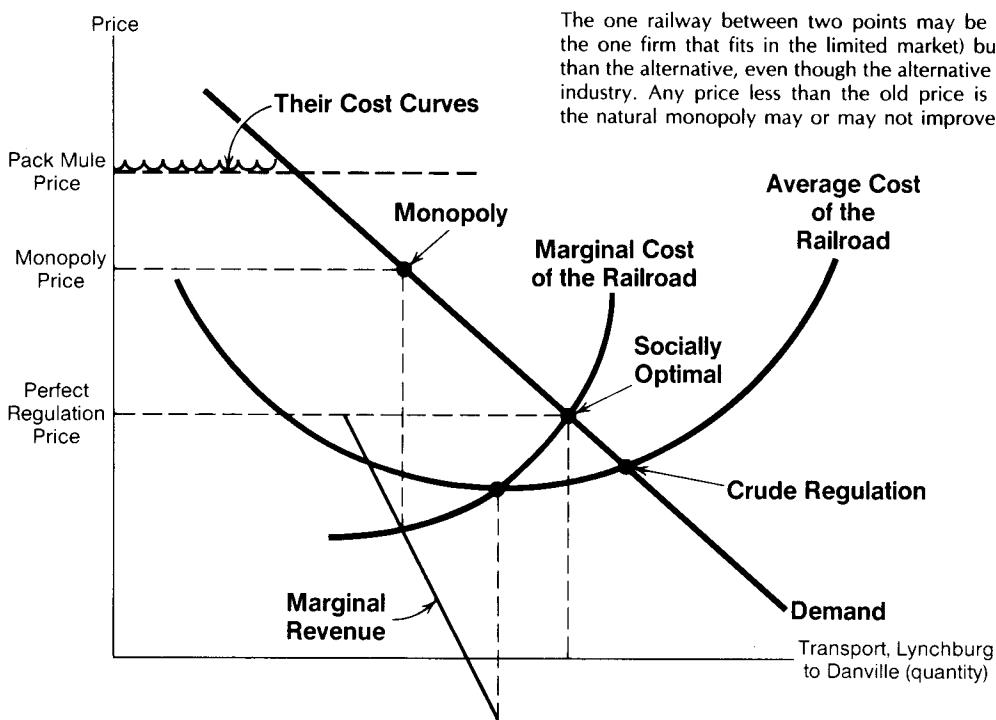
The monopolistic railroad between Lynchburg and Danville does not, of course, produce the Socially Optimal output, where marginal valuation equals marginal cost. Regulation of the railroad as a public utility or as a government enterprise could bring it there, if the regulators were willing and able to perform the calculations. Electric utilities staffed by economic geniuses, such as Électricité de France, are said to do exactly this, charging their customers the marginal cost and allowing the demands of the customers to pick off the point Socially Optimal. In the more common case the utility is compelled to price at average cost (including a normal return to capital). Consumers of the utility would be delighted with the low price at Crude Regulation, but the society as a whole is buying too much of a good thing. The social loss from Crude Regulation, in short, can be greater than that of monopoly.

If My Grandmother Had a Patent

In any event, there is often no obvious alternative to lend meaning to the assertion that "monopoly is worse than competition." Would my grandmother be better or worse off if she had wheels? The following is an example that ends with the same point, after applying the model of simple monopoly.

⁵ Joseph A. Schumpeter, *Capitalism, Socialism and Democracy*, 3rd ed., first published in 1942 (New York: Harper Torchbook, 1962), p. 101.

Figure 17.15
Natural Monopoly Can Be Good for You



Q: A patented invention reduces the cost of making hand calculators. The owner of the patent can charge a royalty for each calculator that uses the invention. *True or false:* The invention will not reduce the price of calculators to consumers, because the best royalty per calculator from the point of view of the owner of the patent is a royalty just below the amount of cost reduction. The manufacturers will be willing to pay the royalty and the owner will make as much as possible.

A: The patent bestows a monopoly on the owner, similar to the state's monopoly in taxing liquor. Suppose that the Preinvention Marginal Cost curve of calculators is flat, as is the Postinvention Marginal Cost (see Figure

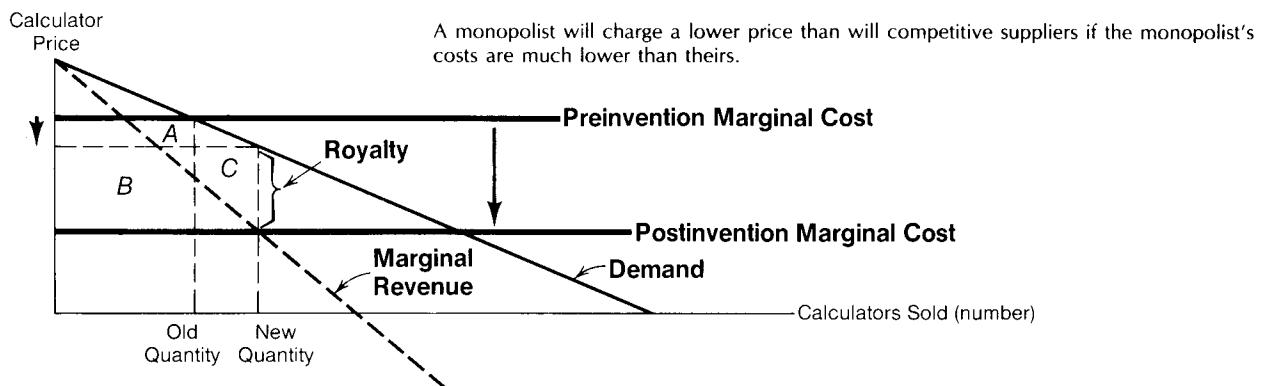
One can ask whether it is desirable to have such patents on invention at all.

T or F: Since price does not equal the marginal cost of production if the production process is patentable (that is, patents give monopoly power), patents should be eliminated.

17.16). Charging a royalty just below the amount of cost reduction will give the owner area $A + B$. But if at the Old Quantity the marginal revenue is greater than the new Postinvention Marginal Cost, then a better royalty is the lower one, Royalty, giving $B + C$. Such a royalty equalizes marginal cost and benefit, as though the owner of the patent bought the calculators manufactured with the new technology and resold them to consumers (compare the state's "buying" liquor at the pre-tax price and "selling" it at the posttax price to customers). And it results in a larger quantity of calculators sold at a lower price. So, false in general (though in particular cases it could be true): The best royalty does *not* necessarily take all the difference in cost.

A: One is free to imagine voiding the patent on hand calculators, leaving the industry pricing at Postinvention Marginal Cost. This would seem to be the obvious alternative to contrast to the monopoly patent, and is rele-

Figure 17.16
Royalties on Patented Inventions Are Not Always So High That the Price of the Product Stays the Same



vant when the patent expires (17 years under American law). But one is *not* free to imagine a world in which patents are regularly voided prematurely, or not granted in the first place, *yet in which the same amount of*

knowledge about calculators is produced as in the actual world. The knowledge is produced because it is patentable. Therefore, false.

The Moral Monopoly, then, should not be diagnosed and treated as though it were one disease. A doctor who treats fever alone, disregarding the causes, does her patients wrong. Likewise, economists and lawyers who attack with the same analysis and policy all markets having the same "structure" (that is, one seller or few) do their society wrong. An analogous case is slum clearance, the policy of all enlightened nations since Britain initiated it in the nineteenth century. Slums are bad relative to ideal communities, just as monopoly is bad relative to ideal marginal cost pricing. The instinct of the social engineer is to clear the slums, possibly replacing them with clean buildings, designed according to a uniform plan. The result has been housing projects worse than the original slums. The enthusiasts for clearance have overlooked the possibility that the original slums might have been better than the *actually* available alternatives or that the slums had *causes* that might reassert themselves in more virulent form when the old community was disrupted. Analogous remarks can be made of monopoly. Sometimes even monopoly is better than the actually available alternatives. And it always has causes, differing from case to case, that must be brought into the cure.

Summary The model of monopoly, then, is widely applicable. It applies to any situation in which one mind is facing an entire demand or supply curve instead of a given price. The question is always, what is the monopolist's marginal revenue? Only changes in the monopolist's marginal revenue affect the optimal position, as when imposing a lid on OPEC's price affects its marginal revenue and induces it to supply more oil. The assumption necessary for such exercises, however, is sometimes questionable: It is that the monopolist can be imagined to be transported to a world in which it is *not* a monopolist. If my grandmother had

wheels, she might or might not be a tram; but she certainly would no longer be my grandmother. Too quick comparisons between monopoly and competition are to be avoided, as are the policy prescriptions that go along with them. Monopolies are bad, surely—Chapter 19 will count their badness. But it is an error to try to eliminate their badness without understanding their causes, a point that applies to any issue of social policy.

EXERCISES FOR SECTION 17.2

1. *True or false:* Since monopolists are just looking for maximum profits, the economist's model that has them looking for the output at which marginal cost equals marginal revenue is irrelevant to real behavior. (*Hint:* What is the connection between looking for maximum profits on the one hand and the marginal condition on the other?)
2. One shouldn't get overly enthusiastic about the contrasts between the economist's model of monopoly and the model of supply and demand:
 - a. If the whole demand curve for a monopolist rises at every price, what happens to the monopolist's price and quantity? How does this compare with what happens in a competitive market when the whole demand curve rises?
 - b. What happens to price and quantity if the marginal costs of the monopolist fall uniformly? How does this compare with competition?
3. Does a price ceiling, if effective, *always* cause the price of stagecoach rides (see Figure 17.11) to rise?
4. Discuss the element of if-my-grandmother-had-wheels in the following:
 - a. If the authors and publishers did not hold copyrights to books the price of books would be lower.
 - b. If the steel industry were more concentrated the customers would be worse off.
 - c. If the railways had not been invented the income of the nation would have been about the same.

PROBLEMS FOR SECTION 17.2

- 1. Imagine Britain in 1900 facing a downward-sloping demand curve for its loans to foreigners (the interest rate earned and paid is the price, pounds sterling worth of loans is quantity). Imagine, too, a marginal cost curve of the British in supplying such loans.
 - a. What is the *competitive* equilibrium?
 - b. Could Britain do better for itself if the British government restricted the amount supplied? Show the gain to the optimal tariff (tax) on exports of capital (loans to foreigners).

True or False

2. Because the company is a monopoly, a tax of t cents per kilowatt-hour of electricity sold by Consolidated Edison will fall on the company, not on its consumers.
3. If the city council sold the right to exercise a natural monopoly of a new cable TV franchise in Philadelphia to the highest bidder, the problem of monopoly would be eliminated.
- 4. If the city council sold the right to exercise a natural monopoly of the cable franchise to the company that promised to charge the lowest fares, the problem of monopoly would be eliminated.
5. If an electric utility has average costs that in the range of present demand always fall with increasing output, then marginal cost is below average cost at any output, and a policy of marginal cost pricing (to achieve the socially optimal output) will lose money.

CHAPTER

18

Measuring Monopoly

What to Read For

Why does it matter how common monopoly is? Is profit a good measure of its commonness? Does the purchaser of an existing monopoly make monopoly profits on his purchase? What is *dissipation of rents*, and how does medical education show it? Is advertising an indication of monopoly? Is the number of firms in an industry a good measure of monopoly power? On what does the elasticity of demand facing one firm depend? What is the *Fundamental Equation of Monopoly*? How can one measure the marginal cost in the *Fundamental Equation*? At what price does a monopolist at home sell in a foreign, competitive market? What is the Rule of Rational Life for a monopolist in two markets? What is the *Fundamental Equation of Discriminatory Monopoly*? Are college scholarships examples of price discrimination? Can a monopolist price discriminate if the customers can resell the product? What is *multipart pricing*? Does charging different prices to different customers always indicate monopoly?

Is Monopoly Common? The Debate and Its Importance

Monopolies, then, are different from competitive industries. An economist looking at an industry will wish to know with which case she is dealing. If she is looking at an entire economy she will want to know which case is typical of the whole. How will she know?

The economist will have a difficult time knowing from the testimony of other economists. The theory of monopoly was developed in the late nineteenth century to deal with allegedly rare cases of monopolies on salt, natural gas, and mineral water. The economy was supposed to be on the whole competitive. Most economists still accept the competitive model as the best rough characterization of how the economy works.¹ Since the 1930s, however, many other econo-

¹ When 198 economists in the mid-1970s were asked to assess the statement that "The 'Corporate State,' as depicted by Galbraith, accurately describes the context and structure of the U.S. economy," half flatly disagreed and only a fifth agreed without provisions. Furthermore, 80% of the 25 economists

mists have come to believe that the economy is on the whole monopolistic, pointing to the fewness of companies in autos, soap, steel, cereal, soft drinks, brewing, and elsewhere and to the restrictions on entry to unionized occupations, heavily advertised industries, and industries regulated by the government. Competition in this view is an exception today and was an exception historically, a brief interlude between an age of local monopoly in the eighteenth century and of national monopoly in the twentieth. A realistic model of the economy, these economists would say, must speak of monopoly capital and the balance of countervailing powers.

The dispute matters. True, most propositions in economics survive transplantation to a world of monopolies. An outward move in demand would still bring price increases, an outward move in supply would still bring price decreases. The Law of Demand would still hold, although no longer confronted by a simple law of supply. There would still be more than one way to skin a cat, although the right way would not in general be chosen. Comparative advantage would still be society's best guide to specialization, although market prices would no longer be themselves a guide to comparative advantage. And so forth.

But if monopoly were pervasive, there would be less reason to *respect* the pattern of output the economy produced. On the contrary, the pattern would be demonstrably wrong, and a case could be made to change it, perhaps even by using the coercive power of the state. It is said that if people buy General Motors cars because they have no alternative or if they work at a certain job because trade union rules have mandated it or if they invest in chemical and insurance companies because the owners are politically powerful, then interventions to restore competition in the car market or the job market or the investment market are not only acceptable but desirable. One's convictions about the prevalence of monopoly, in short, can affect one's politics.² And in any case they can affect one's economics.

**Profit Is a Measure
of Monopoly, Unless
It Is Rent**

The most obvious way of resolving the dispute is to look around the economy at the level of profit. Presumably monopolies make profits above normal. The industries with above-normal profit could be placed in the "monopoly" column, those with normal profits in the "competitive" column.

You can see that the procedure would run into difficulties if all industries were in fact monopolies, because it would be unclear how any one monopoly would stand out as unusually profitable. Even supposing the whole world is not monopolized, the level of profit is not a reliable guide, on many grounds.

For one thing, the ordinary workings of a competitive economy make some nonmonopolists rich. The owner of a handsome face, a mellifluous voice, a

employed as full professors at seven leading economics departments (as distinct from those at other universities, or in government or business) flatly disagreed, more so even than economists employed by the corporations themselves. J. R. Earl, Clayne L. Pope, Gordon C. Whiting, and Larry T. Wimmer, "A Confusion of Economists?" *American Economic Review* 69 (May 1979): 28–37, especially pp. 30, 36.

² Two important works attacking the notion that competition is the way the economy works are Paul A. Baran and Paul M. Sweezy, *Monopoly Capital* (New York: Monthly Review Press, 1966); and John Kenneth Galbraith, *The New Industrial State* (Boston: Houghton Mifflin, 1967). Related writers in sociology and political science have attacked the analogous notion that democratic pluralism is the way politics works. For instance, C. Wright Mills, *The Power Elite* (New York: Oxford University Press, 1956).

charming air, and luck becomes a movie star (or a president) without any intent or ability to engage in price searching. Fertile land, likewise, earns economic rent, even if its owners compete with each other in atomistic markets. Furthermore, the temporary quasi-rent to the entrepreneur who is first on the scene with polyurethane roller skates or a vaccine for tooth decay is the very way in which a competitive economy attracts entry to desirable activities. Yet a snapshot of the economy might easily mistake the *quasi-rent* for *permanent monopoly rent*. To be a monopoly rent the excess payment above what is required to draw the resource from some alternative employment must be useless, serving to reward neither agility of response nor rare merit. Only some rents are monopoly rents. So the level of rents (that is, profits) is not much of a guide to the level of monopoly.

For the Secondhand Owner of a Monopoly, the Profit Is a Cost

And the monopoly elements in the rents are not easy to spot. Aside from the occasional governmental monopoly on salt or postal service, few monopolies are so arrogant as to have a line in their annual statements called "monopoly profit." Even without any desire to hide them, the profits from price searching are liable to be buried in accounting categories such as goodwill or salaries. They will look like normal costs instead of abnormal profits. For the person who buys the right to a monopoly from an earlier holder, indeed, the monopoly profits *are* normal costs. In the market for the monopoly right he competes with many others. Therefore, the *secondhand* buyer of a patent on cellophane or a license to operate a cab in San Francisco earns only a normal return.

Q: Woody Fleisig buys the Chicago Bulls basketball team for \$10 million and acquires thereby the rights to whatever he can earn from the services of a group of athletes after paying for their salaries, their equipment, and the rent on the stadium. Woody, in common with many other sportsman millionaires, proceeds to lose money, say, \$5 million. *True or false:* This is evidence that professional basketball, far from being a closed monopoly (as some gloomy-minded economists sometimes argue), is competitive. After all, monopolies are profitable.

A: If professional basketball were competitive, in the sense of having free entry, the permission to enter (that is, the franchise, the right for which Woody paid \$10 million) would be valueless. The ownership of the re-

sidual income from the Chicago Bulls after paying salaries and so forth would be valueless, the taking of risk aside. The ownership has no real social opportunity cost. The fact that Woody was willing to pay someone \$10 million indicates that he expected that the value of the stream of returns to ownership was \$10 million (after making the appropriate allowances for the value of any reductions in his income taxes that he might have anticipated from losing money on the property). That he did in fact lose money merely indicates either that he has a devilishly clever tax lawyer or that he had overly optimistic expectations about the future *rise* in the monopoly profit. Therefore, false. The person who earns the whole monopoly profit is the person who established the monopoly, not the one who later buys the rights to its profits, once established.

The profits from monopoly can be buried so deeply in cost that literally no one earns above-normal returns. The principle here is the zero-profit condition and the jargon is *dissipation of rents*. Since monopoly rents are valuable, a monopolist is willing to spend valuable resources acquiring and defending them. The lawyers, accountants, chemists, economists, teachers, inspectors, and so forth hired to seek rents may well earn nothing above their value in alternative employment. Their employment in rent seeking is like standing in line, that is, a sheer waste of resources, benefiting no one. The monopoly rents can be dissipated entirely in wasted motion.

Medical education is a good example. The doctor in the emergency room who puts your finger, sprained playing football, into a metal splint has been to college for four years, to medical school for another four, and to internship for one or two years. The expense of the four medical school years is very high, far above that for any other form of education. Even though the fees he charges are large, the individual doctor may find himself earning only a normal return on this enormous investment and will therefore quite understandably resent being called a monopolist. That the athletic trainer with an eighth-grade education does a better job of fixing the sprain at a quarter of the cost, however, suggests doctoring as a whole is indeed a monopoly, although one that throws away some of its monopoly rents on superfluous education.

The Height of a Barrier to Entry Is a Measure of Monopoly, Unless It Is Advertising

The barrier to entry that the educational requirement of medicine erects is itself evidence of monopoly. So are other barriers to entry. You might infer from locks, burglar alarms, and guard dogs that a house contained valuable jewels. Likewise you might infer from licenses, membership requirements, and patents that an industry contained valuable profits. The barrier, however, may be a real cost of doing business, not a mere protection to monopoly.

A case in point that has long been controversial in economics is advertising. Contrary to a widespread opinion, most advertising is merely informative, not argumentative. To be sure, the ring-around-the-collar ad on television is uniquely irritating and devoid of intellectual content, but in fact it is dwarfed in message time and dollars by the classified ads in the St. Louis *Post-Dispatch*, the billboard that tells you of a place to sleep on your trip to Montreal, or the radio jingle that informs you as a new denture wearer of the miracle of Dentu-Cream. Advertising executives will tell you that the industry's proverb is "you can bring a horse to water but you can't make him drink." They can back up their modesty about the power of advertising with stories of failed ad campaigns for failed products, such as Fruit Float or the Edsel.

Even admitting the informative nature of most advertising, however, the mere existence of advertising indicates something less than perfect price taking. In a trivial sense, even the sign on the grocery store is indicative of price searching by the store, because if it were literally a price taker it could sell all it wished at the going price and would have no incentive to spend money on increasing the amount it could sell. In a less trivial sense, large advertising budgets for soap and cigarettes indicate that the companies face downward-sloping demand curves and would like to push their marginal revenue curves out further, a meaningless activity for a price taker, who has a flat marginal revenue curve. That the companies spend so much on advertising indicates that the curves are indeed pushed out (or at least prevented from being pushed in by the advertising of competitors). The large expense of advertising becomes a requirement of the trade, a barrier to entry. The significance of this line of reasoning, however, is much disputed. Some economists believe that the barrier is high and leads to large monopoly distortions; others do not. The evidence so far is ambiguous.³

³ A view from the antiadvertising side is William S. Comanor and Thomas A. Wilson, "Advertising and Competition: A Survey," *Journal of Economic Literature* 17 (June 1979): 453-476; from the pro-advertising side, Harold Demsetz, "Accounting for Advertising as a Barrier to Entry," *Journal of Business* 52 (July 1979): 345-360.

The Number of Firms Is a Measure of Monopoly, Unless There Is a Competitive Fringe

If the size of monopoly profits and the height of barriers to entry are difficult to measure, the number of competitors in an industry is easy. Count them. One or few implies monopoly; many implies competition. But as was shown in the last chapter, *the number of firms now in an industry is always less than the number that would come in at a higher price, including those waiting on the fringe*. The share of the market taken by the three largest auto firms, say, may be very high but will result in little deviation from the competitive price if at a little above the going price five foreign firms will enter. In other words, the *concentration ratio* (the market share of the top N firms) looks like a straightforward measure of monopoly power, but in fact it is made doubtful by the neglected fringe of potential entrants. Low concentration (many sellers) might show that monopoly does not exist—although what would one say about monopolies of doctors and plumbers?—but high concentration cannot show that monopolies do exist.

Elasticity of Demand Is a Measure of Monopoly Power

The very failures of concentration as a way of measuring monopoly, however, lead to another, better measure, as follows. One among few sellers of steel will know of both existing *and* potential entry and will allow for it. The entry will in fact increase effective elasticity of demand. The elasticity of demand facing Inland Steel is a measure of Inland's monopoly power. The mathematics here is that of Chapter 7 on elasticities, which you should bring back to mind. There it was shown that the elasticity of demand facing *one* of the suppliers of steel, say Inland Steel, rises as the elasticities of market demand and of supply from U.S. Steel, Bethlehem, and so forth become high and as Inland's share becomes small (the elasticity is defined here to be positive):

$$\epsilon_{IN}^D = \frac{Q}{Q_{IN}} (\epsilon^D) + \frac{Q_{OTHERS}}{Q_{IN}} (\epsilon_{OTHERS}^S)$$

Elasticity of inverse of market ratio of elasticity
demand facing = Inland's market \times elasticity + others' output to \times of supply
Inland Steel share of demand Inland's output of others

There are four variables on the right-hand side. The presence of the fringe can be viewed as raising Q/Q_{IN} , Q_{OTHERS}/Q_{IN} , or ϵ_{OTHERS}^S , all of which raise the elasticity that Inland faces. The economic common sense here is that Inland's elasticity of demand is high—and its monopoly power low—if Inland is a small producer (high Q/Q_{IN} and Q_{OTHERS}/Q_{IN}) facing a fringe that supplies a lot more steel if prices rise to monopoly levels (ϵ_{OTHERS}^S high). The elasticity is a measure of Inland's monopoly power. If it is low, Inland faces a sharply downward-sloping demand curve, approaching the condition of a single seller.⁴ If it is high, Inland faces a flat demand curve, approaching the condition of one perfect competitor among many.

Notice that in the present case *it does not take a very large fringe, or a very small share, to make the elasticity very high*. If the elasticity of demand

⁴ The argument neglects the awareness that Inland will have of its effect on other steel makers and the awareness others will have of their effect on Inland. Chapter 21 talks about this.

overall were 1.5 and the elasticity of supply of others were 1.0, an inland share of 1 in 15 of every ton of steel produced would give an elasticity of

$$\epsilon_{ln}^D = \frac{15}{1} (1.5) + \frac{14}{6} (1.0) = 36.5$$

The Fundamental Equation of Monopoly

But is 36.5 high or low? The way in which elasticity of demand affects a monopolist provides an answer, and further justifies the elasticity as a measure of monopoly power. Most important, it makes it possible to measure the power without elaborate information on all the various elasticities of demand and supply. The marginal revenue of a monopolist, recall, is $MR = P - Q \Delta P / \Delta Q$. In equilibrium the monopolist arranges matters so that marginal cost equals this marginal revenue:

$$MC = P - \frac{\Delta P}{\Delta Q/Q}$$

It is one short step—dividing both sides by P —to a most useful idea:

The Fundamental Equation of Monopoly

$$\frac{MC}{P} = \frac{P}{P} - \frac{\Delta P/P}{\Delta Q/Q} = 1 - \frac{1}{\epsilon}$$

Ratio of marginal cost to the price charged by a monopolist	= inverse of the = 1.0 – elasticity of demand facing the monopolist
---	---

This equation provides a measure of the gap between price and marginal cost. (Recall that the monopolist produces an output level at which price is greater than marginal cost.) If the elasticity is very, very high, say, the 1000 it might be for a seller in a large auction market competing against many hundreds of others, then the ratio of marginal cost to price is very near to 1.0—namely, $1.0 - 1/1000 = 0.999$. Marginal cost is virtually equal to price. If the elasticity is low, say the 1.5 it might be for the government acting on behalf of the steel industry to set quotas on steel that maximized the American steel industry's profits, then the ratio is well below 1.0—namely, $1.0 - 1/1.5 = 0.33$. Marginal cost is well below price.⁵ Ask the question again. Is 36.5 high or low? Well, it implies a ratio of marginal cost to price of $1.0 - 1/36.5 = 0.97$. A monopolist who knew the marginal cost to within plus or minus 3% would be an unusually discerning person. An economist who could second-guess her knowledge to within plus or minus double that figure would be a genius. In practical terms, then, 0.97 is indistinguishable from competition. An elasticity of demand of 36.5 is low, and says that the monopolist has very little monopoly power.

The Ratio of Marginal Cost to Price in Monopoly Is Not Impossible to Measure

In any event, the ratio of marginal cost to price provides an attractive measure of monopoly power.⁶ Its attractions are two. First, because it is related to the elasticity, the ratio is a measure of the influence that the monopolist believes he has. In other words, it measures behavior. Second, because it is the ratio of the ideal price (marginal cost) to the actual, the ratio measures the hurt the

⁵ Notice that the formula reflects the condition that elasticity must be greater than 1.0 for a monopolist. If ϵ were less, say, 0.5, then the implied ratio would be negative, $1.0 - 1/0.5 = -1.0$, which is ridiculous.

⁶ Abba P. Lerner, "The Concept of Monopoly and the Measurement of Monopoly Power," *Review of Economic Studies* 1 (June 1934): 157–175.

monopolist's behavior imposes. In other words, it measures happiness. Detecting monopoly by measuring its profits or barriers to entry or numbers is inconclusive. Here is a conclusive measure: Measure the ratio of marginal cost to price.

The fly in the ointment is that marginal cost is very difficult to measure. Marginal cost is necessary for the calculation. The total cost may rise in response to a rise in output in ways that the firm itself will find difficult to know. And the firm will have in any case no incentive to make whatever it does know public.

The solution is to find a second market in which the monopolist firm is a competitor (that is, facing an infinitely elastic demand). In that market the firm sets its price equal to marginal cost, *revealing its marginal cost*. If a monopoly is temporary, such as a temporary coalition of gasoline stations in the neighborhood raising prices, one may judge its marginal cost from the episodes of competitive pricing.⁷

An even better circumstance for self-revelation of marginal cost is that of the monopolist selling in two markets, in one of which he is a competitor.

Q: The German steel cartel (a monopoly) in 1890 supplied all the market in Germany (a tariff protected the German market from interlopers) and also exported to the world market. How would you measure the degree of its monopoly at home?

A: In the *world* market it took price as given, because German steel was only a small part of the world's total.

Consequently, it sold to the world until its marginal cost equaled the world price. The marginal cost was common to production for either market, world or home—marginal cost is marginal cost no matter where the steel is sold. The world price divided by the German home price was therefore the degree of monopoly at home.

**The Uses of the Case
of Elastic Foreign
Demand**

Q: Suppose that no foreigners could sell steel in Germany no matter how high the price became but that the German monopoly could sell steel abroad at the world price if it wished. *True or false:* If the German monopoly did sell abroad, then the price it set at home would have been rigid with respect to changes in the costs of making steel.

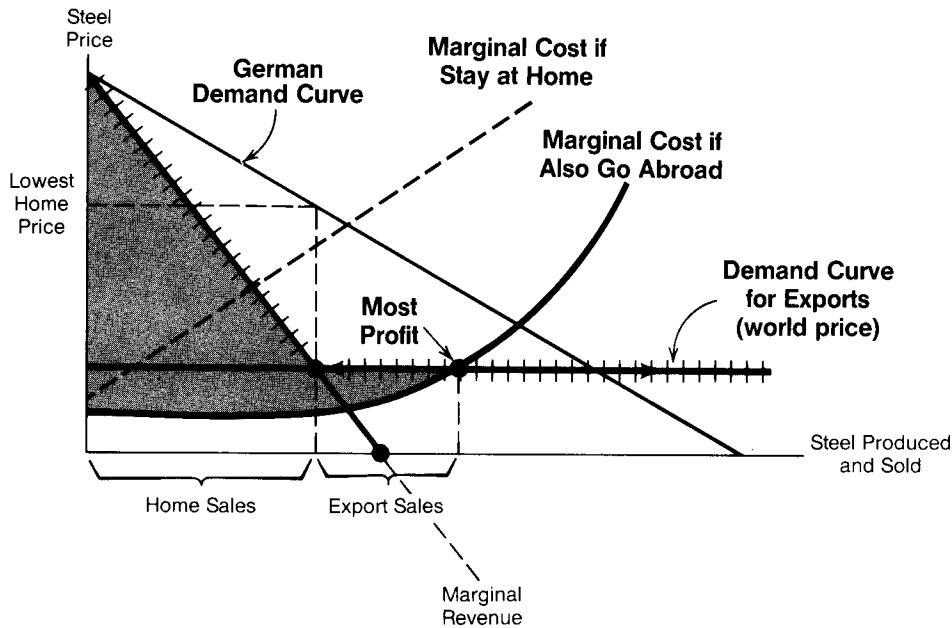
A: Look at Figure 18.1. That no foreigners may ever sell in Germany means that the German monopoly faces

the whole German demand curve. That it may sell abroad means that the monopoly also faces a flat Demand Curve for Exports. If it does sell abroad, it is evidently on a marginal cost curve like Marginal Cost if Also Go Abroad instead of Marginal Cost if Stay at Home. The two marginal cost curves are to be viewed as alternatives, not the marginal cost of producing for the two different markets (on the contrary, the cost is supposed to be the same, from a common plant). Look at the kinked Marginal Revenue. It is kinked because the mo-

⁷ This way of implementing Lerner's measure of monopoly power first occurred to John T. Dunlop, "Price Flexibility and the 'Degree of Monopoly,'" *Quarterly Journal of Economics* 53 (August 1939): 522–534. It has been used to examine the competitiveness of the iron and steel industry in the nineteenth century, which had temporary monopolies, by Peter Temin, *Iron and Steel in Nineteenth-Century America* (Cambridge, Mass.: MIT Press, 1964), p. 187; and D. N. McCloskey, *Economic Maturity and Entrepreneurial Decline: British Iron and Steel 1870–1913* (Cambridge, Mass.: Harvard University Press, 1973), pp. 24–28.

Figure 18.1
Self-interest Puts a Floor on the Marginal Revenue of a Monopolist Able to Sell in a Competitive Market

A monopolist at home facing a foreign market in which he is a competitor may sell in both markets, at home and abroad. If the monopolist does, he will set the marginal cost of output equal to marginal revenue in both markets and therefore will set them equal to each other. Because the one marginal revenue is just the price the monopolist faces, the foreign price drives the model. If it is too low (relative to marginal costs, such as the high Marginal Costs if Stay at Home), no exports occur and the case is one of simple monopoly.



nopolist will sell all his additional output in the foreign, constant-marginal-revenue market beyond the point where domestic marginal revenue has been driven down. The Marginal Cost if Also Go Abroad marginal cost intersects the Marginal Revenue at Most Profit, the shaded area of profit is clearly being maximized there. At such a point, marginal cost is equal to marginal revenue in both directions of sale, home and abroad. The rule of rational life permits no other conclusion. The marginal cost of a thing must be equal to the marginal

benefit in each of its uses if profit is to be maximized. The price at home is fixed at Lowest Home Price because at that price the marginal revenue abroad (namely, the price abroad) equals the marginal revenue at home. Changes in the position of the Also Go Abroad marginal cost curve will over a wide range have no effect at all on the best price to charge at home. In other words, as asserted, the home price will tend to be rigid with respect to cost changes at home.

A more general case introduces a useful fine point about monopoly diagrams and says something about the limits on monopoly power.

Q: Suppose that in 1880 in Britain steel rails were made in Glasgow and Sheffield, 170 miles apart. Suppose that

the Sheffield makers viewed the Glasgow price in Glasgow as given: They believed themselves to have no mo-

nopoly power in Glasgow. So did the Glasgow makers in Sheffield. Suppose finally that steel rails sold for 100 shillings a ton in Glasgow and could be transported between the two cities for 5 shillings a ton.

1. Within what limits must the price in Sheffield lie?

2. Suppose that the Sheffield makers form themselves into a monopoly at home. What is their curve of marginal revenue? What prices do they charge at home, given various curves of marginal cost?

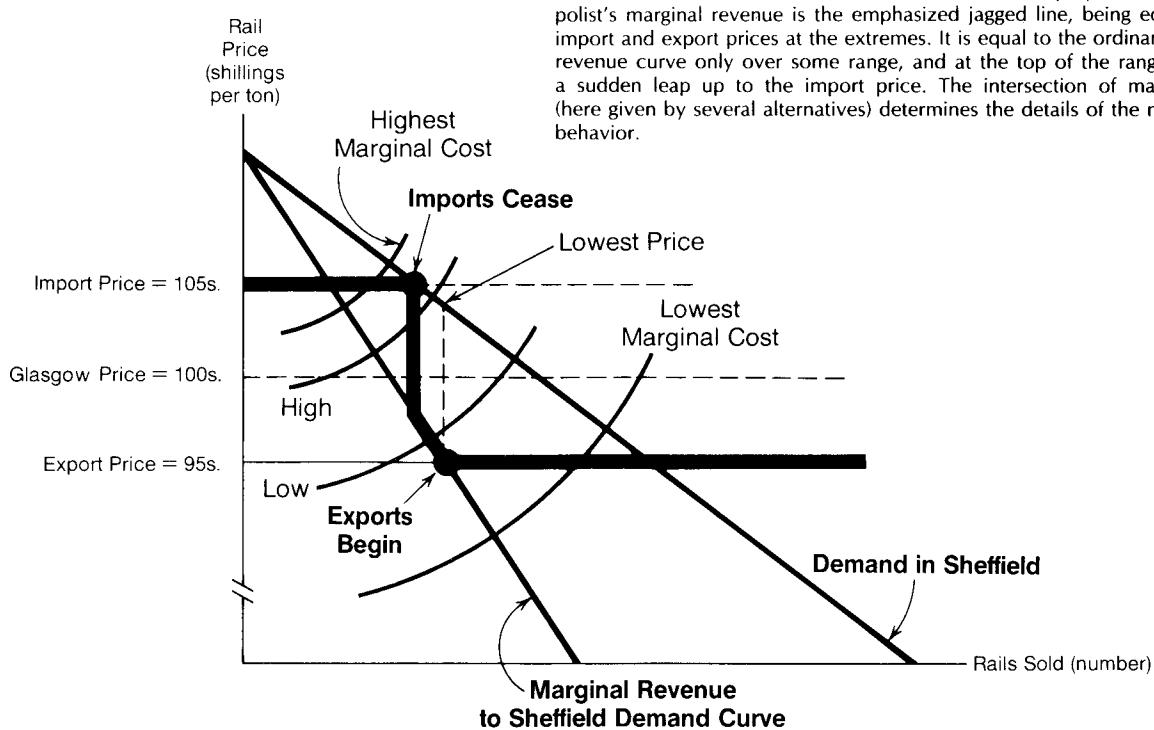
3. How many shillings is the *lowest marginal cost* that Sheffield will ever produce at? How many shillings is the *highest price*? What therefore is the lowest possible ratio of marginal cost to price? In view of the elasticity formula for the ratio, what is the lowest possible elasticity of demand the Sheffield monopoly faces?

4. If transport costs to or from a market in which a monopoly is not a monopoly is $t\%$ of the foreign price, what is the most monopoly power that it can have?

A: 1. Clearly the Sheffield price must lie between 95 and 105 shillings, for at a price in Sheffield a little below 95 shillings the Sheffield makers do better to sell their rails in Glasgow and at a price a little above 105 shillings the Glasgow makers invade the Sheffield market (see Figure 18.2).

2. The marginal revenue curve needs to be constructed step by step by imagining various different curves of marginal cost. Start with Highest Marginal Cost. In such a case the relevant marginal revenue is the 105 shillings set by import from Glasgow. Now consider High marginal cost. The marginal revenue leaps

Figure 18.2
Imports and Exports Limit Monopoly Power



down to the curve marginal to the Sheffield demand curve, as it did in the earlier problems on the stage coach and OPEC. To supply more than is demanded at Imports Cease is to force more onto the market than will be bought at 105 shillings. The price must fall and must spoil the price on all previous amounts. Naturally, then, the marginal revenue beyond Imports Cease is much lower than it is before. Over a wide range of High marginal cost curves, the price is rigid at exactly 105 shillings, and the quantity supplied by Sheffield makers is the amount at Imports Cease. At Low marginal cost the curve intersects the Marginal Revenue curve and it becomes worthwhile to supply more than the amount at Imports Cease. The price must therefore be a little lower. Finally, at Lowest Marginal Cost Sheffield is such a cheap producer that it exports. But it never

produces more for the home market in Sheffield than it does at Exports Begin. To produce more would drive marginal revenue at home below marginal revenue abroad (the export price), an irrational thing to do. The Sheffield price never gets below Lowest Price.

3. The lowest possible *marginal cost* is 95 shillings, by the argument just given. The highest possible *price* is 105 shillings, so the lowest ratio is $95/105 = 0.90$. Using the Fundamental Equation of Monopoly a few pages back, it must be true that (at most) $0.90 = 1 - 1/\epsilon$. Solving for the elasticity of demand, ϵ , yields an elasticity of 10.

4. The most monopoly power is clearly twice the percentage transport cost. For many items—gold, wheat, automobiles—such limits on monopoly power will be tight.

The Rule of Rational Life for Two Markets: The Fundamental Equation of Discriminatory Monopoly

Knowledge of the elasticity a monopoly faces in one market, then, is enough to know its marginal cost. This makes it possible to measure against market price its monopoly power in all markets. If it faces an infinite elasticity of demand in one of the markets, the fixing of marginal cost is simple. It is whatever price the monopolist gets in that market.

But even if it does not face infinite elasticity, the Rule of Rational Life holds. The monopoly shifts outputs among the various markets in which it sells until *marginal revenues in all markets are equal to each other and equal to the common marginal cost*. Doing so will in general lead to unequal prices that cannot be explained by unequal costs, called *price discrimination*. For example, airlines charge lower prices for children than for adults, even though the marginal cost of providing a seat to Emily Redfield, who is 15 years old, is identical to that of providing one to Louise Bolduc, who is 40. For the child's seat the equation for the ratio of marginal cost to price is as usual $MC/P_c = 1 - 1/\epsilon_c$. Likewise for the adults. Multiplying each by its price will connect demand elasticity to marginal cost. In equilibrium, then, that the two marginal revenues are equal to the common marginal cost implies the following:

The Fundamental Equation of Discriminatory Monopoly

$$P_c \left(1 - \frac{1}{\epsilon_c}\right) = P_A \left(1 - \frac{1}{\epsilon_A}\right)$$

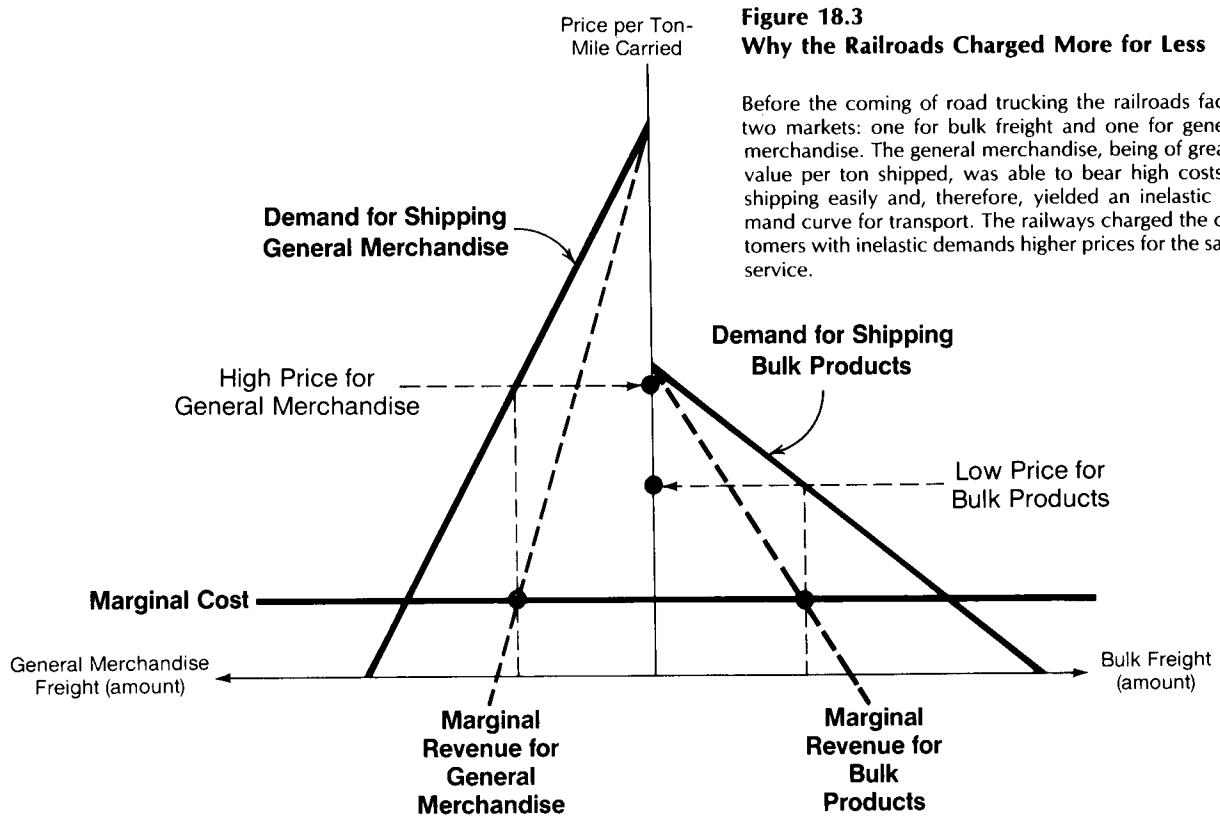
Price for the child	×	1.0 – the inverse of the elasticity of demand by children	=	Price for the adult	×	1.0 – the inverse of the elasticity of demand by adults
------------------------	---	--	---	------------------------	---	--

Children have a more elastic demand than do adults. The child's ride is usually not a business trip, for example, but a trip to Granny's that can easily be forgone. If the elasticity of demand by adults were, say, 1.5, the elasticity by children might be 3.0 or 4.0. Putting these values into the formula gives $P_c(\frac{2}{3} \text{ or } \frac{3}{4}) =$

$P_A(\frac{1}{3})$. The adult, relatively inelastic market pays 2 ($= \frac{2}{3} \div \frac{1}{3}$) or 2.25 ($= \frac{3}{4} \div \frac{1}{3}$) times more for the same seat. Try it out.

A useful diagram exhibiting a monopolist that sells in two (or in general many) markets draws the demand curves back to back, sharing a common price axis. If the marginal cost is constant, the diagram serves to determine how much is produced as well as how it is divided between the two markets. A railroad in 1920, for example, faced *low-elasticity* customers shipping general merchandise (freight such as clothing and farm machinery with high value relative to the shipping cost and with few alternative shippers). But it also faced some *high-elasticity* customers shipping bulk products (freight such as coal and wheat with low value relative to the shipping cost, and with many alternative shippers). It maximized profit by setting Marginal Cost per ton-mile, assumed to be the same for both, equal to Marginal Revenue in both markets (see Figure 18.3). The result was a high price for general merchandise and a low price for bulk products. (The year 1920 is chosen because shortly thereafter long-distance trucking stole the general merchandise from the railroads, "skimming the cream" as it was called. Unfortunately, this led to successful requests to bring trucking under the high-price rules of the Interstate Commerce Commission).

The general principle is that the inelastic demand pays a higher price, because it will tolerate the higher price.



Q: Marshall Field, a department store in Chicago, offers to pay up to \$30 for an old piece of luggage as a trade-in on the purchase of a new one. The old luggage is worthless to Marshall Field, which throws it away. The only effect is to reduce the price to the buyer with old luggage by \$30.

1. What do you suppose the difference in elasticity of demand is between a new buyer of luggage (newly-weds, say, with no old luggage) and an old buyer (who already has some serviceable but battered luggage)?

2. Now explain the \$30 discount and other trade-ins involving goods with no secondhand value (refrigerators, televisions, and so forth).

A: 1. The elasticity of demand for the new buyer is probably lower than is that of the old. The old buyer already has luggage, which competes with a possible purchase.

2. In other words, Marshall Field is price discriminating. It detects a high-elasticity buyer (to whom it charges a \$30 lower price) by his ownership of old luggage. Likewise with trade-ins elsewhere. The old refrigerator in your house competes for the new, making you (unlike the first-time buyer) less willing to stand still for a high monopoly price. So you pay less.

**Price
Discrimination Is
Evidence of the
Existence of
Monopoly**

Price discrimination, then, is evidence for the existence of monopoly power, serving even to measure its degree. A doctor, for example, commonly charges lower prices for poor than for rich patients. Putting it so suggests that her motive is charitable. But putting it the other way, that the doctor charges higher prices for rich than for poor patients, suggests an alternative explanation: that she is able to price discriminate, charging a higher price to the presumably less elastic demanders with high incomes.⁸

T or F: Giving partial college scholarships that depend on the student's income is price discrimination.

A: The low-income student will not come to the college (read "buy the product") without a scholarship (read "without a lower price"). The low-income student's elasticity of demand is higher. The same is true even of scholarships based entirely on "merit" (merit being excellent College Board scores or high school grades or, for that matter, unusual ability to stuff a large ball in a small basket while dressed in underwear and sneak-

ers). People with such abilities have many colleges from which to choose. That is, they have a high elasticity of demand for the services of the University of Iowa. The economics of scholarships often puzzles observers. They wonder how it could be that a college can make more in tuition money by "spending" scholarship money. The answer is that the college is not in fact spending anything; it is merely offering discounts to students with especially elastic demand curves, in the style of a discriminating monopoly. So, true: Scholarships are price discrimination.

The exercise of discriminatory monopoly is impossible if the good can be resold. If the child could resell his airplane seat to an adult, if the new owner of luggage could resell to the old, if the poor patient could resell his medical care to a rich patient, or if a scholarship student could resell her place to a nonscholarship student, the two-price system would break down.

Q: Seen on Bronco Superfine Toilet Tissue: "Commercial Pack, Not for Resale to the Public." Why not?

A: Because with resale the Bronco Company could not charge two prices, a high one to retail customers at the grocery store and a low one to commercial customers

with many alternatives available to them. In like fashion, makers of a plaster substance used in both taking casts of teeth and in fixing walls added poison to the wall-fixing version. If by law, by poison, or by the personalized nature of the product a resale is made impossible, then price discrimination is made possible.

⁸ See Reuben Kessel, "Price Discrimination in Medicine," *Journal of Law and Economics* 1 (October 1958): 20-53.

Other Methods of Discrimination

The underlying reason a monopoly makes more profit by discriminating than by charging a uniform price is that freeing it of the constraint to charge a uniform price must make it better off. So it is with other pricing schemes. If the monopoly is free to offer different meals to different consumers it can extract more money. By contrast, if it is a simple, single-price monopoly, it can only charge one price to all consumers and let them buy what they wish at the one price. If the monopoly has the power to go further, especially the power to stop consumers from reselling among each other, then it can present consumers with more complex deals. The deals must make it still richer if it in fact uses them. The more complex deals are all "price discrimination," and all share with two-person price discrimination the feature that different consumers are treated differently.

The simplest complexity is an entry fee, Mafia "insurance" premium, membership fee, tribute payment, installation charge, or cover charge that varies with the person. Businesses, for example, are charged more for installation of telephones than are household customers. In the extreme *such a charge can extract all the consumers' benefit from trade*, by adjusting the fee to each consumer's consumers' surplus. The charge that the marginal consumer would accept, however, is of course zero, since he has by definition no consumer's surplus to extract. Therefore if the marginal consumer can trade with more eager customers, a variable fee cannot be sustained. If a club charges membership fees graduated by income, for example, the low-income members will sell their memberships (if they are transferable). No high-income member will in fact pay the high fee, since he will be able to more than compensate a low-income, low-fee member for giving up his membership. The variable fee will disintegrate into a single and simple monopoly.

A closely related scheme, often adopted by the same monopolists who charge discriminating fees, is to charge lower prices per unit to those who buy a lot. The offer of one box of Kleenex for \$1, two for \$1.50 is such a scheme. Electricity, for instance, is commonly sold according to a schedule of declining rates as more is purchased. The similarity between such *multipart pricing* and the discriminating fee is illustrated in Figure 18.4. The Sock-It-to-Them Electricity Monopoly can charge Lars Sandberg the shaded area as a fee and then sell him electricity at marginal cost. Alternatively, the monopoly can present Sandberg with the schedule called Imperfect Multipart Price. The schedule charges him more for early amounts than for late. It induces him to buy out to More Profit and yields more profit to the monopoly than does the Simple Monopoly Price (which gives only the inscribed rectangle of profit). With perfect knowledge the monopoly could arrange the perfect multipart price, namely, Sandberg's demand curve itself, selling him the last unit (55 kilowatt-hours) at marginal cost. The profit from such a scheme would be identical to the largest possible fee, scooping out all the consumers' surplus.

Another and better interpretation of multipart electricity pricing is not in fact that it charges less for Sandberg's "last" unit of consumption than for his "first" but that it simply charges Sandberg, as a residential consumer, less than Baack Bauxite Refining, which consumes vastly more electricity. The parts of the multipart price can be thought of as applying to different demand curves, not to different parts of one person's demand curve. The case is interpretable as the elementary one of two consumers with different elasticities. If high-volume consumers have higher elasticities at the same price than do low-volume consum-

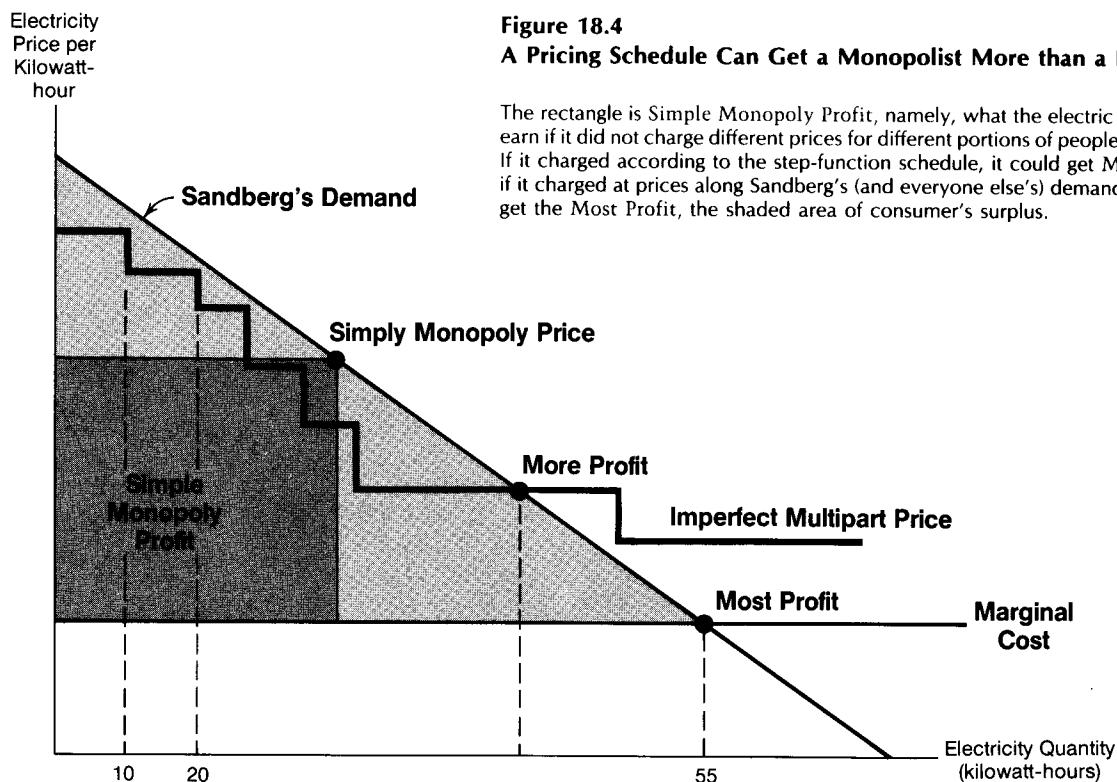


Figure 18.4

A Pricing Schedule Can Get a Monopolist More than a Mere Price

The rectangle is Simple Monopoly Profit, namely, what the electric company could earn if it did not charge different prices for different portions of peoples' consumption. If it charged according to the step-function schedule, it could get More Profit. And if it charged at prices along Sandberg's (and everyone else's) demand curve, it could get the Most Profit, the shaded area of consumer's surplus.

ers, then it will be rational for a monopoly that can prevent resale to charge the high-volume consumers less. That the monopoly expresses the price as a schedule is really irrelevant. It is in fact using volume as a meter to detect high-elasticity demanders, just as low age is used to detect high-elasticity demanders of airfares or movie seats.

The Apparent Discrimination May Be a Cost

The presence of quantity discounts, however, does not invariably indicate monopoly. After all, it is possible that large quantities are in fact cheaper per unit to sell, because fixed costs of ordering are spread over the larger volume. The high-income member of a club, likewise, might well be a more expensive member to serve, making for example more demands on the staff. In other words, a defense to a claim of price discrimination is always that the products sold at different prices to different consumers have in fact different costs. The defense will sometimes be persuasive.

The more general point is that it is difficult to detect monopoly. Even discrimination among customers, the most public evidence of monopoly power, can sometimes be explained as a competitive result of differences in cost. No wonder, then, that the evidence on the prevalence of monopoly is not all in.

Summary

The question of whether the economy is predominantly monopolistic or competitive has been asked in a great variety of ways by economists over the past 50 years. The answer is important, affecting the enthusiasm with which one views

market outcomes. The ways of measuring monopoly are many, but all are in some respect inconclusive. Profit above normal is one measure, but it is often confounded with normal rent or dissipated in seeking to acquire and protect the monopoly. The height of barriers to entry is another measure of the profit, advertising being one example, and itself direct evidence that the advertiser is not a price taker. But the significance of the barrier is difficult to measure. The number of firms in the industry, a third possible measure of monopoly, is easy to measure but is irrelevant. The size of the *fringe* is what is relevant to the use of monopoly power. A fourth measure looks directly at the behavior of a monopoly, noting that it sets its price above marginal cost. The trick, then, is to measure marginal cost. Price discrimination between elastic and relatively inelastic demands is a good source of evidence on marginal cost, as when a monopoly charges its export customers less than its domestic consumers. The idea is a rich one, applicable to all manner of markets, from the steel cartel in Germany to the airlines in America. Price discrimination comes in many packages, all of which fall apart if the victims can resell the monopolized good. But such uses of the measure of monopoly are more successful than is their use to answer the original question of prevalence. The question remains unanswered. Is the modern economy predominantly competitive? We are not sure.

EXERCISES FOR CHAPTER 18

1. Will the following people earn more than normal returns on their investments? If not, where will the monopoly profits be hidden?
 - a. C. K. Harley, the purchaser of a liquor store in Chicago, the license for which sells at \$20,000.
 - b. Stephen Easton, who bought stock in Morton Salt at a time when Morton had substantial monopoly power.
 - c. Brad Lewis, who went into lengthy and expensive training to become a union electrician at high monopoly wages in Michigan.
2. What would be the elasticity of demand facing Inland Steel if Inland produced 1 in every 25 tons of steel produced, the overall elasticity of demand were 0.5, and the elasticity of supply from other steel companies were 2.0? If Inland exploited this power, what would be its ratio of marginal cost to price? Within what error is the ratio virtually 1.00?
3. Where or when will these show a price equal to marginal cost?
 - a. A steelmaker in Pittsburgh selling at home to the European market.
 - b. A conspiracy of gasoline dealers that is only occasionally effective in setting the monopoly price.
4. If the price of gasoline is \$1.50 when the local conspiracy of gasoline dealers is working well and \$1.20 when it's not working at all, what do the conspiracy judge their elasticity of demand to be? (*Hint:* Suppose that their marginal cost is unchanging, and is only revealed when their conspiracy collapses; use the Fundamental Equation of Monopoly.)
5. If a monopolist thought he faced an elasticity of demand of only 2.0, what ratio of marginal cost to price would he arrange?
6. If the elasticity of demand of ordinary students at the University of Iowa were 1.33 and the elasticity of demand of basketball players were 20, what would be the ratio of prices (tuitions and other costs) charged to ordinary students and basketball players, supposing (contrary to fact) that the University of Iowa is a grasping monopoly? (*Hint:* Use the Fundamental Equation of Discriminatory Monopoly.)

7. Explain the support for price discrimination in the following:
- Senior citizen discounts in movies, on presentation of evidence of being over 60.
 - Discounts on Enormous Family Size washing detergent for clothes.
 - Discounts on a new one if you turn in your old TV.

PROBLEMS FOR CHAPTER 18

1. In the stock market the right to the future profits of a firm is priced. If George Marr and Co., Inc., is expected to earn, say, \$1000 for the next 15 years and then to go broke, the total value of the company's stock (ignoring the lower value of distant earnings) will be $15 \times \$1000$, that is, \$15,000. *True or false:* If the stock market is functioning correctly, dividing income per year by the value of the stock of a firm will always give the competitive rate of return, even if Marr and Co. is a monopoly.
2. The Justice Department accused Citibank and some other big U.S. banks in October 1978 of conspiring to make money by forcing down the price of the dollar. What one number characterizing the market for dollars would you want to have as an economist deciding whether or not the alleged conspiracy could have been effective? (*Hint:* Look at the Inland Steel problem in the text.) And what value would you like the number to take as an attorney for Citibank attempting to rebut the Justice Department's accusation?
3. In the text of the last section it was argued that the owner of a patent on a cost-reducing invention for hand calculators would not always charge all the cost reduction as a royalty; that is, the owner would sometimes charge a royalty that allowed price including the royalty to fall from its preinvention price. To find out when "not always" and "sometimes" occur, view the owner as a monopolist in the market for hand calculators, making calculators at some fixed marginal cost and selling them at a markup. Suppose that the older, more expensive technique was not available. According to the elasticity formula for monopolies, what would be the profit-maximizing markup? What is the markup if the older technique is available and the optimal markup is greater than the ratio of the old price to the new marginal cost? Try out some values for the elasticity and the old-new cost ratios. What do you conclude about the likelihood of passing cost savings on to consumers?
4. Consider the Sony Television Division selling televisions at home in Japan at a high monopoly price and in the United States at a low competitive price.
 - Draw Sony's entire marginal revenue curve. Watch for kinks.
 - Draw a U-shaped marginal cost curve that would induce Sony to sell some abroad. At the equilibrium point what is the total revenue (use the area under marginal revenue)? What is the total variable cost? What is total profit (net of variable cost, that is, ignoring fixed cost)?
 - Sony is often accused of *dumping* in the United States. The definition of "dumping," however, varies. According to the model of a monopoly just developed, would Sony's behavior fit a definition that dumping is "charging a lower price to foreign than to home customers"? How about "charging a price to foreigners that is below *marginal cost*"?
 - Dumping is more usually defined as selling below *average cost* (whether average total or average variable cost), since average cost is easier to observe. Taking the extreme, would it ever be rational for Sony to price its televisions in the United States below average *variable cost*? That is, could Sony still be making money? Use U-shaped marginal and average cost curves and what you know about the necessary relations between average and marginal curves.
- A local monopolist sells in two markets, home and abroad. At home he supplies 100% of the quantity consumed; abroad he sells 0.0000000001%. The price he charges

at home is observed to be 50% above the price he gets abroad. *True or false:* Therefore, the elasticity of the demand curve he faces at home is 3.0.

6. Alcohol refined from vegetable matter (such as corn) is commonly used as automobile fuel in Brazil. It is drinkable, being pure grain alcohol. The government, however, makes it poisonous by adding a little gasoline. Why?

7. A standard example of discriminatory monopoly is the case of International Business Machines a long time ago. As the case is usually described, IBM would not sell its large-scale electronic computers, only lease them at so much per month (the same for any user). It required that its customers buy computer cards from IBM, at a price above what a competitive firm selling cards would charge. The practice is alleged to have been a devilishly clever method of price discriminating (that is, charging different prices to different customers). The customer who used her leased machine a lot would also use a lot of cards and would therefore pay a lot to IBM *per unit of computing service*. The customer who used it a little would pay a little per unit of computing service. The big user, the analysis goes, pays more than the little user, making the scheme a good example of price discrimination.

a. If the argument is put, as it has been so far, in terms of the price of a unit of computing services—say, one multiplication of two five-digit numbers—is it correct? That is, is the big user charged more per unit of computation?

b. How can one restate the argument in correct form? Suppose that the product on the horizontal axis is the number of computing machines instead of the number of computations? What is on the vertical axis? How much is a big user willing to pay compared with a little user?

True or False

8. That all the major coffee canners raised their prices at the same time and in the same amount is evidence of monopoly in the coffee canning industry.

9. The opening of an opportunity by a monopoly at home to sell abroad as a competitor will if anything raise the price at home.

10. Barbers charge less for children because it is less costly to cut the hair of children.

11. The practice of bringing out the hardback edition of a book well before an identical paperback is an example of price discrimination.

o 12. A discriminating monopolist will charge in each market the price that maximizes receipts in that market.

13. American Motor Company's offer to give back \$200 of the price of small cars to buyers over the age of 65 reflects their concern for the aged.

o 14. Since it can be shown diagrammatically that straight-line demand curves with a common price at which the quantity demanded is zero have at the same price the same elasticity, such demand curves should not be used in an analysis of price discrimination between two markets.

CHAPTER

19

The Welfare Economics of Monopoly

What to Read For

What social purposes are used to defend monopolies? What is the infant industry tariff? Can static economic theory decide whether or not monopoly is good for technical innovation? How is a monopoly's behavior represented in an Edgeworth box? Can the monopoly itself be made better off by exchange beyond the monopoly point? What does the cost of making agreements have to do with the efficiency of a monopoly? Is a perfectly discriminatory monopoly inefficient? Is the loss from sacrificed exchanges in simple monopoly large? Is this loss the largest loss attributable to monopoly?

Monopoly May Subsidize Some Desirable Activity (or May Not)

Monopoly has its defenders, chiefly the monopolists themselves. The most elementary and common defense of a monopoly is that it is good for some other social purpose. The Postal Service's monopoly of first-class mail is defended on the grounds that the Service can thereby subsidize unprofitable customers, such as remote farmhouses. The United Automobile Workers' monopoly of labor sold to automakers is defended on the grounds that the union can thereby protect workers from arbitrary treatment by the automakers and can in the end improve productivity. The government's monopoly of violence is defended on the grounds that the government can thereby supply the citizens with many valuable services (such as the construction of a new dam where none is wanted or the enforcement of monopoly prices for interstate furniture moving).

Q: Restricting entry to medical school (by restricting the number of places) is often defended on the grounds that it improves the average quality of doctors and therefore improves medical care. Even if the restriction does improve the quality of doctors, this question will show, the argument is not decisive. Is it socially desirable that only the students with the greatest aptitude and affection for academic work be allowed into medical school?

Compare a high-quality (and expensive) doctor with a high-quality (and expensive) automobile. Is the amount of transportation maximized by allowing only Rolls-Royces to be made?

A: What the apologists for the monopolists are forgetting is that the bright students attracted by the high monopoly rewards (coupled with barriers to entry that make

only very bright and hardworking students admittable to medical school) have alternative employment elsewhere. The social benefit of their IQs and energy might well be higher elsewhere. (Does your local sawbones *really* need to have gotten straight As in organic chemistry at Cal Tech?). And, of course, a high quality of doctors does not imply that the quality of medical care,

given its low *quantity*, is high. We could improve the quality of automobiles by requiring that everyone drive a Rolls-Royce, turning over to the Rolls-Royce company the task of certifying new entrants. The automobiles driven would be of high quality, to be sure. But there would be so few of them that the quality of transportation would fall.

One must always ask, in other words, whether a grant of a monopoly is the most efficient way to a social end. The standard case in point is the *infant industry tariff*, that is, a very high tax imposed on foreign goods to give a young industry at home a chance to grow up free of foreign competition. The high tariffs on cotton textiles and iron imported into the United States in the nineteenth century were commonly defended on such ground. Protect our monopoly for 10 (or 20 or 100) years, said the American industries, and we will *eventually* produce cotton cloth and iron rails even cheaper than the British. The argument is inconclusive. For one thing, the "eventual" low prices might not justify the many intervening years of high prices. This is especially true if the lack of foreign competition makes it possible to sustain prices at home above even the high American cost. For another, there may be a more direct and less expensive way of teaching Americans to make cloth and rails as cheaply as the British. The government could subsidize research, set up a demonstration plant, or send technicians to Britain to study. In the nineteenth century such alternatives were politically unimportant, but in the twentieth century they are in fact taken seriously by some developing countries (most developing countries give in to the appeals for tariffs, though). In the few instances in which the governments have been able to resist agitation to use tariffs to stop foreign competition, the countries have been made better off. They have gotten the industries they should, but by more efficient means than official monopolies.

Monopoly May Encourage Economic Progress (or May Not)

A closely related defense of monopoly is the one argued in 1942 by Joseph Schumpeter, the great theorist and historian of industrial capitalism:

The introduction of new methods of production and new commodities is hardly conceivable with perfect—and perfectly prompt—competition from the start. And this means that the bulk of what we call economic progress is incompatible with it. . . . In this respect, perfect competition is not only impossible but inferior, and has no title to being set up as a model of ideal efficiency.¹

He means that monopoly, like a patent, protects the rewards of innovation from erosion by competition, giving more incentive to seek innovation. He admitted, however, that "it is certainly as conceivable that an all-pervading cartel system might sabotage all progress as it is that it might realize, with smaller social and private costs, all that perfect competition is supposed to realize."² From a theoretical point of view, in other words, it is unclear whether monopoly or competition is best for innovation. On the one hand the monopolist might reap the benefits of innovation; on the other the monopolist might have less

¹ Joseph Schumpeter, *Capitalism, Socialism and Democracy*, 3rd ed. (New York: Harper Torchbook, 1962), pp. 105–106.

² *Ibid.*, p. 91.

incentive to initiate the development of new technologies. In the static world of most economic theory the issue can hardly be studied, much less resolved.

T or F: Because it was a monopolist in many markets and had no incentive to keep costs down, the old American Telephone & Telegraph Company had higher costs than would a competitive telephone industry.

A: Set aside the problem of my grandmother and her wheels. Lower costs translate into larger profits no matter what the state of competition. A dollar of additional value to AT&T stock is the same as a dollar of additional

value to Douglass North's local telephone service. Therefore, false. Only a misapplication of diminishing marginal utility would suggest that a monopoly swimming in profits would want them less than thirsty little competitors for your telephone business. Its monopoly might have dulled the vigor and raise the costs of the telephone company as a matter of fact. But as a matter of the simple logic of static economic theory the result is not necessary.

Likewise, "planned obsolescence," (an alleged conspiracy against durable products) is often attributed to monopolies. We have all heard (false) rumors that such-and-such a big company has kept the long-lived light bulb or the long-lived car off the market. But the market form has little to do with the logic of "planned obsolescence."

Q: Alec Guinness was *The Man in the White Suit* (1951), who invented an indestructible and unsoilable fabric only to find that clothing manufacturers would not use it. To use it, they said, would ruin their business, which depends on the wearing out of old clothing. Would the fabric be used?

A: If the fabric lasted forever, it could sell at the value (allowing for changes in style) of the infinite series of ordinary pieces of clothing that it replaces. A monop-

oly of clothing manufacturers would be delighted to sell it at this price, or in any event at the price that maximized profit. A competitive clothing industry would be forced to sell it at the lower price of its cost of production—forced, that is, by the attempts of each manufacturer to steal a share of the market. In either case the fabric would be used. That is, yes. The fabric would be used, contrary to the bit of ersatz economics in the movie.

The logic is not decisive, or course, for in such matters logic unassisted by fact is never decisive. The facts may be that monopoly *does* nonetheless have a good or bad effect on innovation. The logic implies only that the usual postulates of economic thinking do not lead to such a conclusion. In any particular case the postulates may be wrong. Postulates are like that in economics. If the executives and owners of the telephone company did in fact think of themselves as sheltered, then innovation would be retarded. If each clothing manufacturer did in fact think his own refusal to adopt the miracle fabric would stop its general use, then again innovation would be retarded. On the other hand, if each monopoly did in fact think of its monopoly profits as a source of funds for investment in laboratories, then innovation would be advanced. Such irrationalities are possible. The point is merely that the usual theory of rationality cannot easily accommodate them.

Monopoly Does Produce Too Little Output

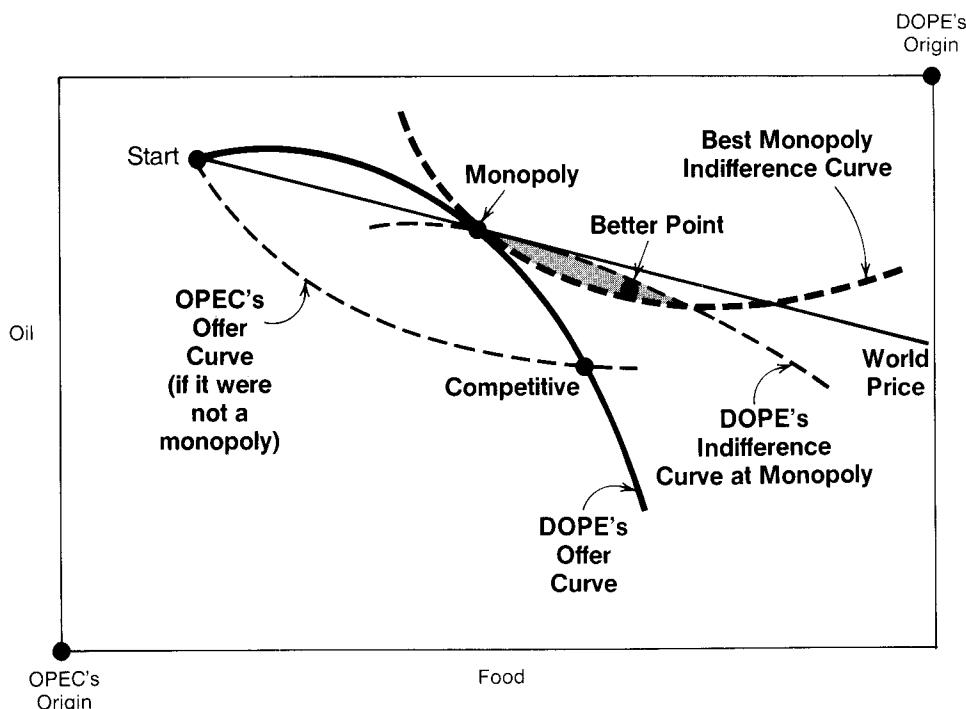
Whether or not the usual theory can show that monopoly retards progress in the development of new techniques, it certainly can show that monopoly produces an inefficient level of output, given the existing techniques. An earlier section used supply and demand curves to show this roughly. The rough argument, you recall, is very simple. A monopolist spoils his price if he produces

the competitive level of output. So he produces less. But the competitive level is efficient. So monopoly produces less than the efficient level of output. The Edgeworth box can now be dusted off and set working to show this in detail.

Recall that the Edgeworth box shows an exchange of two goods between two types of people. Suppose, for example, that the Organization of Petroleum Exporting Countries (OPEC) gives its oil in exchange for the food of the Disorganization of Provender Exporters (DOPE) (see Figure 19.1). The one mind of OPEC faces the many competing minds of DOPE. OPEC maximizes its happiness in view of its effect on the World Price ratio between oil and food. Each member of DOPE (and therefore the entire group, since the members do not act together as a monopoly) maximizes its happiness in view of what it takes to be a given, unalterable World Price. The constraint on OPEC's maximization is merely that it stay on DOPE's Offer Curve. In other words, OPEC chooses the point Monopoly

Figure 19.1
The First Astounding Fact: The Exercise of Monopoly Leaves a Region of Unexploited Exchange

The exercise of monopoly power by OPEC consists of maximizing OPEC's happiness *subject to the constraint of the DOPE's Offer Curve*. Therefore, the monopolist's indifference curve at the Monopoly point will be tangent to DOPE's Offer Curve. Therefore, it will *not* be tangent to DOPE's indifference curve. Therefore, there will exist a lens shape of mutually advantageous exchange unexploited. Note that the unexploited exchange is *mutually* beneficial. That is, even the monopoly can be made better off if further exchanges could somehow be arranged.



on DOPE's Offer Curve that puts OPEC on the Best Monopoly Indifference Curve. (Exercise 2 at the end of the chapter goes through it step by step.)

Notice that at Monopoly the DOPE's Offer Curve, but not its indifference curve, is tangent to an OPEC indifference curve. That's the point: OPEC's monopoly power leads to less exchange than at Competitive. Notice, too, that at Monopoly the definition of an offer curve implies that the World Price line going through the point (and therefore *through* rather than tangent to the offer curve) is tangent to DOPE's Indifference Curve. That is, at Monopoly the slopes of the two indifference curves are *not equal*. That is, *at monopoly there is still a shaded lens of mutually advantageous exchange left to be taken*. That's the point: Monopoly is inefficient.

**Even the
Monopolist Is Worse
Off**

This is the first of three odd features of monopoly that the Edgeworth box makes plain. There are at Monopoly still *mutually* better deals. Monopoly leaves *even the monopolist firm* worse off than it could be if all opportunities for mutually advantageous exchange were taken. The monopoly is of course better off at the point Monopoly than it would be if (like DOPE) it ignored its monopoly power and did not conspire to raise the price. But both the victim and the monopoly could be made even better off if they could somehow agree to seize the opportunity for more exchange, moving to a Better Point within the shaded lens.

**Transaction Costs
Are the Cause of
Inefficiency**

The second odd feature is that the cause of single-price monopoly and its inefficiencies is *merely the cost of making agreements*. It was just said that "if the monopoly and its victims could somehow agree to seize the opportunity" all would be well on the efficiency front. The agreement, however, is expensive to make and to enforce. For example, if OPEC wished to move from Monopoly to a Better Point within the shaded lens, it would need to enforce an agreement to charge a lower price for the *additional* oil sold. Only the lower price would induce the victims to buy more. But to enforce the agreement, OPEC must be able to keep old oil (sold at the high price) distinct from new oil. Otherwise each member of OPEC will have an incentive to sell more old oil by calling it new, taking advantage of the high price and thereby spoiling it. And customers will have an incentive to buy only new oil, spoiling the market for old oil. All in all, the price discrimination necessary to get to the Better Point is difficult and expensive to maintain.

A more extreme way of stating this second odd feature is the following:

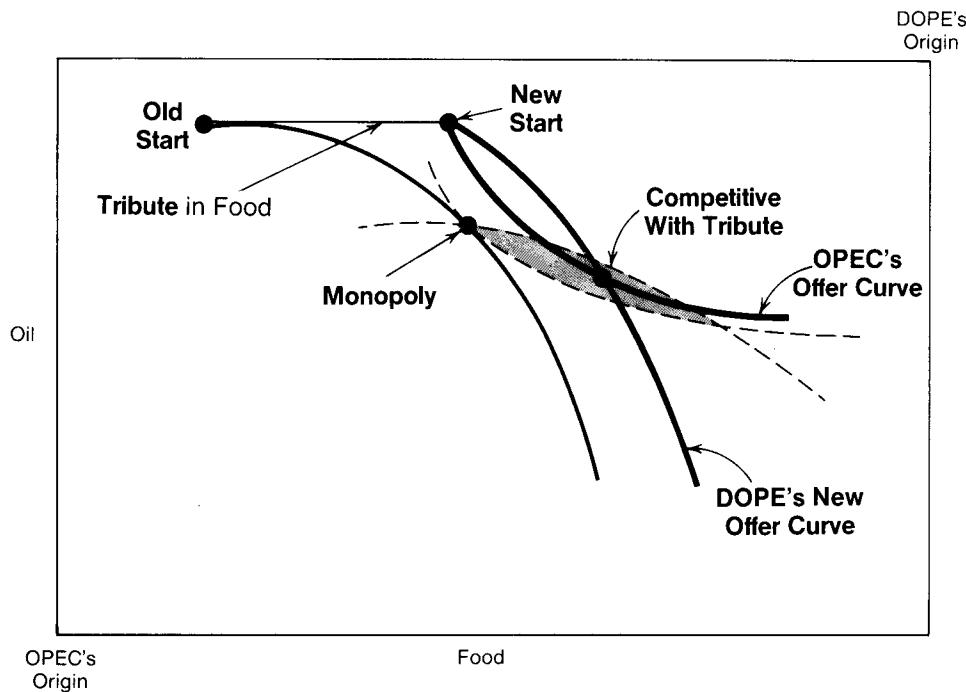
T or F: If the victims of OPEC would agree to pay tribute to OPEC in exchange for an agreement by OPEC to sell its oil competitively (that is, along OPEC's own offer curve, instead of price searching along the victims' offer curve), both OPEC and its victims would be better off. (*Hint:* See Figure 19.2. Note that the tribute would shift the Old Start point of the initial endowment; imagine the two offer curves of the two parties coming out of the New Start.)

A: True. After the Tribute is paid the heavily emphasized offer curves govern the behavior of OPEC and its vic-

tims. The size of the Tribute, you see, can always be manipulated so that the equilibrium point Competitive with Tribute falls somewhere within the cigar shape coming out of Monopoly. That is, the agreement results in an improvement for both parties over what they got at Monopoly. The equilibrium, in fact, is a competitive equilibrium, on the contract curve. Only the costs of making the agreement prevent all goods (including oil) from being sold competitively.

Figure 19.2
The Second Odd Feature: Only Transactions Costs Stand Between Monopoly and Competition

There always exists a bribe, or Tribute, that the victim could pay the monopoly (in return for an undertaking by it to pretend it is a competitor) that would make both better off. The monopoly is made better off by the bribe, and the victims are made better off by the lower price. To be more precise, the bribe-and-marginal-cost-pricing scheme brings the little society into the lens-shaped area of mutually advantageous exchange coming out of the former Monopoly point. If arranged properly the deal can get the society to an efficient point, from which there are no further lenses. But arranging deals properly is expensive. In other words, it is merely transactions costs that allow the inefficiencies of monopoly (or any other inefficiency) to persist.



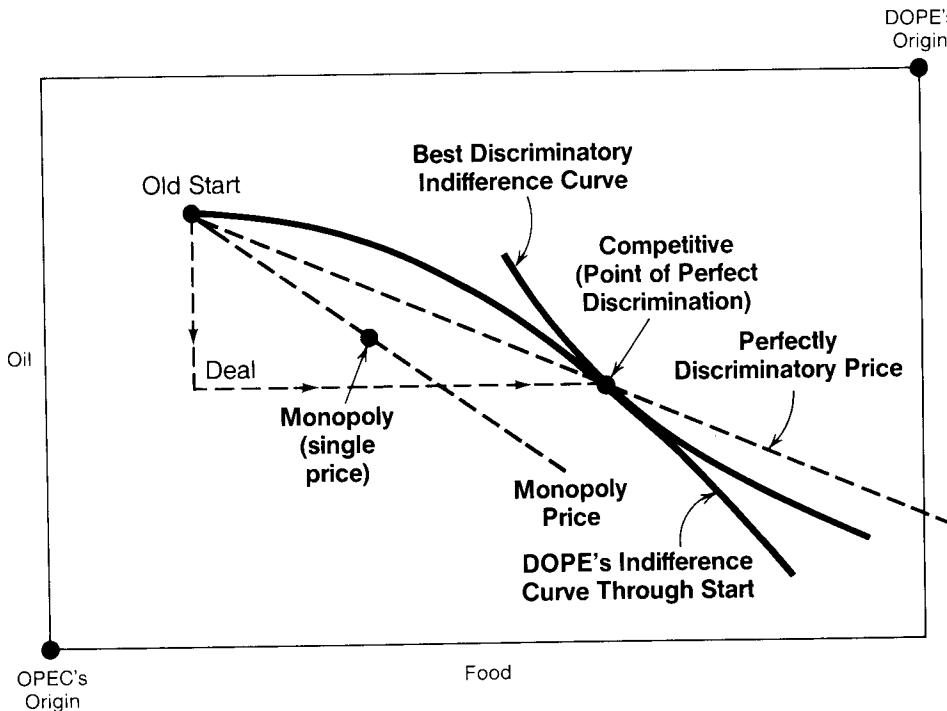
Perfect Price Discrimination Is Efficient: Edgeworth Box

The third odd feature of monopoly is closely related. It is that the opposite extreme from price-taking competition, namely, multiprice monopoly, is efficient. In other words, only *clumsy* attempts to extract profit from victims are inefficient. Skillful attempts in fact arrive on the contract curve (like price-taking competition). And being on the contract curve is efficient.

There are various ways of demonstrating this, each corresponding to a different set of monopolistic institutions. For Example, in the Figure 19.2 just used, one can imagine that OPEC sets the Tribute (or, in other contexts, the entry fee or membership fee) so high that at the Competitive equilibrium the victims are no better off than they were at the Old Start (see Figure 19.3). Such victimization is success in life for a monopoly. It can clearly extract no more than achieving

Figure 19.3
The Third Odd Feature: Perfect Discrimination Is Efficient

A perfectly discriminatory monopoly will be able to get to its Point of Perfect Discrimination namely, that point at which the monopoly's own happiness is greatest and the happiness of its victims is at least no lower than it was at Old Start. The deal the monopoly could arrange to achieve such a result would be an all-or-nothing offer that moved directly to the Point of Perfect Discrimination. Alternatively, the monopoly could extract as much as possible (that is, along the demand curve) from the first little bit of oil it sold, then as much as possible from the second, and so forth, arriving finally at the same point.



the point where it is as well off as possible subject to the constraint that its victims, who can refuse to trade altogether if it comes to that, are no worse off than at Old Start. But such a point is on the contract curve and is efficient. It may not be just or good. But it is efficient. Likewise, consider the following.

T or F: A trade union selling hours of labor in exchange for goods can arrive at the point of perfect discrimination by setting both the wage and the number of hours worked.

A: True. Look at Figure 19.3. Imagine OPEC's origin to be the trade union's origin and imagine DOPE's origin to be the buyers'. The trade union starts with many hours of labor (measured along the oil axis) but little

money (along the food axis). If it merely sets a monopoly price and lets the buyers choose how many hours they want, it will end up at the Monopoly point. But if it also specifies the amount of hours to be bought and can enforce this specification, then it can achieve the dashed line of the Deal that brings it to the Point of Perfect Discrimination. Notice that the price per hour is lower than under simple monopoly (the slope of the price line is lower) but that the number of hours is

larger. That is, perfect discrimination offsets the tendency of monopolies to produce too little. Notice, too, that the same point could be reached by charging the

entry fee, installation charge, Mafia "insurance" premium, membership fee, tribute payment, or cover charge discussed in Chapter 18.³

Perfect Discrimination Is Efficient: Supply and Demand

The argument can be translated into supply and demand. The simplest case is one in which the income effect on the consumption of, say, electricity is zero for both buyers and sellers. (The analytical purpose of the assumption is, as usual, to prevent supply and demand curves from moving about when there are changes in the distribution of income.) In the top panel of Figure 19.4 the condition is exhibited by the vertically parallel indifference curves (drawn as dashed lines), yielding a vertical Contract Curve (drawn as the heavy line).

As a result the curves of Demand and Marginal Cost in the bottom panel remain the same no matter how the gains from trade are divided between the monopolistic seller of electricity and its victims. Therefore, one can compare the three possibilities without confusion. The points of Competition and Discrimination are in the diagram of supply and demand exactly the same. Perfect monopoly under these assumptions leads to the same point as perfect competition. Schemes of price discrimination (which electricity companies in fact use) may bring the society closer to an efficient outcome than would less subtle monopoly.

- Q:** 1. In the bottom panel, what is the area of competitive profit to the electricity company? Simple monopoly profit? Perfectly discriminating profit?
 2. What is the sum of consumers' and producers' surplus under conditions of perfect competition, simple monopoly, and perfect discrimination? Therefore, what is the social loss from simple monopoly?

- A:** 1. A competitive electricity company moving along its own offer curve (marginal cost curve), would earn the triangular area *ORE* in Figure 19.4. The simple monopoly, by restricting output and raising the price, could

earn *ARE*, which will be larger than *ORE* if *A* is larger than *O*. The perfectly discriminating monopoly, however, could earn *all* the gain from trade, namely, *PARETO* (Vilfredo Pareto was an Italian sociologist and economist who long ago systematized the idea of efficiency).

2. *PARETO* is the sum of producers' and consumers' surplus for both perfect competition and perfect discrimination. The distribution of income differs in the two cases, but its size is the same. The sum for simple monopoly, however, is only *PARE*. The shaded area *TO* is the social loss.

Why the Efficiency Loss from Monopoly Appears to Be Small

There is nothing surprising here. Simple monopoly restricts output, sacrificing opportunities for trade. The sacrificed opportunities are the cost.

If one knew the shapes of the curves involved, one could calculate the size of the loss. When calculated in this way the national loss from monopoly is usually small.

- T or F:** The percentage loss from monopoly is on the order of half the percentage reduction in output multiplied by the percentage difference between price and marginal cost.

- A:** True. The argument is familiar from earlier discussions of taxation and rent control. In Figure 19.5 the

shaded area of monopoly losses is, geometrically speaking, half the emphasized rectangle. (*Proof:* Note that the demand and supply curves cut in half the two subrectangles above and below the dashed line.) Take the (observed) monopoly price and the (unobserved) competitive output to be 1.0. Having fixed the scale, the marginal cost is therefore $1.0 \times$ marginal cost as a pro-

³ Wassily Leontief, "The Pure Theory of the Guaranteed Annual Wage Contract," *Journal of Political Economy* 54 (February 1946): 76–79.

Figure 19.4
If the Income Effect Can Be Ignored, Perfect Competition and Perfect Monopoly Have the Same Equilibrium Point

The exaction of a perfectly discriminatory monopoly is income to the consumers. If the demand for the product is *unaffected* by income, then perfect discrimination and perfect competition, both being points of marginal cost equal to marginal valuation, will overlap precisely.

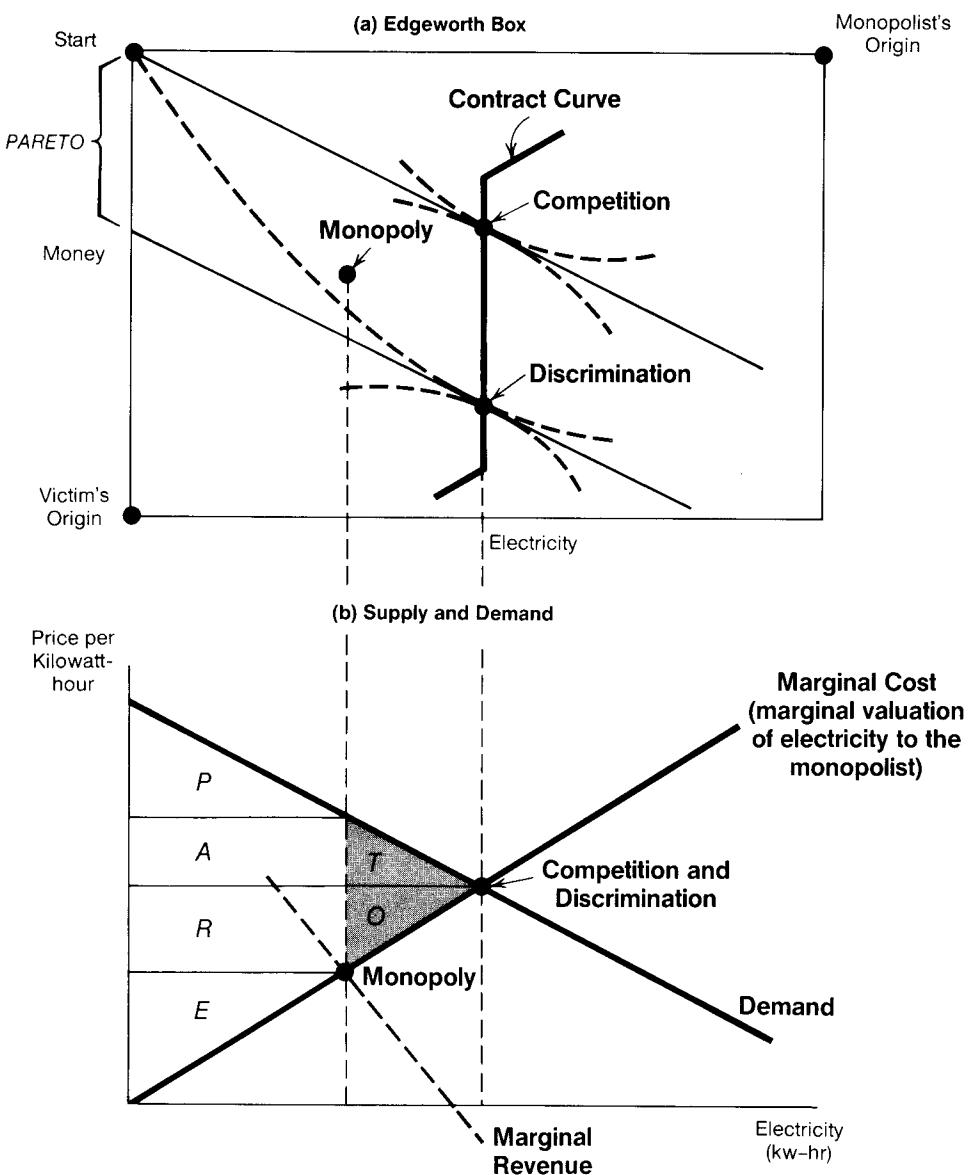
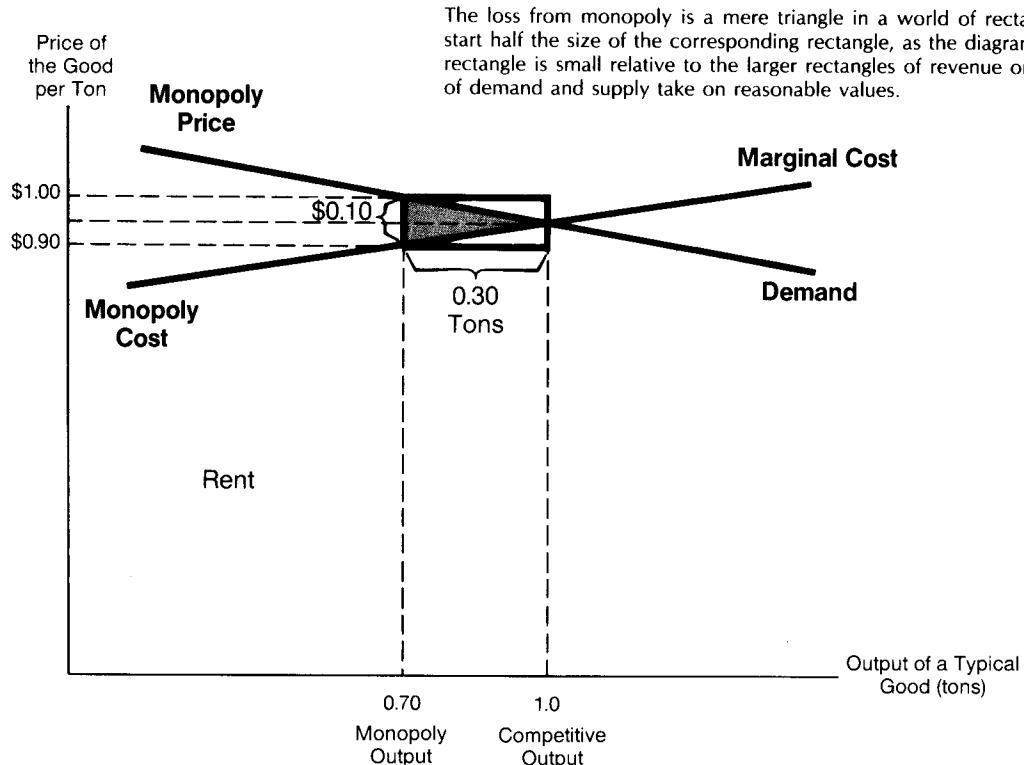


Figure 19.5
Why the Efficiency Loss from Monopoly Appears to Be Small



The loss from monopoly is a mere triangle in a world of rectangles. That makes it for a start half the size of the corresponding rectangle, as the diagram shows. Furthermore, the rectangle is small relative to the larger rectangles of revenue or cost when the elasticities of demand and supply take on reasonable values.

portion of price (that is, the measure of monopoly power developed earlier). And the monopoly output is $1.0 \times$ monopoly output as a percentage of competitive output. So the area of the rectangle is the product of the differences between the two. The typical marginal cost might be 90% of price. The typical elasticity of demand would be calculated by averaging the high elasticities facing the many competitive firms in the economy with the low elasticities facing the monopolistic firms. According to the fundamental equation of mo-

nopoly, if the typical ratio of marginal cost to price were 0.90, then the elasticity on average would be 10, a not unreasonable figure. If the corresponding fall in output were as much as 30%, the size of the shaded area would be $\frac{1}{2}(0.30)(0.10) = \0.015 . The whole national income earned in this typical industry is the price times the quantity, or $(\$1)(0.70 \text{ tons})$, or $\$0.70$. In other words, the typical efficiency loss from monopoly is small, a mere $\$0.015/\$0.700 = 2.1\%$ of national income.

Arguments such as this persuade many economists that the efficiency losses from monopoly (or, indeed, from other distortions, such as excessive regulation of trucking or taxes on food) are small.

Why Appearances Are Deceiving

The many economists, however, are mistaken. They are also mistaken, for the same reason, when they assert that perfectly discriminatory monopoly has no efficiency costs. The mistake is the familiar one of supposing that the competitive or monopolistic structure of markets does not itself have market causes, that it simply falls on the community like rain. The point is that monopoly of any

sort earns supernormal returns. Therefore people will seek these returns. And in seeking them they may spend resources, *wasting resources in addition to and quite possibly much larger than any little triangle of efficiency loss*. That is, the seeking of monopolies is itself an industry, and it might well use up resources.⁴

The argument is identical to the argument about the waste from queuing. If seekers of a monopoly of import licenses for oil or offshore drilling leases simply bribe the relevant officials with cash, there is in the first instance no additional (deadweight) loss. But if they hire lawyers, offices, restaurants, secretaries, and so forth, all employable elsewhere in the economy in some useful purposes, then additional loss is possible. It could be as high as the whole excess of marginal earnings of resources in the monopoly over their marginal cost, that is, as high as the whole amount of profit. In terms of the Figure 19.5 just used, the whole excess is \$0.10 at the margin, or the rectangle of profit over rent, namely, $(0.10)(0.70)$, or \$0.07 out of a total national income earned in the typical industry of \$0.70. The 10% loss is in this extreme case almost five times the loss calculated as a triangle of inefficiency.

Consider the following, which is a good example of the importance of viewing monopoly as behavior instead of structure.

Q: Edwin Land was the first to invent an instant camera.

He and his Polaroid Corporation continued to improve it, taking out over a thousand separate patents as they did so. Kodak wished to muscle in on this territory. To do so, however, it had to invent another sort of instant camera, different enough in its principles to avoid infringing the Polaroid patents. What was the social loss of Polaroid's monopoly? What single accessible figure would you want to have to measure it?

A: The social loss and the single figure you would want to have to measure it is the entire, enormous cost of developing a new type of instant camera. Little or nothing is gained socially by having two types. But the existence of a monopoly profit exploited by Polaroid gave the incentive to spend millions seeking it.

Summary

The defenses of monopoly are as varied as are the monopolists. The more subtle of the defenses are not easy to overcome: Monopoly might indeed, for example, be a spur to invention. On a theoretical level all that can be said in reply is that it might not be, too. What can be said with confidence is that simple monopoly reduces output and, more deeply, that it leaves unexploited opportunities for exchange between the monopoly and its victims. Even the simple monopoly is worse off than it might be. Note the word "simple" in "simple monopoly." It means "single price" or "unable to discriminate by way of charging different prices to different consumers." By the same Edgeworth box that shows that simple monopoly is inefficient, one can show that perfectly discriminatory monopoly is efficient; that is, it arrives at a point with no unexploited opportunities for mutually advantageous exchange. The notion that discrimination tends to efficiency has some practical use. For example, it rebuts the natural but erroneous opinion that because monopoly is bad more and stronger monopoly must be worse. A multipart tariff for electricity may therefore be better than a single-price electric company charging one, high, clumsy price to all.

⁴ Gordon Tullock, "The Welfare Costs of Tariffs, Monopolies, and Theft," *Western Economic Journal* 5 (June 1967): 224–232, and Richard A. Posner, "The Welfare Costs of Monopoly and Regulation," *Journal of Political Economy* 83 (August 1975): 807–828.

The main use of the notion, however, is theoretical. It alerts the economist once again, for example, to the great split between equity and efficiency in economic thinking. An inequitable society filled with fee takers may well be efficient, or an efficient society may be equitable, or any other pairing. And the logic of extreme monopoly reminds him that the inefficiency from less extreme monopoly comes from the transactions costs of making further deals between monopolist and victim. The amount of the inefficiencies (the triangle of social loss) is in a sense a measure of the transactions cost. More than they realize, perhaps, Marxists and other critics of capitalism are correct to view monopoly returns as a measure of the importance of friction in the system. The triangle turns out in fact to be small. But this does not end the argument between defenders and critics of the efficacy of markets. The triangle of social loss is a great underestimate of the whole social loss. The loss from the pursuit of monopoly profits, as distinct from the loss from exercising the monopoly power once attained, can be very large. Its size depends on how monopolies come to be monopolies, the subject of the next chapter.

EXERCISES FOR CHAPTER 19

1. Devise cheaper ways (namely, direct subsidies in the form of money) to achieve the supposedly desirable social end that is achieved by a monopoly.
 - a. Of local telephone service in order to subsidize rural telephones.
 - b. Of the right to import butter into France in order to help French farmers.
 - c. Of makers of military tanks in order to achieve reliability.
2. Show in Figure 19.1:
 - a. Why the intersection of the two *offer curves* is indeed the Competitive point. (*Hint:* Think back to what an offer curve is: It gives the amounts a *price taker* will take at various different prices.)
 - b. What indifference curves are tangent to each other at Competitive. Why this is therefore a point on the contract curve. That is, why it is therefore *efficient*.
 - c. Why OPEC is on a higher indifference curve and DOPE is on a lower one at Monopoly.
3. For a monopoly whose marginal cost is 80% of its price (the price is taken to be \$1 per ton) and whose output is 75% of the competitive (price equals marginal cost) output (the output is taken to be 1.0 ton), what is
 - a. The size of the shaded rectangle of efficiency loss in the style of Figure 19.5?
 - b. The national income earned in the industry?
 - c. The efficiency loss as a percentage of the income earned?
 - d. The net revenues (profits) of the monopolist?
 - e. The amount the monopolist would waste in lawyers and the like to protect the profits?
 - f. The share of these costs of monopoly as a percentage of the national income earned in the industry?
 - g. The ratio of monopoly protection costs to monopoly-inefficiency costs?

PROBLEMS FOR CHAPTER 19

- 1. The large profits of major league baseball teams are often defended on the grounds that much of the profits, extracted from the players, go to pay for their training in the *minor* leagues (the major league subsidizes the minor leagues). *True or false:* Even if this is true, it is irrelevant to deciding whether or not to break up the monopsony power of

the owners over the players (Compare the training of professional baseball players with that of professional tennis or golf or economics players).

2. Does a monopoly of razor blades have an incentive to make the life of the razor blades less than it could make it? Consider what the consumer will be willing to pay for a blade that gives 10 shaves as against one that gives 20 shaves.

3. Macmillan brings out new editions of books such as *The Applied Theory of Price* very frequently, making earlier versions worthless in the secondhand market. *True or false:* This is obviously a case of planned obsolescence (that is, making things wear out so that the maker can make more money).

4. In movie theaters only large candy bars are for sale. *True or false:* The practice is an example of setting both the price and the quantity of the victim of a discriminatory monopoly. (*Hint:* Remember that discriminatory monopoly sells a larger output than single-price monopoly.)

- **5.** The medical profession is alleged to be a cartel. Supposing it to be a perfect one, illustrate the social loss from the cartel in the market for medical services, identifying separately the opportunity cost of the lost output and the willingness to pay for it. If a society found the loss disagreeable, what would be the obvious social policy for it to pursue? How much would doctors lose if this social policy were in fact pursued? [Use supply—*that is, marginal cost—and demand curves.*]

Having said this, suppose that the policy were impossible (because, say, doctors are too powerful politically to stand still for the loss of cartel profits or doctors are too highly esteemed for society to suffer their impoverishment without protest). Show that a subsidy by the state to the firms (usually individual doctors) producing medical services can, even in the presence of a cartel, yield the socially optimal output of medical services. Distinguish real opportunity cost from the marginal cost facing the cartel in this demonstration.

The point, as usual, is one of “second best.” It is that when one is forced to accept conditions yielding a bad allocation of resources, adding another condition that would be bad in itself (subsidizing a firm, thus distorting its pattern of incentives) will sometimes make the situation better, not worse.

6. The text asserts that “if seekers of a monopoly . . . simply bribe the relevant officials with cash there is in the first instance no additional loss” from monopoly over the usual triangle of surplus lost.

- a. Explain what is meant by the assertion.
- b. Suppose that there is competition to become the “relevant officials” by studying law or buying advertising. In such a case is there excess loss?
- c. Suppose now that potential officials can compete for high office only by offering themselves at lower pay. What happens to the loss?

7. Suppose that an American manufacturer of military tanks has a monopoly in the United States and is initially not permitted to export his product to other countries. Now suppose that he is permitted to export it (imports of foreign tanks are still not permitted) and suppose that he makes a small portion of the outside world’s supply of tanks. (Use supply and demand.)

- a. Compare his profits before and after he is permitted to export.
- b. Compare American welfare before and after he is permitted to export.
- c. What is the effect on profits and welfare of a technological change specific to the American manufacturer, both before and after exports are permitted?

- **8.** The right to use a patented invention can be bought from the inventor in one of two ways: either with a fee paid each time the invention is used or with a lump sum. Thus a new type of plastic could earn the inventor either \$2.56 per ton produced (by you exclusively, perhaps, or by anyone else who contracts to work the patent) or a flat payment of \$100,000

(again, with or without exclusivity), with no charge per ton. Both ways of licensing patents are in fact in use.

a. Explain why the fee per unit is worse than the lump sum on at least two grounds: the expense of enforcing the contract and the social waste of underusing knowledge.

b. Explain why an inventor might nonetheless wish to license by fee per unit if she did not know exactly how valuable the invention would be to its users. Ignore considerations of risk and insurance, but consider ignorance itself. Is the extra cost incurred by the fee per unit, according to (a), really inefficiency?

c. Inventions are commonly distinguished as product or process, for example, electronic calculators (a product sold to ultimate consumers) or a can-making machine (a process). And each of these can be novel or routine: A better vacuum tube is a routine process innovation; a transistor is a novel process innovation. These distinctions are not especially penetrating, memorable, or useful. But in accepting them for the moment, show that by the argument of (b) one would expect that routine process inventions would be licensed by lump sum and novel product inventions would be licensed by fee per unit. *Factual note:* One would be right, according to recent researches by S. N. S. Cheung of the University of Hong Kong.

True or False

9. A good way in which to improve the welfare of society would be to subject monopolies to regulations limiting their rate of return to capital to some low figure.

10. Monopoly in one industry will create excess supply in another.

- o 11. If one knew the elasticity of demand for medical services, the present price of those services, their present real opportunity cost at the margin, and the elasticity of this opportunity cost, then one could estimate the resources wasted by the presence of a medical monopoly in the United States.

20

Monopolistic Competition and the Economics of Location

20.1 Monopolistic Competition as Competition Among Local Monopolies

What to Read For What are the desirable features of a competitive equilibrium? How does *monopolistic competition* violate them? How do you construct the demand curve facing a local store? What do you have to assume about how neighboring stores react to various different prices set at the store? How does the demand curve shift as more firms enter? Why do they enter? What is the long-run equilibrium? What is product differentiation? What are the three objections to the theory of monopolistic competition? Is product differentiation always bad? Is the assumption about how neighboring stores react a reasonable one? Has the theory of monopolistic competition been rich in applications?

Who's in Charge Here? The analysis of one seller assumes that there is one mind in charge. The monopoly may have literally one mind, as does the monopolistic seller of a unique painting, or it may have many as does the American Medical Association. But it behaves as though it had one. However it is organized, the industry maximizes its profits jointly. Each participant surrenders to the common object of making as much as possible using any selfish advantage to be gotten by competition.

The question is how an industry behaves when there is *more* than one mind in charge but fewer than the many minds of perfect competition. The answer to the question is the subject of this and the next chapter. It has turned out to be a long and frustrating search in the dark by economists. The answer is that there is no answer, only a number of partial answers, each more or less useful, none very satisfactory. Life is hard.

The Model Without Transport Costs

The analysis works inward from either extreme: perfect competition (a large number of sellers) or perfect monopoly (one seller). This chapter takes up the large number.¹ Suppose that there are 150 sellers of groceries arranged every two miles along Route 66 through Oklahoma. The situation is supposed to reflect a pretty common one, of a large number of competitors in an industry each with *some* monopoly power—in this case, *local* monopoly power. Each has the usual array of cost curves, assumed to be the same for each. Suppose that the potential customers are scattered evenly along the highway and that each demands one bag of groceries a week regardless of the price that must be paid. The customer's demand curve, in other words, is perfectly inelastic. The idea here is to keep things simple.

Inelastic though her demand is, however, the customer still wishes to pay the lowest price. If it costs her and other customers *nothing* to drive to each of the 150 stores, then each store would be forced to compete in the entire Oklahoma market with every other store. The results would be the usual ones of competition: The price of a bag of groceries would almost exactly equal the marginal cost and the minimum long-run average cost of the bag at the stores. That is, the two features of competitive equilibrium would show themselves: (1) marginal valuation would be equal to marginal opportunity cost; and (2) no one would be earning supernormal profits. Read that sentence again slowly. Each firm would be operating at the minimum point of its average cost curve, supplying 1/150 (or 0.0067) of the socially optimal output of groceries.

The Model with Transport Costs

Optimality breaks down, however, if there are (as there are) costs of driving. The cost of driving to a store can and should be viewed as part of the price of the bag of groceries. Someone next to the store pays less in transport costs if not in the price at the store than someone two miles away. In other words, when there are transport costs, there are *local* monopolies. The customers near David Haddock's store on Harvard Avenue in Tulsa do not have the option of visiting at the same low cost all the other stores along Route 66, such as those in Oklahoma City or Joplin (Missouri), each 100 miles away. Assuming that Haddock cannot distinguish between customers who have come merely from Yale Avenue a mile away and those who came all the way from Reed Park five miles down to the highway, he cannot charge different prices to different customers in proportion to their willingness to pay. (If he could, he would charge Esther Schultz, a customer from Yale Avenue, more. Such a close customer would be willing to pay over to Haddock as much as *all* her saving in transport cost in order to continue buying at such a convenient location for her.) Price discrimination, in other words, is assumed to be impossible, leaving the store-keeper as a simple, single-price monopolist.

The Demand Curve Facing a Local Monopolist

But a monopolist he is. We want to know what Haddock's demand curve is to figure out how he behaves. The demand curve he faces is downward sloping. *Its slope depends on how his neighboring monopolists react to his activities.*

¹ The exposition here follows that of Gordon Tullock in *Toward a Mathematics of Politics* (Ann Arbor: University of Michigan Press, 1967), Chapter V, "Single Peaks and Monopolistic Competition," especially in the matter of identifying transport models with monopolistic competition. Tullock's book is a good one, and readable.

The simplest assumption is that in the short run they do not react at all. Remember this step in the argument: The assumption about what rival monopolies do is the key.

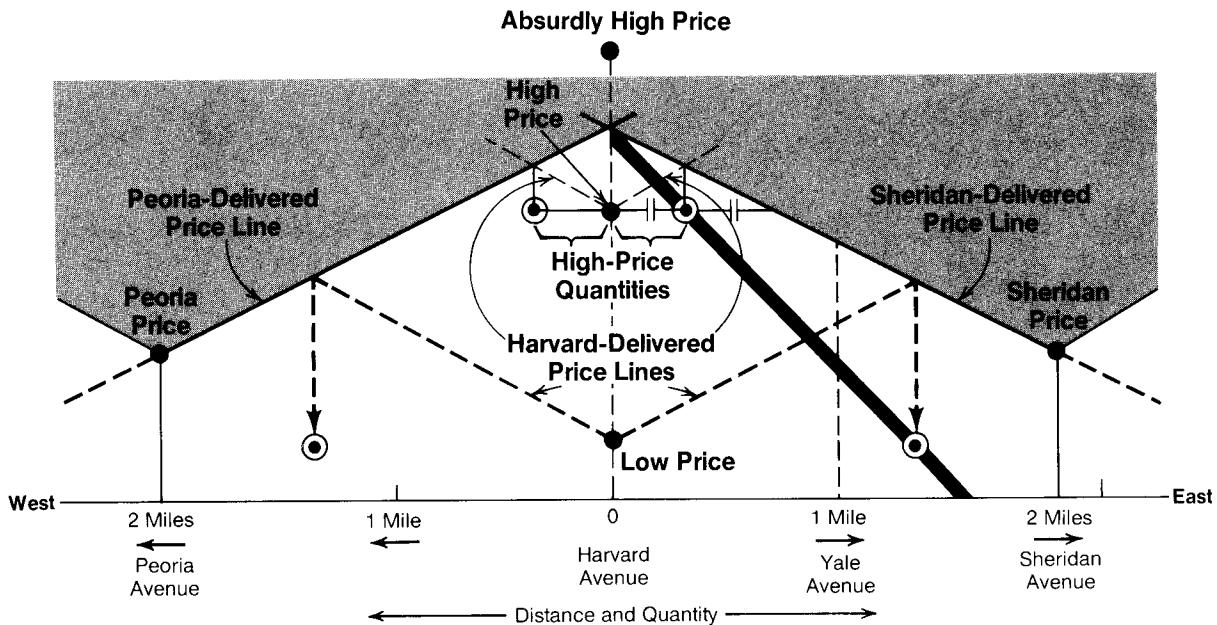
Even under such a simple assumption, the derivation of Haddock's demand curve is somewhat involved. It goes as follows: Imagine Haddock sitting in his store at the corner of Harvard Avenue and Route 66, with a store competing with him two miles east along Route 66 (at Sheridan Road) and another two miles west along Route 66 (at Peoria Avenue). Get this image literally in your mind. Obviously no one will bypass the Sheridan store, say, coming from far east to Haddock's store. Haddock's potential customers therefore are *only* the ones between Sheridan and Peoria Avenues, two miles east and west. A certain price at his store will attract a certain number of these customers in both directions. How many? The answer depends on what his two immediate neighbors are doing. What they are doing can be summarized in their price at their stores, represented as the points Peoria Price and Sheridan Price in Figure 20.1. Look at it. Don't faint. It's not impossible to understand if you take it step by step.

The prices are taken to be the same at both Peoria Avenue and Sheridan Avenue merely to shortcut some reasoning, reasoning that concludes that in the final equilibrium the prices *will* be the same. The shaded "tent" over the Harvard Avenue location is a result of adding transport costs on to the Peoria and Sheridan prices. It is a tent from Haddock's point of view, but from the Peoria or Sheridan point of view it is a shaded *transport cone* balanced on each price. The sides are straight because the customers are supposed to be scattered evenly along the road, resulting in the same addition to sales for each additional mile. It gives the price to customers at various distances from the two stores, including transport costs. The farther a customer is from, say, the Sheridan store, the more the customer must spend on transport to buy from it. Obviously a customer at, say, Yale Avenue (look at the right side of the diagram) would buy from the Sheridan store as long as it was the low-cost supplier. In other words, Haddock's price at Harvard Avenue plus transport cost cannot be outside the tent formed by his competitors: If it is the other stores will undersell him, and he will get no sales.

Now consider the problem of constructing the demand curves Haddock faces. Remember: That's what all this is about—constructing Haddock's demand curve, which is typical of one of these local monopolists. A demand curve is a locus of combinations of prices charged at the store and of quantities demanded by consumers. If Haddock charges *Absurdly High Price* in Figure 20.1, he will get no customers. Even customers across the street from his Harvard Avenue store will prefer to go to either the Peoria or the Sheridan store. But if he charges only *High Price*, he gets some customers. In fact, it turns out that he gets all the customers out to where the dashed lines have the same slopes as the tent and intersect the tent. It is assumed that the same cost of transport per mile is incurred by Haddock's customers as by all other customers. The margin of Haddock's market is determined by the tent, because beyond the tent Haddock's price including the transport cost to the customer exceeds that of his competitor, and therefore the customers farther out go to his competitor instead. Since customers are distributed evenly along the highway in proportion to distance, the distance can be identified with quantities. At *High Price*, then, Haddock sells the *High-Price Quantities*. They are marked as horizontal lines, to either side

Figure 20.1
The Demand Curve Bisects the Distance Out to the Tent of Transport Costs

The quantity sold at Harvard Avenue depends on (1) the price charged at Harvard (such as Low Price); (2) the Harvard-Delivered Price Line, which gives the total cost, including transport costs, to customers of Harvard; (3) the Sheridan- and Peoria-Delivered Price Lines, which give price to customers of neighboring stores; (4) the density of customers along the road; and (5) the fact that customers choose the cheaper delivered price. The store at Harvard Avenue can sell to customers out to the point where the dashed Harvard-Delivered Price Line cuts, say, the Sheridan-Delivered Price Line. The resulting demand curve is the heavy line equidistant from the Harvard axis and the Sheridan- (or Peoria-) Delivered Price Line, on the suppositions that transport costs (the slopes of the lines) are equal and that the Sheridan price is given and fixed at the level it would attain in equilibrium.



of his store. We're almost there: We have found what quantities demanded come from setting a certain price. Notice the equality of the horizontal lines (indicated by slashes). That they are equal can be seen by noting that their triangles are identical. The consequence is that the quantity purchased bisects the distance out to the tent.

The argument is identical at other prices, such as Low Price. The demand curve facing Haddock in either direction is constructed by linking up the circled points. *The points themselves are halfway out to the tent sides.* This is the key result. The emphasized line, for example, is the demand curve in the Sheridan direction.

Pause for a moment to note that the construction of the demand curve facing Haddock is in accord with common sense. If transport costs fell, for instance, the tent would collapse, too, until at last the tent and the demand curve con-

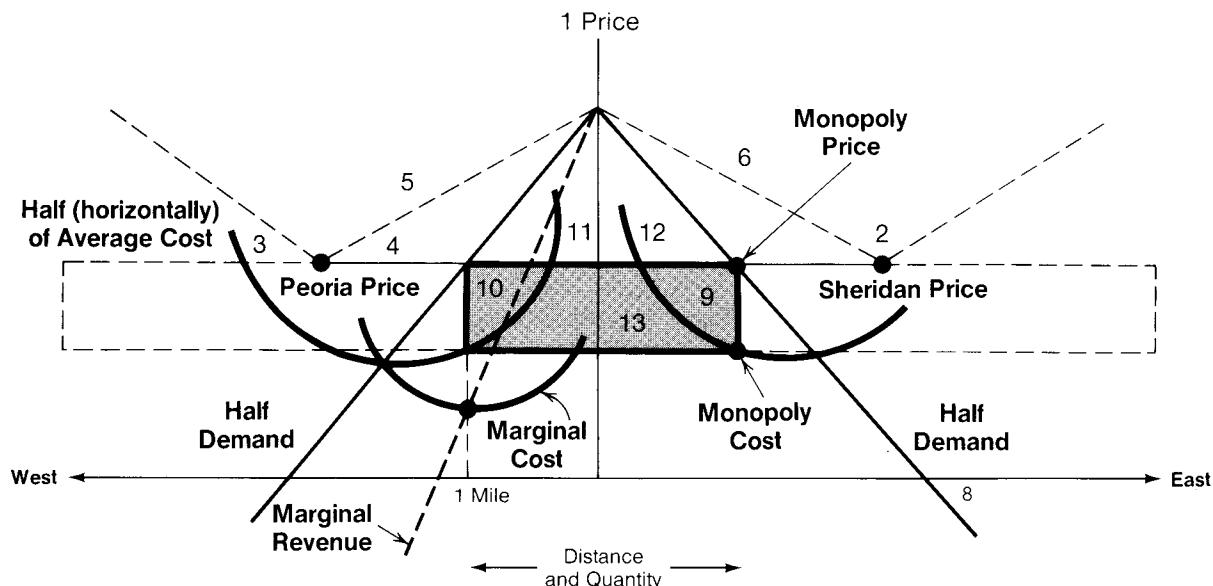
verged as a flat line at the price abroad. Local monopoly would decay, as it did to some degree with the spread of the automobile early in this century. On the other hand, if transport costs were very high, the tent would slope upward more steeply out of the pegs at Peoria Price and Sheridan Price, and the local monopolist would face a more inelastic demand curve. As transport costs rise, the top of the tent gets higher and the demand curve facing the local monopolist gets steeper (that is, more inelastic). Services with high transport costs, such as elementary school education, have local monopolies. In the limit the local monopolist would be left entirely alone to exploit his local victims, transport costs forming a wall around the market.

The Maximum of Profit for the Local Monopolist

Having found the demand curve, the completion of the theory of the local monopolist requires comparing his costs to the demand. This, too, is pretty tricky, but not hard if you take it step by step. Haddock's cost curve for supplying bags of groceries is supposed to be of the usual U shape. It can be split into

Figure 20.2
Monopolistically Competitive Equilibrium When All Firms Are Identical and Entry Is Barred

The numbers in the diagram are the sequence in which you should draw the lines. You'll follow it better if you do. In equilibrium with N identical and evenly spaced stores, each store chooses quantity such that marginal cost = marginal revenue, each store charges the same price, and by these policies the stores evenly divide the territory between them. Notice that the equilibrium does *not* occur at the minimum point of the average cost curve and that the price does not equal marginal cost.



two halves to represent the simultaneous servicing of outputs in two directions, east and west. Haddock chooses his monopoly price as the one that maximizes profit, that is, the vertical distance between Monopoly Price and average cost multiplied by the horizontal distance of quantity (see the shaded area of profit in Figure 20.2). As usual, marginal cost is equated with marginal revenue (as is shown in the left-hand side).

As the diagram is drawn, Haddock's choice of price is the same as everyone else's (namely, the Peoria price and the Sheridan price). This is true in equilibrium, which is why it is taken as the example here. It reflects again the assumption of symmetry underlying the analysis, namely, that if all grocery stores are identical, then each must in the end act in identical ways. The same diagram can serve therefore to represent any of the 150 stores from Joplin, Missouri, to Erick, Oklahoma. Since each firm is motivated identically, each will be motivated to charge the same price in equilibrium and to get the same one mile of customers on either side. The equilibrium, in short, is an allocation of the entire 300 miles of Route 66 in Oklahoma into local monopolies covering two miles of customers each.

Entry Leads to the Tangency Solution

Q: Is the equilibrium portrayed a long-run equilibrium?

A: No, it is not. *Supernormal profits are still being earned.* In pursuit of the profits, more grocery stores will try to enter the industry, squeezing in beside the other stores. It would be unpleasant for Haddock if another person named Ulen, for example, set up a store on Yale Avenue, at the very edge of Haddock's market. Haddock could make it less unpleasant by relocating in the long run away from Ulen, as Ulen's competitor on the other side could by relocating toward the east. The coming of Ulen would spread ripples of movement east and west, like parishioners shuffling over on a church bench as one more joins the row. The result

in church is more people to a bench, that is, smaller spaces between each person. So, too, on Route 66. The neighboring stores are closer or more similar in some other dimension. The crucial point is that *their closeness shifts inward the demand curve* at any given price abroad facing Haddock, as from Old Demand to New Demand in Figure 20.3. The diagram exhibits half of the two-sided figure used earlier.

The innermost curve, Equilibrium Demand, is just that, the only demand curve that leaves *no supernormal profits* to attract further entry. Its (negative) slope is, like the others, twice the cost of transportation (since the demand curves always bisect the tent side and pole). The grocery industry on Route 66 settles to rest.

The outcome is striking. If transport costs were zero, the industry would arrive at the Competitive Point on the bottom of the U curve of Average Cost, with each of a smaller number of grocery stores producing at a larger output and at the lowest possible cost (full capacity). But in the presence of transport costs, the industry ends at the Tangency Solution, with a large number of Ma and Pa grocery stores producing at less than full capacity and at a high price. A theory that joins free entry to a monopoly with neighbors implies an equilibrium of profitless and inefficient little enterprises, like dry cleaners, local restaurants, corner drugstores, or gas stations. They are monopolies, to be sure, but don't get much out of it for themselves or for society.

Price

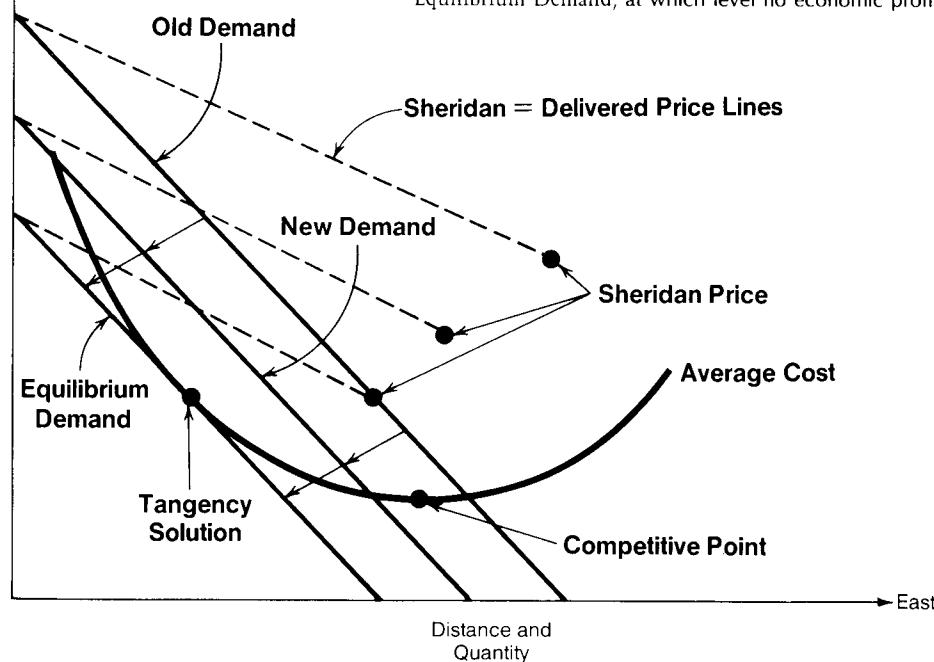


Figure 20.3
Monopolistically Competitive Equilibrium When Entry Is Free

The economic profits earned at Old Demand induce entry. The more stores enter, the closer stores are to their neighbors, and the lower the delivered price line and demand curve each store faces. Entry continues until the demand curve facing each store falls to Equilibrium Demand, at which level no economic profits are earned.

Monopolistic Competition as Local Monopoly

The theory is called *monopolistic competition*. Invented by Edward Chamberlin in the 1920s, it has had a remarkable career of elaboration and criticism, if not much of actual use.² "Its thesis is that both monopolistic and competitive forces combine in the determination of most prices, and therefore that a hybrid theory affords a more illuminating approach to the study of the price system than does a theory of perfected competition supplemented by a theory of monopoly." The theory is identical to the problem of the grocery stores on the road, except that distance in physical space is also interpretable as "distance" in "product space." Various breakfast cereals, for example, are more or less close to

² Edward Chamberlin, *The Theory of Monopolistic Competition: A Re-orientation of the Theory of Value* (Cambridge, Mass.: Harvard University Press, 1933), especially Chapter V, submitted as a Ph.D. thesis at Harvard in 1927. The quotation is from the Preface to the first edition. Readers of Chamberlin will wonder where his famous *dd* and *DD* curves appear in the present exposition. The answer is that the demand curve here is Chamberlin's *dd* curve, that is, the curve drawn on the assumption that neighboring monopolists keep their prices constant no matter what Haddock does. The *dd* curve is the heart of Chamberlin's argument. The curve drawn on the other assumption, that neighboring monopolists match Haddock's prices, would be Chamberlin's *DD* curve. For given locations, it is simply a vertical line halfway between Haddock's place and the neighbor's place. The market always divides in half if Haddock's and the neighbor's price are the same. It is vertical because of the assumption here that each customer's demand curve is perfectly inelastic.

each other, Kellogg's Corn Flakes and Post Toasties being very close, perhaps indistinguishable in chemical content, whereas Corn Flakes and Ralson-Purina Corn Chex are farther apart and Corn Flakes and Nabisco Shredded Wheat still farther. A standard Ford is closer to a standard Chevrolet than it is to a Cadillac or a Honda.

The type of product is variable. Firms can choose where to locate along the road or where to locate in product space. The emphasis on the product as well as on the price as a variable subject to economic choice is one of the lasting contributions of the theory of monopolistic competition.

It emphasized, too, that firms can create the image of differences in products as well as the reality. *Product differentiation* has come to be in economics another word for advertising. The firm in this view can create artificial distance between its products and its competitors, by persuading people that only at Wendy's do they have much beef and only at Schlitz do they make a beer that witty folks drink. What is distasteful about the result of monopolistic competition is the combination of the low output characteristic of monopoly with the proliferation of tiny, pointlessly differentiated firms characteristic of competition, all supported by advertising at the customer's expense.

***The First Objection:
Product Variety Is
Not Bad***

The spatial model suggests, however, that one must not be too quick to damn the differentiation of product. The differentiation along the road, after all, is real distance, expensive to get through. If the differences are not completely phony, then it is *not* desirable for there to be fewer grocery stores, even though each would produce and price at a lower point on its average cost curve. The consumers prefer a store close to them with a high price at the store to a store far away with a low price. In other words, the "lowest point of the average cost curve" is misdefined in the argument because it ignores the unpleasant cost of travel (that is, the pleasures of nearness). As Gordon Tullock has put the matter,

Prices will never get down to the level of perfect competition. The customers can hardly complain about this since it arises out of an effort to please them. Their unwillingness to go long distances to take advantage of small differences in price gives the monopoly power. . . . [There is according to the traditional way of putting it overinvestment in stores. They] are not operating on the low point of their average cost curves, and . . . the prices are higher than would prevail if perfect competition reigned. [Such perfection], however, involves ignoring the preferences of the consumer, and hence is a most unusual type of optimum.³

The argument is identical to the one in the last chapter about natural monopoly: If the monopoly is natural, it may be good for you. If you yourself value a different product, product differentiation also may be good for you. The welfare economics of the theory of monopolistic competition is dubious.

***The Second
Objection: The
Competitors Would
Allow for Their
Interactions***

The behavior posited at the heart of the theory is also dubious. The heart of the theory is the combination of downward-sloping demand (characteristic of monopoly) with large numbers of firms *each of which can ignore its interaction with any one other firm* (characteristic of competition). If firms do *not* ignore their interactions, the taking of the neighboring prices as fixed for purposes of

³ Tullock, *Mathematics of Politics*, pp. 68–69.

analysis will be false. But it is plain that Haddock and his near neighbor Ulen *would in fact allow for, not ignore, their interactions*. A decision by Haddock to alter his price would have powerful effects on Ulen, changing his behavior and in turn changing Haddock's, which would in turn change Ulen's, and so on into a hall of mirrors. The hall will be studied in the next chapter. At present it is enough to doubt Chamberlin's notion "that any adjustment of price or of 'product' by a single producer spreads its influence over so many of his competitors that the impact felt by any one is negligible and does not lead him to any readjustment of his own situation."⁴

Along a literal road the notion is certainly false: Haddock has necessarily two neighbors so close that their mutual effects are great. One might try to rescue Chamberlin's argument by generalizing the road to a space of many dimensions. But such a generalization would seem merely to multiply the number of close neighbors, multiplying the problem with them. Two analytical ploys are possible. If one multiplies the close neighbors enough, the price facing the monopoly in question becomes fixed by competition, *but the analysis reduces to perfect competition*. If one pushes the neighbors farther away, to make the "impact felt by any one" truly "negligible," the group equilibrium is no longer important and *the analysis reduces to perfect monopoly*. The middle ground that Chamberlin attempted to occupy is a canyon.

**The Third Objection:
The Theory Has Not
Borne Fruit**

The theory of monopolistic competition, then, has difficulties in the behavior it assumes as well as in the inferences about goodness or badness it draws. You will notice, furthermore, that the preceding pages have had no problems to solve. This is the third and most serious difficulty with the theory of monopolistic competition. Unlike perfect competition or monopoly, or even the oligopoly models to be discussed shortly, the theory has not been very fruitful. It can be argued, and has been most vigorously by many economists, that monopolistic competition is a better description of many industries than is either monopoly or competition by itself. But along with the superiority of descriptive accuracy comes no striking proposition about how such industries behave that would make it worthwhile to develop a third thing between monopoly and competition. One might as well develop a model of "male-dominated competition," on the (true) argument that, unfortunately, males hold a disproportionate number of the positions of authority in the economy. The descriptive accuracy of the model would be greater than a model that did not mention sex, but only in the single respect, since it seems unlikely (although on reflection perhaps not impossible) that the economy would behave much differently with more females in charge. The theory of male-dominated competition would probably not be fruitful in new and unexpected applications of economics. Neither has been the theory of monopolistic competition.⁵

Summary

The best way of representing the theory of monopolistic competition is to take literally its talk of "near" and "far" products. If the assumption about how each competitor believes the neighbors will react (namely, not at all) is true, then the model of grocery stores along a road yields a monopolistically competi-

⁴ Chamberlin, *Monopolistic Competition*, p. 83.

⁵ See R. Robinson, *Columbia Essays on Great Economists: Edward H. Chamberlin* (New York: Columbia University Press, 1971), for a more favorable view of the achievement.

tive equilibrium. Too many grocery stores crowd onto the road, each protected in its profitless enjoyment of a local monopoly by the cost of transportation to another store and each producing too little at a high cost. The idea of the distance between different stores along a road can then be extended to "distance" between different products in the minds of consumers. The results are the same: firms that are too little, too expensive.

The analogy with distance, however, suggests that product differentiation is not always artificial and pointless. After all, a store on 125th Street is farther from downtown than is a store on 42nd Street. Some people do prefer a big automobile with four-wheel drive and a vinyl roof to a small one with an airbag and a sun roof. Furthermore, the assumption that a competitor believes that the neighbors will not react to his decisions is unreasonable. If the neighbors are close enough to matter, they are close enough to react, and the competitor in question will be intelligent enough to perceive that they will react. Finally, and most important, the theory has borne little fruit. It is a better description of the world than is perfect competition or perfect monopoly, but it is not a better instrument for scientific study.

EXERCISES FOR SECTION 20.1

1. Construct a third point along the demand curve of Figure 20.1. That is, start at a middling price, and ask how much Haddock would sell in the Sheridan direction. Where would his marginal customer be?
2. Which of the following industries do you suppose is monopolistically competitive? Of those that are, what justification in consumer preferences is there for product differentiation?
 - a. Stationery stores.
 - b. Corn farms in Nebraska.
 - c. Automobile manufacturers. (*Hint:* Would they ignore their interaction?)
 - d. The hundreds of dressmakers.
 - e. The hundreds of colleges.

20.2 The Theory of Location

What to Read For

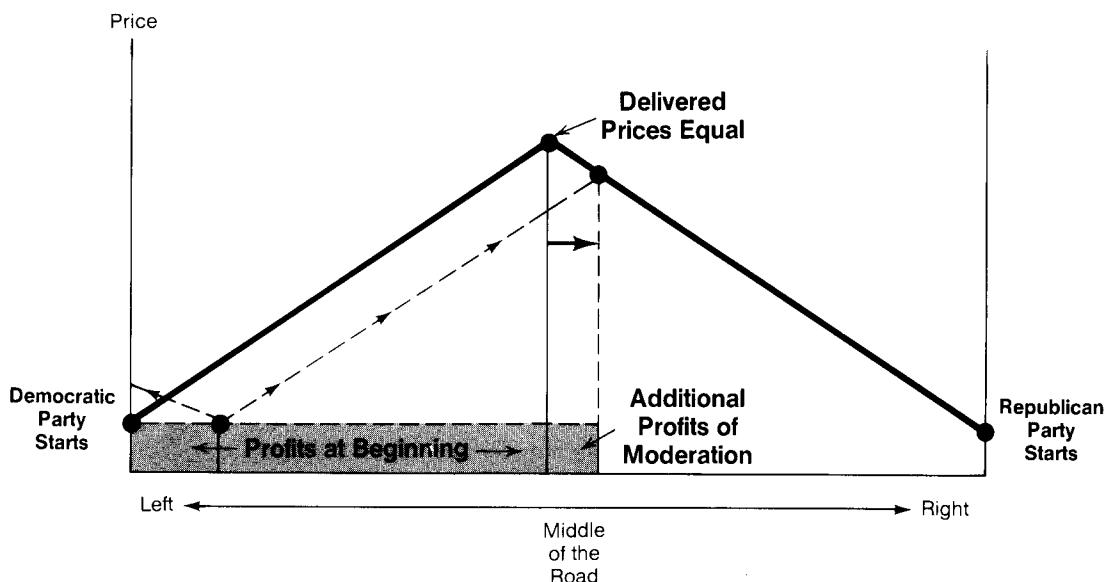
How can the theory of location be applied to politics? Why do American political parties look so similar? How does transport cost to a central city determine land rent away from the city? What is a bid rent curve?

**The Hotelling Model
of Two-Party
Competition for
Votes**

If the theory of monopolistic competition itself has not been fruitful, many of the devices developed to explore it have been. Above all, the idea of competition over space has been fruitful. A good example comes from a famous article by Harold Hotelling (1929), which was published independently of and before Chamberlin's book but which discusses the same "neglected fact that a market is commonly subdivided into regions within each of which one seller is in a quasi-monopolistic position" in which "a tiny increase in price by one seller

Figure 20.4
Why Parties Move to the Middle in a Two-Party System

Each party sells to that part of the market that gets a lower delivered price from it than from its competitor. If parties start at the extremes, either party can increase its sales by moving toward the other.



will send only a few customers to the other.⁶ He applied it immediately to a very important—though noneconomic issue—namely, why the two big American political parties are so similar.

Q: Suppose that two soap stores called the Republican Party and the Democratic Party start at the right and left ends of a road called Political Spectrum Boulevard. They sell soap at the same, fixed price at the two stores. Because transport costs per mile are the same in all directions, the price consumers pay rises from the price at the stores in proportion to distance from the stores. The consumers, who are distributed evenly along the road, have perfectly inelastic demands for soap (in the spirit of the earlier grocery store problem). The two stores are assumed to have no costs, so that their profits are simply the number of miles of road they service multiplied by the price paid at the store: All revenues are clear profit.

1. Draw a diagram showing how the market for soap

divides between the two stores when they are located at opposite extremes.

2. Suppose that the Democratic Party assumes that the Republican Party will not change its location if the Democratic Party moves a little. Show the incentive for the Democratic Party to move to the right.

3. Suppose that the Republican Party makes the same assumption. Where will it move? Would it ever be desirable for it to stay sitting on the right extreme? Where will the sequence of moves end?

4. Can you imagine what feature of the American party system is explained by the model?

A: 1. The market divides where the delivered price of soap is equal from either store, namely, at the Middle

⁶ Harold Hotelling, "Stability in Competition," *Economic Journal* 39 (1929): 41–57, reprinted in G. J. Stigler and K. E. Boulding, eds., *Readings in Price Theory* (Homewood, Ill.: Irwin, 1952), pp. 468–470.

of the Road in Figure 20.4. Consumers to the right go to the Republican Party, those to the left to the Democratic Party.

2. The incentive for the Democratic Party to move to the right is that its sales will rise because the point of market division determined by the slope of transport costs will move right (as indicated by the dashed line). The Democratic Party will capture the additional market share. It will keep the profits to the left of its new location *because those customers have nowhere to go*.

3. But the Republican Party has exactly the same opportunity, and it will take it. Its incentive to do so is increased by the extermination that will come inevitably if it stays put on the right end of the road: The Democratic Party will march all the way to the right,

seizing all the consumers. But since both parties do it, the moves cancel each other out, neither party gaining in the end. The parties end up sitting a bit to either side of Middle of the Road, having lost in the process nearly all their identity as left- and right-wing soap sellers.

4. As Hotelling put it,

The competition for votes between the Republican and Democratic parties does not lead to a clear drawing of issues, an adoption of two strongly contrasted positions between which the voter may choose. Instead, each party strives to make its platform as much like the other's as possible. . . . Each candidate "pussyfoots," replies ambiguously to questions, refuses to take a definite stand in any controversy for fear of losing votes.⁷

Improvements on the Simple Model Are Persuasive

T or F: With three political parties located in the beginning at the two extremes and the exact middle and constrained to stay in the same relative position, the two left- and right-wing parties will exterminate the middle.

This happened in fact to the Liberal Party in Britain between the election of December 1910 and the election of October 1924. The Conservative Party on the right and the new Labour Party on the left crushed the Liberals by moving in on them. The final blow was the adoption in June 1924 of the policy of free trade by the leader of the Conservatives, the last of many examples of the other two parties' stealing Liberal positions and becoming thereby more similar. The Liberal Party lives on to the present, but as a mere shadow. For the same reason, middle-of-the-road third parties in the United States (if not, to be sure, in France and many other countries) have been short lived.

The assumption in the simple version of the Hotelling model that the only competition is by location, not by price, is also crude, but amendments to it also yield fruit.

Q: Imagine two stores located next to each other in Hotelling fashion at the middle of the road. They charge the *same* price at the store, each selling only in one direction (the left store to the left only, for instance). What happens if one of the stores lowers its at-store price below that of the other, and the other sticks to the high price? However stable it may be as a *location* equilibrium, is the Hotelling solution stable as a *price* equilibrium? View location as "the party's position on

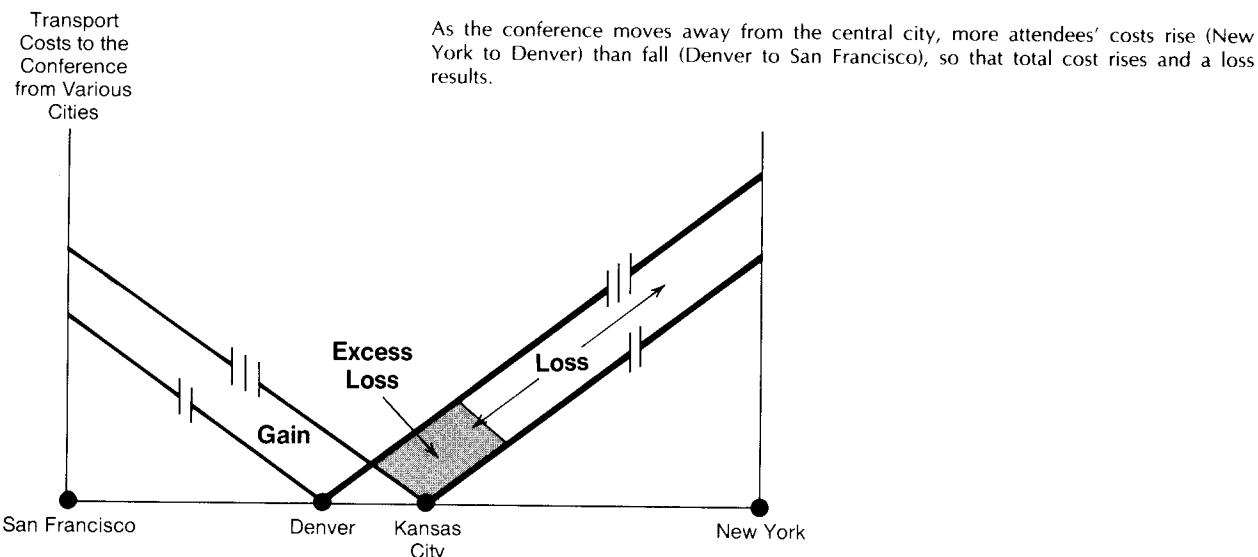
A: The third, middle party has nowhere to go. Trapped between the other two moving in to poach votes, it ends up with no votes. Therefore, true: The two parties move in on the third.

an ideological spectrum" and price as "the attractiveness of the major candidate." Now interpret the presidential election of 1952, in which the Republican Party triumphed by adopting Democratic positions on ideology (accepting the results of the New Deal) and getting a war hero (Eisenhower) to run as its candidate.

A: As a political scientist examining such models has put it, "if the difference in nearness for any given cus-

⁷ Hotelling, in Stigler and Boulding, eds., *Readings in Price Theory*, p. 482.

Figure 20.5
All Conferences Should Happen in the Midwest



tomer is negligible [as it is if the shops have placed themselves next door to each other], one of the competing shops will be able to get all the customers (and thus ruin the other) provided it can maintain its price just a shade below those of its rival. . . . The closer the parties become ideologically [that is, in 'location'], the more people's votes are liable to turn on something other than ideology [that is, on a 'price' such as the attractiveness of the candidate]."⁸ The parties have no

incentive to move away from a Hotelling equilibrium in location, but they do have an incentive to move away from it in price. Having set up shop next door to Democrats, the Republicans had only to shade their price (nominating General Eisenhower instead of the ideologically purer but less personally attractive Senator Taft) to attract voters from the Democratic side of the spectrum.

The Best Location The locational models apply, of course, to problems of literal as well as figurative location.

Q: You are organizing a conference with 50 invited participants drawn at intervals of 60 miles between each from a line across the United States 3000 miles long. You can locate the conference anywhere along the line. You pay the transport costs, which are constant and identical per mile traveled, and wish to minimize them. Where should you locate the conference? (*Hint:* Follow your instinct).

A: Your instinct is correct in telling you that the midpoint is cheapest. The most elegant demonstration is to start with a location at the midpoint and show that movements away from it cause total transport cost to rise. The total transport cost can be approximated by the area under the transport cone rising out of the location chosen. At the midpoint, for example, it is the area under the sloped lines in Figure 20.5. A move from

⁸ Brian Barry, *Sociologists, Economists, and Democracy* (Chicago: University of Chicago Press, 1978), p. 105. The page following contains the analysis of the 1952 election. Barry's Chapter V, "The Economic Theory of Democracy," is a penetrating summary and critique of such theories as Hotelling's and especially of the brilliant extension of his theories by Anthony Downs, *An Economic Theory of Democracy* (New York: Harper, 1957).

the midpoint at Kansas City to, say, Denver farther west will result in a larger area of loss on transport costs than gain. The shaded area is the excess loss. One can alter the assumptions slightly to favor one or another midwestern city. Kansas City is near the physical

midpoint. St. Louis, however, is near the center of population. And Chicago is close to St. Louis and has more airline connections. In any case the minimization of transport costs leads to a midwest location.

The Isolated City

The earliest and most fruitful location argument in economics is that of David Ricardo and, with more emphasis and detail, Johann H. von Thünen, in his book of 1826, *The Isolated State in Relation to the Economics of Land and of Nations*. It simply reverses the sense of the model of grocery stores. Instead of a store essentially sending goods out to customers along the line, farmers along the line send their goods into the central (isolated) city.

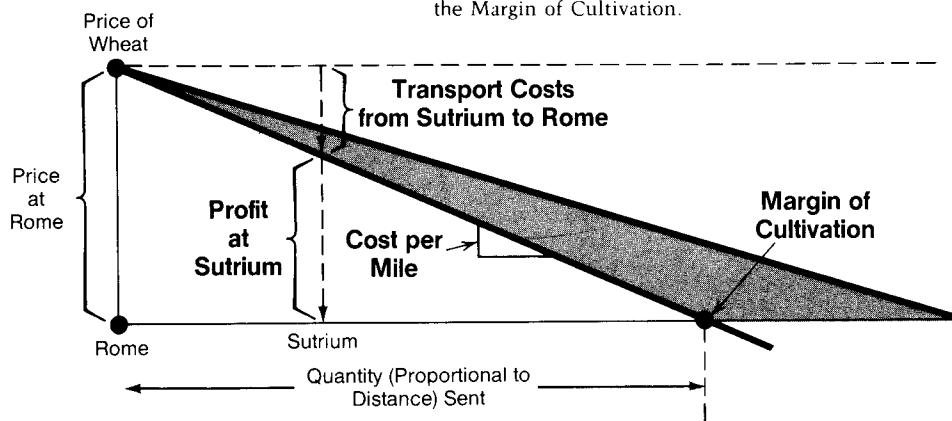
Q: Suppose that the price that citizens of the city of Rome are willing to pay for wheat is fixed; that is, their demand is elastic at some given price. Rome is the only consumer of wheat: In particular, farmers do not consume wheat. Suppose, too, that all roads lead to Rome and that shipping wheat along them costs some fixed amount per mile. Suppose that there is a limited amount of land available at each distance from Rome. Suppose, finally, as an inessential simplification, that wheat costs nothing to produce. In other words, the only difference between the price at Rome and the profit received by a farmer at a remote place such as Sutrium is the cost of transporting the grain from Sutrium to

Rome. Draw a diagram showing the point along a road from Rome at which profit in wheat growing is zero. How much wheat is sent to Rome along the road?

A: The profit per bushel at any point is the price at Rome minus the cost of transporting the bushel from the point to Rome. The point of zero profit is where all the price is eaten up in transport costs. It is the Margin of Cultivation. If wheat produced per mile is a constant, then the amount sent to Rome is proportional to the distance from Rome to the margin of cultivation (see Figure 20.6, ignoring the shaded area for the moment).

Figure 20.6
The Margin of Cultivation Is Determined by Transport Cost

The fixed price paid for wheat at Rome goes either to pay for transport to Rome or as profits to wheat farmers. The extreme geographical limit of cultivation is the point at which transport costs equal price and profits are zero. A fall in transport costs shifts outward the Margin of Cultivation.



Measuring the Benefit of Cheaper Transport to the City

T or F: If the cost of sending wheat to Rome is cheapened by the construction of a better road, then the profitability (rent) of land increases by the same amount as transport cost falls.

A: The transport cost swings out to, say, the dashed line in Figure 20.6. The profit at Sutrium and elsewhere

Similarly, consider the following.

Q: You are attempting to measure the social value of the Ohio canals constructed in the early nineteenth century. You do not have good information on the fall in the cost of transport after the construction of the canals. You do, however, have good information on the *rise in the rent of land* around the canals. *True or false:* The rise in the rent is an estimate of the fall in the cost of transport.

Why Crops Around the City Fall into Zones: Bid Rent Curves

The same intellectual machinery can be turned to the question of what crops are grown where. Wool, for example, is grown in Northumberland or Montana, far from centers of population. Eggs, on the other hand, are grown close to where they are consumed. Why? One's instinct is to answer that wool is cheaper to transport than eggs, being more packable and less perishable. The instinct is useful but not complete. The complete answer views the crops as bidders for the land and asks what each crop is willing to pay.

Q: A certain crop yields Y tons per acre, sells for $\$P$ per ton at London, costs $\$T$ per ton to ship D miles, and costs $\$C$ per ton to raise. How much is it willing to pay at most for an acre located D miles from London?

A: An acre yields Y tons. Each of these Y tons would sell for P at London from which must be subtracted the cost of raising, C , and the cost of transport, TD . The upshot is

$$\text{Willingness to pay for an acre} = Y(P - C - TD)$$

The crop with the highest willingness to pay at each distance, D , wins the land. The equation can be rewritten in a more illuminating form by multiplying and dividing by P :

$$YP \times \left[1 - \frac{C}{P} - \left(\frac{T}{P} \right) D \right]$$

rises as the transport cost falls, simply because everything left over from subtracting transport cost from the price at Rome is the profit. The margin of cultivation, naturally, moves out. The shaded area is the increase in rent. Therefore, true.

A: The increase in an Ohio farmer's willingness to pay for land (that is, its rent) equals the fall in transport, because every dollar of transport of his goods saved is a dollar in his pocket. The fall in transport cost is like a subsidy to the production of buckeyes and popcorn and other Ohioan things. The land in each location is supplied inelastically, with the result that the subsidy falls on owners of land.⁹

Or in words,

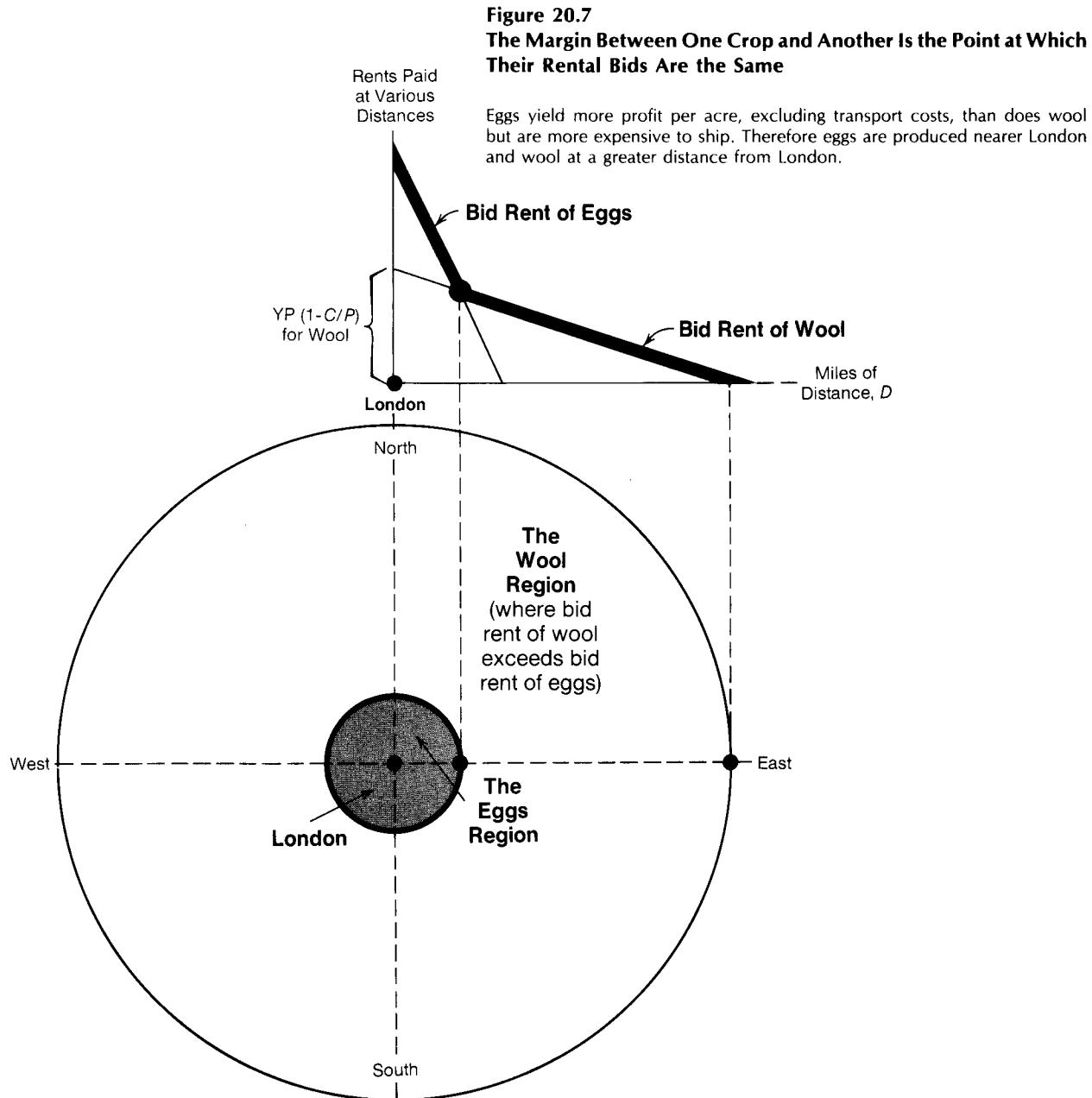
Dollars of yield per acre at London prices	\times	$1 - \frac{\text{cost per ton to raise the product as a fraction of the price at London}}{\text{the product of transport costs of anything for one mile per ton from London as a fraction of the price at London}}$]
--	----------	---	---

The equation is called *bid rent curve* for the crop, since it tells how much the crop will bid to rent land at various distances. Obviously eggs, for example, have a higher dollar yield per acre at London prices than does wool, because you can fit more hens producing valuable eggs on an acre than sheep producing wool. The transport costs per mile per ton for eggs, on the other

⁹ For further details see Roger Ransom, "Social Returns from Public Transport Investment: A Case Study of the Ohio Canals," *Journal of Political Economy* 78 (September 1970): 427-436.

hand, are obviously higher. In other words, in a diagram such as Figure 20.7, the intercept for eggs is higher, but so too is the slope. The results portrayed in Figure 20.7 are rings of cultivation around London, the division between rings being where one bid rent curve

falls below another. Where the bid for wool land exceeds that for egg land, as it does at such a distance that eggs cannot be delivered without excessive spoilage and breakage, the wool ring begins. And so forth, out to the last ring of cheap goods that are cheap to transport.



**The Value of
Distance from a
Good or Bad Thing**

The emphasized, outermost line in the top panel is how much rent successful bidders will pay for the land at various distances from the center. Whenever greater distance costs more, it slopes downward, as here. The center can be a central business district to which people wish to commute daily, in which case the rings are rings of differing land use: very near the center, office buildings; a little farther out, high-rise apartment buildings willing to pay a great deal per acre; still farther out, low-rise apartments and tightly packed single-family houses; and in the outermost ring, suburbia, with low land rents (because no one will pay premium prices to be 50 miles from work) and therefore large lots. The area under the bid rent curve is the rent of all land at various distances. Likewise, the value of other powerful amenities (a park, golf course, or lake) or disamenities (an airport, freeway, or dump) can be measured by the area under their bid rent curves.

Summary

The idea of "distance" between products is more fruitful when applied to literal distance. Hotelling's tale of stores along a road says much about the location of stores and about the location of political stores called parties. And once the diagram of rising cost with distance is developed, it says a great deal about location in general. It says, for example, that the midpoint of a road is the socially optimal location for a single store or a conference. It says, again, that remote locations are valued less than near ones, a simple idea that is the key to explaining the geography of production or to measuring the value of transport improvements or to planning parks or, a negative amenity, pollution.

EXERCISES FOR SECTION 20.2

1. Suppose that Political Spectrum Boulevard is 100 yards long and there is a voter located at every yard.
 - a. If the two parties are at first at the extreme, how many votes does each get at first?
 - b. If the Democrats move halfway toward the midpoint and the Republicans don't budge, how many votes does each party get? (*Hint:* Draw a diagram to scale, keeping all the "transport" costs equal and start with a triangle of equal division. See where the new Democratic line will cut the Republican half of the old triangle. Use a ruler or high school geometry.)
 - c. If the Republicans retaliate by *also* moving halfway toward the midpoint, how does the vote divide?
2. Suppose that a conference is to be held among six people arranged along a 3000-mile line at mile 0, 600, 1200, 1800, 2400, and 3000. If travel costs 20 cents a mile (round-trip):
 - a. Show how much it will cost to have a conference at mile 1500.
 - b. Show how much it will cost to have one at mile 600.
 - c. At mile 0.
3. Suppose that eggs sell for 62 pence a pound in London and wool for 230. The eggs cost on the farm 30 pence per pound and the wool 100 pence per pound. An acre can yield 2143 pounds of eggs and 120 pounds of wool a year. Eggs cost 0.5 pence per mile per pound to transport, wool 0.1 pence per mile per pound.
 - a. How much will an egg farmer be willing to bid for an acre of land (in pence of rent per year) if the acre is 10 miles from London? Use the simple bid rent equation.

- b. How much will the wool grower bid at 10 miles? Who gets the land?
- c. How much will each bid at 100 miles from London? Who gets the land?
- d. At what distance are their bids equal? Where will the kink be in the bid rent diagram?

PROBLEMS FOR SECTION 20.2

- o 1. The Hotelling model of the location of two soap stores assumes that the demand curve for soap of each of the consumers scattered along the road is perfectly inelastic. *True or false:* If it is somewhat elastic, then the two stores will not locate exactly side by side because their moves toward the middle of the road will discourage consumers on the ends from consuming at all.
- 2. What is the political analogy to Problem 1? Who is the “consumer”?
- 3. Imagine the road to be a circle rather than a line. Stores can locate anywhere along the circle. There is no end, for circles do not have ends. *True or false:* The Hotelling equilibrium along a circular road is both stable and socially optimal.
- 4. At the beginning of the nineteenth century a ton of iron made in England required 2 tons of iron ore and 8 tons of coal. By the end of the century it still required 2 tons of iron ore, but only 1.5 tons of coal. Where would ironmaking locate in the two periods relative to the (widely separated) deposits of coal and of iron ore?
- o 5. The chief cost of visiting Yellowstone Park is the time spent getting there. Yellowstone is 4 hours from Butte, Montana, 11 hours from Spokane, Washington, and 17 hours from Seattle. *True or false:* If one knew the percentage of each city’s population that visited Yellowstone in a year, one would have three points on the demand curve for visits per year (with price expressed in hours, which can be translated back into dollars at some wage rate per hour).
- 6. Using geometric methods (and given the simplifying assumptions in the text), show that the ratio of travel costs for a conference located at the end of the line of participants in New York to one located at the midpoint in Kansas City is 2.0. (*Hint:* Draw with some care to be accurate the transport cone out of the midpoint and the half cone out of the eastern endpoint. The areas under the cones are the costs, and if transport costs are constant, the edges of the cones are at the same angle to the line. Search the areas for similar triangles.)

True or False

- 7. If a third store joins the two already selling soap along the road in Problem 1, the equilibrium is instable.
- o 8. The Hotelling location for two firms is socially nonoptimal.
- 9. Land rents fall to zero as transport costs fall to zero.

Competition Among the Few

21.1 Simple Solutions: Bertrand and Cournot

What to Read For How do two sellers alone in a market behave? What if they cooperate fully? What if they maximize selfishly and assume the other will keep his price constant? What if they assume the other will keep his quantity constant? What are the *Bertrand* and *Cournot* solutions? How do you prove the *Cournot* solution diagrammatically? How do you apply it—or its “dual”—to the case of end-to-end railroads?

Starting with Few Competitors The last chapter attempted to develop a good theory that is intermediate between competition and monopoly by beginning with competition (many sellers) and letting in elements of monopoly (downward-sloping demand curves). It failed. The disappointment was softened by the uses found for the bits and pieces of theory constructed along the way. The present chapter picks up the other end of the stick, beginning with monopoly and introducing some elements of competition. It proves to be just as disappointing. Again, however, the ideas prove useful in other directions. And the repeated experience of failure is a maturing one, as it often is. It brings us to accept the hard truth that there is in fact no good theory intermediate between monopoly and competition.

Cooperation as a Solution Begin with a John Hughes, a monopolist of mineral water in Mudlavia, Indiana, who has no costs. Were the water supplied competitively it would be supplied at cost (zero). But Hughes is running a monopoly, not a charity. He sets his supply price or, with equivalent results, the quantity supplied—at the profit-maximizing level. Therefore, his marginal revenue on the mineral water is equal to its marginal cost (zero). The entire theory of *oligopoly* (few sellers) developed in economics can be summarized by imagining what happens if a new seller, Joel Mokyr, sets up shop beside Hughes. Two sellers is *duopoly*, the simplest case of oligopoly. If there is any reasonable story of how three or four or N sellers will behave, it should work for two.

One possible outcome is *cooperation*, with Mokyr saying to Hughes, “Look,

old buddy, we needn't quarrel: You just give me half your monopoly profits and we'll go on exploiting the customers to the optimal extent." Because the former monopoly price is the best that the industry can do, a cooperating pair of duopolists keep it.

The problem with cooperation, however, is achieving and maintaining it. Hughes will naturally resent the upstart's demand for half the profits. And Mokyr might as well ask for 99% of the profits as for 50%. Although a 50:50 split might seem "fair," there is in fact no reason other than social convention for adopting it. Even if some division of the spoils is agreed upon, each duopolist has an incentive later to cheat on the agreement if he can. And if the cooperation somehow survives both the initial bargaining and the subsequent cheating, its very success will attract others, yielding a three-, four-, and five-firm industry, and so on. Each new entrant will dilute the monopoly profit and complicate the already complicated agreement to cooperate.

Market Sharing Leads to the Cooperative Solution

T or F: If Hughes and Mokyr each believes that the other will in the end *match both his price and his quantity*, each will believe that he faces a demand curve that bisects the market demand, and each will voluntarily set the true monopoly price. (*Hint:* How will consumers divide if Hughes and Mokyr charge the same price?)

A: When Mokyr's price is the same as Hughes's, consumers divide randomly (and equally) between them. When neither duopolist attempts to sell more than the other, this equality is undisturbed (see Figure 21.1). The line marked One Half of Market Demand is therefore each

Though complicated and precarious, cooperation sometimes happens. And even when it is not explicit it can occur.

duopolist's demand curve when price and quantity are matched. At any particular price, this curve has the same elasticity as Market Demand. The simplest proof is that a constant elasticity demand curve such as $DP^{-\epsilon}$ continues to have the same elasticity, ϵ , when divided in half ($\frac{1}{2}DP^{-\epsilon}$ still has ϵ as an exponent). Therefore, each duopolist acting as a monopolist along his market share demand curve has the same incentive to set his price at the same markup over marginal cost as would a monopolist of the entire industry. An assumption of market sharing, in short, leads to the cooperative solution. Therefore, true.

Because each duopolist is led by market sharing to act in the (monopolistic) interest of the two together, duopolists (or N -opolists, for that matter) setting up a monopolistic conspiracy often in fact specify that sales will be divided in fixed shares of the total. Each member of the market-sharing conspiracy still has an incentive to cheat, which can be policed by central selling, one office dividing up the market. The only remaining obstacle in the way of quiet enjoyment of the monopoly thus arranged is that such arrangements are against the law.

Noncooperation: The Bertrand Solution or Price Competition

The alternatives to cooperation as a description of how the duopolists will behave are numerous. One is the price war of all against all, otherwise known as competition. That is, Mokyr could try to get all the market for mineral water by selling it for 1 cent below Hughes's old price. The assumption that Mokyr is making is that Hughes will not only not match his price but will not respond at all, or will respond so slowly that Mokyr will make a lot of money in the meantime.

The assumption is in fact a correct one for a single firm among 100,000 to make, for it is literally true that the market (the 99,999 other firms) will not

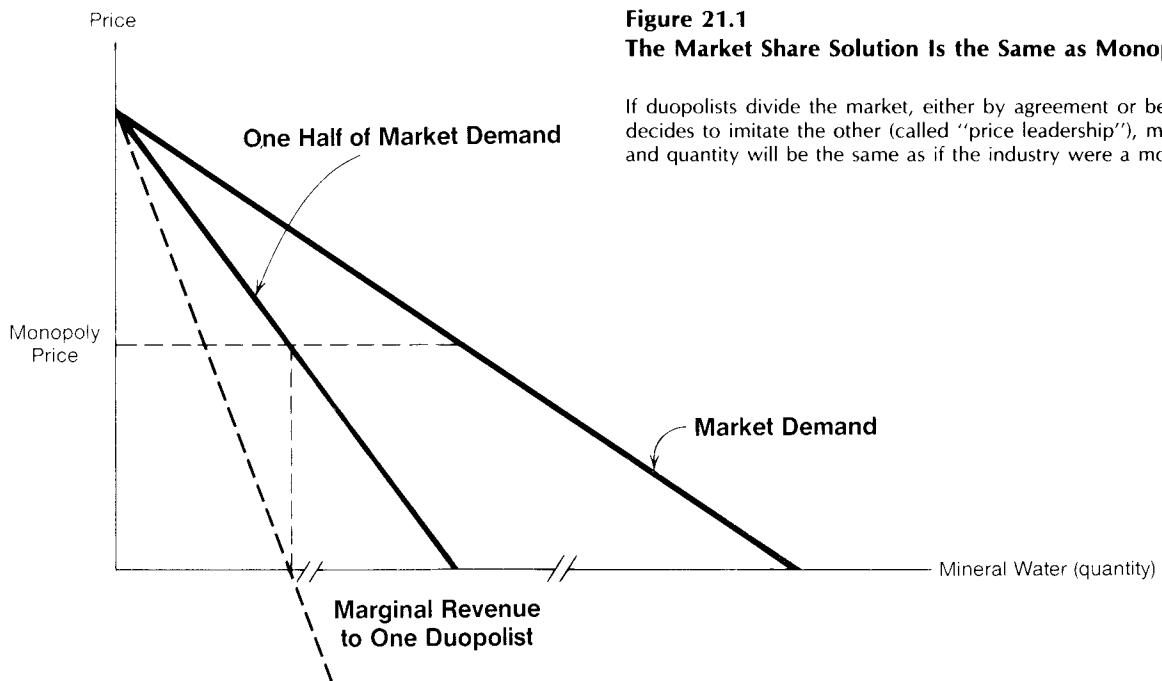


Figure 21.1
The Market Share Solution Is the Same as Monopoly

If duopolists divide the market, either by agreement or because one decides to imitate the other (called "price leadership"), market price and quantity will be the same as if the industry were a monopoly.

respond noticeably to the price cutting of one little firm unable to supply anything but a tiny share of the market. But 100,000 is very far from 2. If Hughes also is alert, he will respond, dropping his price 1 cent below the new price set by Mokyr and stealing back the entire market. In other words, since Mokyr is potentially a large part of the market, his initial shading of the price *does* have a noticeable effect on the other firms in the market—namely, a fall in Hughes's output from all the market demand to nothing—and therefore evokes a response.

The assumption that the other duopolist will keep his price fixed if one changes his price is in fact an irrational assumption for duopolists to make. If they nonetheless make it and persist in it despite all their unhappy experience to the contrary, the result will be a price war, eventually driving the price of mineral water down to marginal cost, namely, to zero. The analysis is known as the *Bertrand solution* to the oligopoly problem. Some solution.

Noncooperation:
The Cournot Solution or Quantity Competition

The other main solution to the oligopoly problem is the *Cournot solution*.¹ In the Bertrand solution each duopolist takes the other's *price* as given. In the Cournot solution he takes the other's *quantity* as given. The "given" amount is subtracted from the market demand to give the demand curve the duopolist believes he faces. Look at Figure 21.2. The Belief Curve will be parallel to the

¹ After A. A. Cournot (1801–1877), a French economist who proposed it in his *Researches on the Mathematical Principles of the Theory of Wealth*, first published in 1838, English trans. by Nathaniel Bacon (New York: Macmillan, 1897). The "Bertrand" of the solution is Joseph, in *Journal des Savants* (September 1883), which was based on a misunderstanding of Cournot. "Cournot" is pronounced "Coor-No."

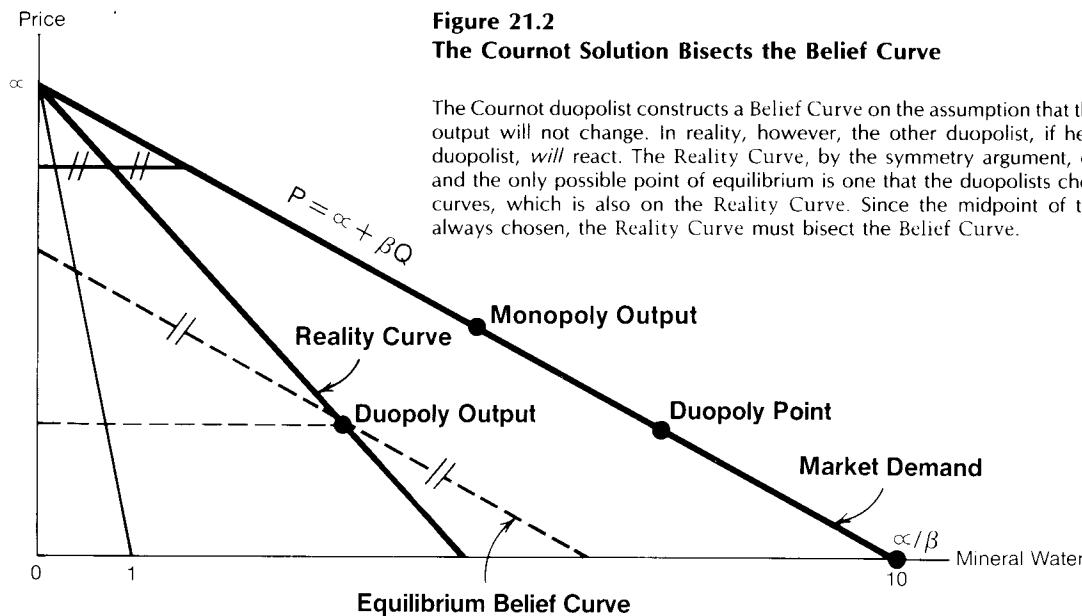


Figure 21.2
The Cournot Solution Bisects the Belief Curve

The Cournot duopolist constructs a Belief Curve on the assumption that the other duopolist's output will not change. In reality, however, the other duopolist, if he is also a Cournot duopolist, *will* react. The Reality Curve, by the symmetry argument, divides the market, and the only possible point of equilibrium is one that the duopolists choose on their belief curves, which is also on the Reality Curve. Since the midpoint of the Belief Curve is always chosen, the Reality Curve must bisect the Belief Curve.

Market Demand, because a constant amount is subtracted at every price. Having a demand curve before him, he will choose the best position along it as though he were alone. Whatever the duopolist believes he faces, however, he actually faces one-half of market demand, his Reality Curve, because in equilibrium two indistinguishable duopolists must have the same price and output. Full equilibrium therefore requires that the point the duopolist chooses along his Belief Curve be sustainable, that is, *be also a point along his Reality Curve*. For a straight-line demand curve and no costs, the condition is easily represented, because the duopolist's optimum is simply the midpoint of the Belief Curve, this being the point of maximum revenue (elasticity = -1 and no costs). That is, one shifts the Belief Curve (which is parallel to the whole demand curve) up and down the Reality Curve (which is the bisection of the demand curve) until the Equilibrium Belief Curve is bisected—the bottom dashed line.

A number of results follow from the diagram. For one thing, the diagram permits an easy extension to three, four, . . . , N sellers. Look at the very steep Reality Curve close to the price axis, equal to a tenth share of the whole demand (when there are ten sellers). The only dashed-line belief curve that will be bisected by such a steep reality curve is a very low one. That is, as the number of sellers increases, the price resulting from Cournot behavior gets smaller, becoming finally the competitive price—here, zero. Furthermore, with Cournot behavior an oligopoly falls neatly between a monopoly and a competitive industry.

The Calculus of Cournot

The calculus of the Cournot argument leads of course to the same result, but the economics is somewhat more transparent. The output of the whole industry is the output from the single firm under examination, Q_i , plus the outputs from the $N - 1$ other firms, each having output Q_j . The distinction in the

subscripts is temporary. In a moment it will be acknowledged that all firms, being by assumption identical, will in the end produce identical outputs, that is, that $Q_i = Q_j$. The reason for the distinction initially is to capture the economic notion that the i th firm maximizes its profits assuming that the other $N - 1$ firms will not change their output. The profit to be maximized is the i th firm's quantity times the market price (since there are no costs of the mineral water). The market price, in turn, depends on the whole output brought to market, which is $Q_i + (N - 1)Q_j$. For a straight-line demand curve $P = \alpha + \beta Q$, with price intercept α and slope β , then, the profit is

$$Q_i P = Q_i [\alpha - \beta(Q_i + (N - 1)Q_j)]$$

(Recall that profits equal total revenue minus total cost. But costs are zero, so profits equal $Q_i P$.) Multiplying this equation out and taking its derivative with respect to Q_i (that is, holding constant the Q_j) gives the following expression, to be set equal to zero to achieve maximum profit:

$$\alpha - 2\beta Q_i - \beta(N - 1)Q_j = 0$$

The Q_j can be eliminated by now recognizing that $Q_i = Q_j$ in equilibrium, which implies that $\alpha = Q_i \beta(2 + N - 1)$ or $Q_i = [1/(N + 1)](\alpha/\beta)$.

This is the quantity of each firm's output resulting from Cournot behavior. Since there are N firms, the whole output of the industry is $[N/(N + 1)](\alpha/\beta)$. When the number of firms is very large, $N/(N + 1)$ is very close to 1, and output is very close to the competitive, zero-cost output, namely, the output at the intercept of the demand curve along the quantity axis, α/β . When, at the other extreme, N is 1 the case is simple monopoly, and output is half $[1/(1 + 1) = \frac{1}{2}]$ the competitive output. The case of duopoly leads to an output equal to two-thirds of the competitive output, tripoly to three-fourths, and so on.

The Application of the Cournot Solution to the Analysis of Large Numbers of Sellers

All this is charming. For large N it is also reasonable. Indeed, the Cournot assumption (that others will keep their quantities constant) is one way of demonstrating the high elasticity (relative to that of the market demand) of the residual demand curve facing one competitor among many.

T or F: In the case of a straight-line demand curve and no costs, the elasticity of the belief curve in equilibrium is always 1.0.

A: True, because the firm will always choose the midpoint of the belief curve (where elasticity is 1.0), because this is the best it can do.

T or F: The elasticity of the market demand curve (the reality curve) in Cournot equilibrium is $1/N$. (*Hint:* Write out the expression for elasticity as $(P/Q)(\Delta Q/\Delta P)$. Put price on the horizontal axis. Substitute $-1/\beta$, that is, the slope of the demand curve, for $\Delta P/\Delta Q$. Also substitute the Cournot equilibrium values of P and Q in terms of N and β for the P and Q .)

A: Taking the hint, fill in the expression for elasticity with $\Delta Q/\Delta P = -1/\beta$, $Q = [N/(N + 1)](\alpha/\beta)$, and solving for P by substituting the expression for Q into the market demand curve

$$P = (\alpha - \beta Q) = \alpha - \left[\beta \left(\frac{N}{N+1} \right) \left(\frac{\alpha}{\beta} \right) \right] = \alpha \left[1 - \left(\frac{N}{N+1} \right) \right]$$

The result is

$$\frac{P}{Q} \left(\frac{\Delta Q}{\Delta P} \right) = \frac{1 - (N/[N+1])}{\alpha/\beta(N/[N+1])} \left(-\frac{1}{\beta} \right) = \frac{1}{N/[N+1]} - 1 = \frac{1}{N}$$

Therefore, true.

T or F: Therefore, the elasticity of the single firm's belief curve is N times the elasticity of the market demand.

A: Since 1 is N times $1/N$, true.

The proof in the three questions is in fact generalizable to any demand curve, whether straight line or not. The proof consists of the simple algebra of elasticities, as follows. If George Marr's firm faces other firms producing Q_0 and the whole market demand is Q , then evidently Marr faces a quantity demanded of $Q - Q_0$. Leaping to elasticities in the style of Chapter 7, his elasticity of demand is

$$\frac{Q}{Q - Q_0} (\epsilon^D) - \frac{Q_0}{Q - Q_0} (\epsilon^S)$$

where ϵ^D is the elasticity of demand of the market and ϵ^S is the elasticity of supply by the other suppliers. But the Cournot assumption is precisely that the elasticity of supply by the other suppliers is zero. Since ϵ^S is zero, the expression reduces to $[Q/(Q - Q_0)]\epsilon^D$. But Marr's company is a typical one among the N . So $Q - Q_0$ (which is Marr's output) is to Q as 1 is to N , which is to say that $Q/(Q - Q_0)$ may be replaced by $N/1$. The elasticity facing Marr is therefore N times the elasticity of demand; that is, it is $N\epsilon^D$. Were the number of firms 100 and the market elasticity 2, the elasticity facing one competitor would be 200 if he made a Cournot assumption.

***The Application of
the Cournot
Solution to the Case
of End-to-End
Railroads and Other
Matters***

It is apparent, then, that Cournot and similar assumptions pop up in many different parts of economics. Here the Cournot assumption has popped up in the theory of the competitive firm facing large numbers of others. In the last chapter it popped up in the Hotelling analysis of political location, with each party making the assumption that the other would not move. And the Bertrand assumption of given prices, a closely analogous assumption, underlies the theory of monopolistic competition. It is clear that the Cournot assumption is hard to beat in its simplicity and the definite results it yields.

The best example is end-to-end railroads. Suppose that the trip of a ton of wheat grown in Scott City, Kansas, and destined for bread in Philadelphia is broken into two legs, the first from Scott City to St. Louis on the Missouri Pacific Railroad and the second from St. Louis to Philadelphia on the Baltimore & Ohio. Consider the decision facing each railroad of what to charge the ton of wheat as freight. The number of such tons transported from Scott City all the way to Philadelphia depends of course on the combined prices for freight, not on how the combined price is divided up between the two railroads. Consumers of railroad services do not care how the price is divided. The separate railroads, of course, do care, very much.

Q: Assume for simplicity that the two railroads have no costs, no competitors, and no other trade but the Scott City–Philadelphia one and assume that the demand for tons transported from Scott City to Philadelphia is a straight line.

1. If the two railroads were merged into one company, what combined price would the company set? Draw the diagram exhibiting the setting.

2. Suppose that the two railroads are separate and that each takes the other's *price* as given, unaffected by what the one railroad decides to do. Draw the de-

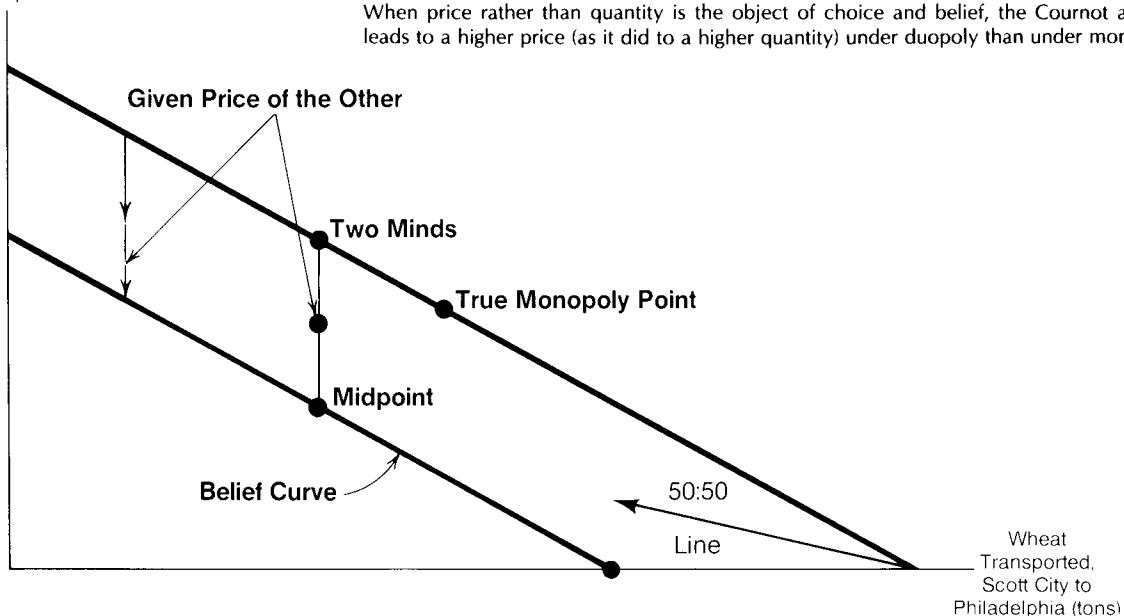
mand curve that the Baltimore & Ohio believes it faces if it believes that the Missouri Pacific's Price for the Scott City–St. Louis leg of the journey is given. In view of the shape and position of this belief curve, what price will the Baltimore & Ohio charge on its leg?

3. Compare the answers to (1) and (2). Which situation—one big railroad or two end-to-end railroads—results in more profits for the railroad industry? Which results in lower prices for consumers? Does the breaking up of vertically integrated monopolies lead always to better outputs, socially speaking?

Wheat Price per
Ton Transported
Both Combined
and Separate

Figure 21.3
End-to-End Monopolies Are the Dual of Cournot Duopolies

When price rather than quantity is the object of choice and belief, the Cournot analysis leads to a higher price (as it did to a higher quantity) under duopoly than under monopoly.



A: 1. The one big railroad would move to the True Monopoly Point in Figure 21.3, halfway down the heavy demand curve (since there are no costs).

2. The demand curve that the Baltimore & Ohio thinks it faces is the inner line in the diagram, that is, the combined demand price minus a Given Price charged by the railroad on the other leg of the journey. Notice the parallels with the Cournot assumption and its diagram: The combined price (like the combined quantity in the Cournot analysis) is fixed by the action of both parties, each party taking the price (the quantity in Cournot) of the other as given and both having to match quantities (prices in Cournot). The parallelism is another example of the pervasive duality of price and quantity in economics.² In any case, the Baltimore & Ohio will choose the midpoint of its Belief Curve, giving a combined price of Two Minds.

3. Regardless of how exactly the railroads divide up the price, it will be higher than the price of the true monopoly. This follows from the way in which the

given price of the other pushes the demand curve facing the one railroad downward beneath the whole demand curve. The midpoint of the Belief Curve will always lie at a lower quantity than will the midpoint of the demand curve, since the Belief Curve is closer in. In consequence the combined profits of the two railroads (here equal to their revenues since costs have been assumed away) are less than they might be. Since a monopoly gets the largest profits, any divergence from the monopoly point reduces profits. The price, in short, is set too high from the industry's point of view.

It is also set too high from the consumer's point of view. Any price is bad for consumers, and any price above zero for a costless good is bad for society as a whole. The higher the price, the worse, and the case of end-to-end companies leads to an even higher price than does a monopoly. The breaking up of a monopoly into parts that supply each other—as the Missouri Pacific supplies the Baltimore & Ohio with freight—is not always going to lead to better (larger) outputs.³

² Hugo Sonnenschein, "The Dual of Duopoly Is Complementary Monopoly, or Two of Cournot's Theories Are One," *Journal of Political Economy* 76 (March–April 1968): 316–318.

³ The full solution to the end-to-end problem requires some assumption about what division of the price is an equilibrium division. That is, there has to be an analog to the assumption in the

The end-to-end problem appears in many other forms. Cournot himself applied it to the making of brass out of zinc and copper. Since zinc and copper are used in fixed proportions, brass is analogous to the journey from Scott City to Philadelphia, each leg of which is used also in fixed proportions. An integrated monopoly would result in a lower brass price, Cournot argued, than would a zinc firm separated from a copper firm. The Cournot assumption is clearly a useful one to begin an analysis.

But is it useful to end?

Summary The analysis of competition among the few can begin with monopoly (one seller) and work up toward many. The first step is duopoly (two sellers), which suffices to illustrate all the promise and problems. If duopolists cooperate, whether explicitly by contract or implicitly by assuming that each will take half the market, then duopoly reduces to monopoly. Any other assumption leads away from the monopoly solution. For instance, if the duopolists make the Bertrand assumption that the other will keep constant his price, the price will fall to the competitive level. For another instance, if the duopolists make the Cournot assumption that the other will keep constant his quantity, the quantity will rise toward the competitive level, stopping well short of it for two sellers and getting very close to it for some large number of sellers. For a large number of sellers, indeed, the Cournot solution provides one way of demonstrating the high elasticity of demand facing a single seller. For a small number of sellers the Cournot assumption provides a way of beginning the analysis: of Hotelling location, of monopolistic competition, of end-to-end monopolies. But all is not well in the small-number case. Does the Cournot assumption make sense?

EXERCISES FOR SECTION 21.1

1. Suppose that the demand curve for mineral water from Mudlavia is $P = \$10 - 0.01Q$, where P is dollars per bottle and Q is the number of bottles.
 - a. One of the prices to choose is the one that maximizes the profit from selling mineral water (the cost, you recall, is zero). Which one is it? \$1 a bottle? \$8 a bottle? \$6 a bottle? \$5 a bottle? Note that maximum industrywide profit is profit when the industry acts as a monopoly.
 - b. What is the profit-maximizing *quantity*? What is the total revenue (here, profit)?
 - c. The Cournot equation derived by calculus in the text says that industry output for a Cournot oligopoly with N firms facing a straight-line demand curve $P = \alpha - \beta Q$ is $[N/(N + 1)] (\alpha/\beta)$. Use the equation to determine industry output, price, and total revenue (profit) for the duopoly of Hughes and Mokyr. How much does each duopolist make? Compare the results with the monopoly case.
 - d. Use the Cournot equation to check the result of (b).
2. What would be the output, price, and profit of the Mudlavia oligopoly if Hughes and Mokyr were joined by Allen, Bittlingmayer, Masuda, Flynn, Gunderson, Leff, Smiley, White, Weir, and Vedder?

Cournot analysis of a duopoly that duopolists charging equal prices will receive each half the total quantity. Formally speaking, the analog is a line bisecting the demand price (which is here drawn as the beginning of the 50:50 Line), in which case the entire diagram becomes an exact dual of the Cournot diagram. Price replaces quantity at every stage of the argument. But aside from its "fairness" a 50:50 division of the price has no persuasive claim to being an equilibrium. The end-to-end problem, therefore, cannot be solved quite so neatly as can its Cournot duopoly dual.

21.2 The Irrationality of Simple Solutions to the Problem of Fewness

What to Read For

What is wrong with the Cournot assumption? Why is pure bargaining relevant to what is wrong with it? What is a reaction function? How is it related to the Cournot assumption? Can it be used as a full representation of the arms race? Could Mokyr do better if he knew Hughes was making the Cournot assumption? What does the principle of outsmarting have to do with the Cournot assumption? Should you believe stock market advice?

The Inapplicability of the Cournot Solution to Rational People

Were the number of firms small, Marr or Mokyr or any manager would be crazy to make a Cournot assumption. This fact rather spoils the prettiness of Cournotesque arguments. The manager is supposed to treat the actions of others as given. But the actions of others change.

The problem is that when numbers get small it is no longer reasonable for a firm to believe that its own actions have a trivial impact on the other firms. On the contrary, to return to the mineral water duopoly, Mokyr's setting of quantity or price radically changes the environment that Hughes faces, and Hughes reacts. The point is that Mokyr would *know* that Hughes would react—contrary to the assumption that in the Cournot or Bertrand solution he is supposed to make. Mokyr would therefore be foolish to make the assumption that Hughes will keep his quantity or price unchanged. He would be doubly foolish to persist in the assumption after a couple of rounds of reaction from Hughes had shown him that Hughes *does* react. To assume foolish behavior in the science of rationality is at best aesthetically displeasing, at worst wrong.

In other words, the difficulty with the Cournot or Bertrand solution to the problem of competition among the few is that the few are by virtue of their very fewness placed in an environment of pure bargaining. Duopolists do depend on each other, and it would be foolish not to recognize the mutual dependence. Interdependence recognized is bargaining. That is, the situation is similar to those early in Chapter 5 on exchange, in which West Germany trades pipes for gas from the Soviet Union or you trade money for a car from Dario Comi. To the extent that the transactions occur in isolation, the parties do not face prices given by the market. The outcome is unpredictable. Likewise, Hughes and Mokyr are in essence bargaining with each other as they set prices and quantities. The Cournot solution solves the difficulty of bargaining by assuming it away.

Pure Bargaining as the Game of Chicken

An extreme case of the difficulty of analyzing pure bargaining is the game of chicken. This manly sport, alleged to have been popular among adolescent motorists in the 1950s, consisted of driving two cars at high speed on a collision course. The first to turn aside lost his honor. If neither turned aside both lost their lives. If Don Paterson, for instance, knows that Marvin McInnis will turn aside when the cars are 100 feet apart, Paterson can plan confidently to drive past the 100-foot mark—to say 80 feet—and win. But if McInnis knows that

Paterson (thinks he) knows—and McInnis would know if he were fully rational, because it is his words and actions that give away the knowledge—then McInnis can stay on line until 60 feet and win. But if Paterson knows that McInnis knows that Paterson knows that McInnis will turn aside at 60, Paterson will by the same reasoning know to stay until 40 feet. And so forth. But zero feet is not the inevitable outcome, either, for both wish to avoid it. Bluff, a reputation for insanity, quickness of reflexes, understanding of the other party's psychology all play a part.

T or F: If McInnis chains himself to the car and locks the steering wheel irrevocably straight and tells Paterson that he has done so, McInnis will win.

A: True, so long as McInnis cannot be outflanked in turn. McInnis has removed the decision to chicken out from his own hands. A rational Paterson will not play a game of choice with a machine that has no choice but to go straight. On the other hand, Paterson could

do the same to himself in his own car. The question would then be which competitor was sincere in his hell-bent plan. The competition simply rises to a higher plane, the trick being now to convince your adversary that you are *really* irrevocably committed to a straight course. Having done so, it then rises to a still higher plane. And so forth, on into the hall of mirrors of one intelligence trying to outsmart another.

The games of adults have a similar structure.

T or F: If the United States wants to dominate the Soviets, it should announce that it will bomb Moscow the next time the Soviet Union even shows up late to a diplomatic cocktail party, much less invades an ally of the United States.

A: Uncertain. If the Soviets believed that the United States was so touchy and violent that it would in fact carry out the threat, then Soviet insults of all sorts would cease forthwith. We would have persuaded the Soviets that "we mean business." On the other hand, it would be difficult to persuade them to the belief, and to do so might be expensive in little wars. And

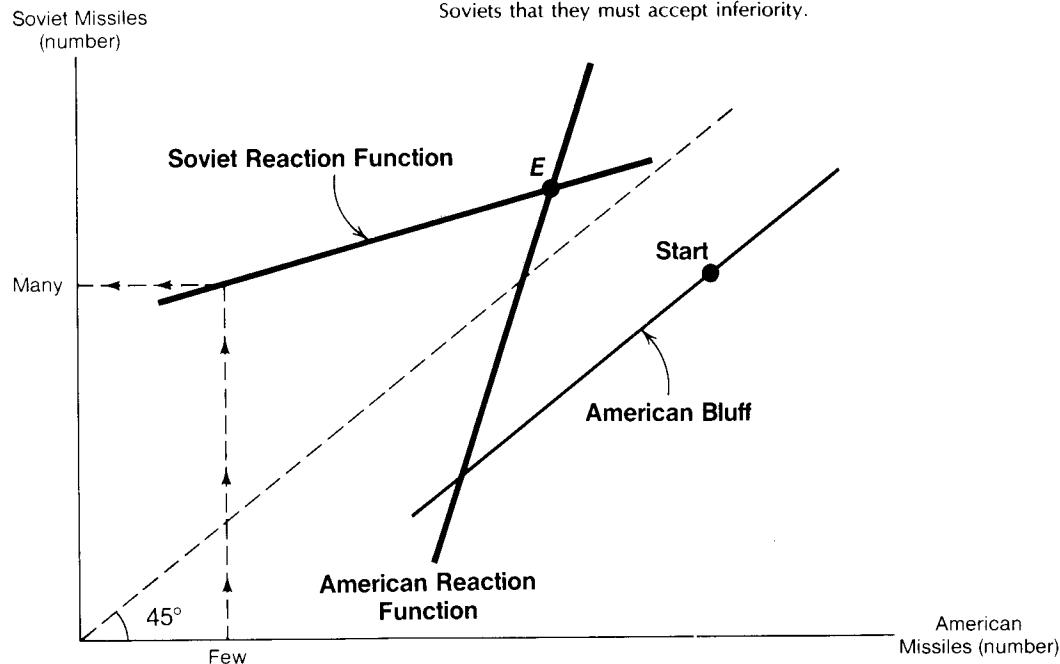
the same or offsetting strategies are open to the Soviets, who must be assumed to be at least as well informed as we are about the possibilities of international poker. For example, the Soviets can promise, and can attempt to convince us of the sincerity of their promise, to bomb New York and Iowa City if we bomb Moscow. To this we can reply that we will bomb Kiev and Odessa as well. And so on, back into the hall of mirrors. It is like the childish game of insults: "You're a silly head"; "So are you, twice over"; "Anything you say a thousand times over"; "Anything you say a million times over"; and on to exhaustion.

The Arms Race Analyzed à la Cournot, and Why It Cannot Be So Analyzed

It is plain in a general way that the Cournot "solution" cuts short the chain artificially. For instance, the international arms race can be illuminated by its analogy to a duopoly problem under the Cournot assumption, but only as a very crude first step that must be followed by other steps. The 45° line in Figure 21.4 represents a condition of equal numbers of Soviet and American missiles. If Soviet strategists insist always on having more missiles than the Americans, then the Soviet *reaction function* would be everywhere above the 45° line. The existence of few American missiles, for instance, results in many Soviet missiles (follow the dashed arrow). The reaction functions, if fixed, embody a Cournot assumption, the Soviets assuming that American missiles will stay at Few while the Soviets adjust their own missiles up to Many. They never learn that the Americans also have a reaction function. And this is not the only problem.

Figure 21.4
The Arms Race as a Cournot Problem

Suppose that the Soviets naïvely attempt to have more missiles than the United States, without realizing that the United States will continue to build more missiles if the Soviets do. Then the arms race will never reach equilibrium, unless the United States is content to have fewer missiles than the Soviets, or unless the United States can persuade the Soviets that they must accept inferiority.



T or F: If the United States is equally insistent on superiority, the arms race has no end.

A: If you draw in an American curve below the 45° line, you will see that there is no point of equilibrium. Therefore, true.

Still, the diagram seems usually to give a result.

T or F: As the American reaction function is drawn in the diagram, the arms race is stable at E .

A: True, as you can see by starting at, say, Start and seeing how each party moves.

The fundamental trouble with such an analysis is that the Americans and Soviets would be irrational to react so predictably to the behavior of the other, just as a duopolist would be irrational to react so predictably to the outputs of the other. The reaction curves are themselves at choice. Suppose, for example, that America wished the equilibrium to be below the 45° line, with American superiority in numbers of missiles. Beginning at Start, America could announce to the world that the light line through Start was in fact how it would react. If the Soviets were buffaloed by the announcement, they would change their curve to allow it to intersect with the American bluff, since the Soviets no more wish to spend all their income on weapons (the ultimate in a runaway

arms race) than do we. The American announcement is like McInnis's chaining his steering wheel straight ahead. But it has the same problem of outflanking: The Soviets can announce a similar, even more belligerent, policy. We are back to bargaining.

The Principle of Outsmarting

Anytime that it is possible to think of a better strategy, in short, the analysis is not finished. This *principle of outsmarting* in the analysis of game playing is the same as the principle of entry in the analysis of the firm. If the participants can still earn profits by intelligent entry, or by intelligent rethinking of their strategy in the game, it behooves the economist to imagine the consequences of entry or rethinking. If Hughes of mineral water fame, for example, acted like a Cournot duopolist, his competitor Mokyr could make more money by acting differently. The wise duopolist knows not merely the other's output (as the Cournot solution has it) *but his policy*. With such superior knowledge of his environment, Mokyr can obviously do at least as well as he can by stumbling about in Cournot fashion.

The point is easily demonstrated by giving another (the third) representation of the Cournot argument (the other two, you recall, are the argument from bisection of the belief curve and the argument from the intersection of reaction curves). If Hughes takes Mokyr's output as given at, say, Duopoly in the top panel of Figure 21.5, then Hughes will take the dashed line to be his own marginal revenue. That is, Hughes foolishly thinks that his demand curve begins where Mokyr's present output leaves off. The situation illustrated in the top panel is in fact the Cournot duopoly solution. Each duopolist sells one third of the competitive output for a total of two-thirds. Since both have the same output, Mokye's profit (shaded) is the same as Hughes's. Notice also one geometric fact that will prove useful in a moment. Because the curves involved are here assumed to be straight and Hughes acts like a monopolist along what he thinks is his demand curve, the Actual Price is half the Solitaire Price that would occur if Mokyr produced the output in the absence of Hughes. Therefore, the area of Mokyr's profit is half the emphasized rectangle inscribed under the demand curve. As long as Hughes acts like a monopolist, Mokyr's profit will always be half of such an inscribed rectangle of total revenue.

Mokyr wishes to do better. He can if he exploits his understanding of Hughes's behavior. He gets half of an inscribed rectangle, and knows it. So his smart move would be to arrange matters to get the biggest rectangle.

Q: What solitaire price (that is, what output) should Mokyr choose in the bottom panel to make the best of the situation?

A: Mokyr should choose the output that maximizes the area of the rectangle inscribed at Solitaire Price: If you get half of something, you want to make the something as large as possible. The largest rectangle is obviously

the monopoly rectangle. So he stops sitting there like a fool producing the one-third of competitive output that is the Cournot output for him; he produces instead one-half of the Competitive Output. In the event Mokyr gets Mokyr's Maximum Profit, which is larger than Mokyr's Duopoly Profit, Hughes, of course, is hurt, and consumers, it happens, are helped, because output is closer to competitive.

To repeat, then, *the Cournot and related solutions violate the principle of outsmarting, or of entry, because it is possible to think of a feasible way in which a Cournot duopolist could improve upon his situation.* The principle of outsmarting is a principle of humility, in the following sense. Imagine that

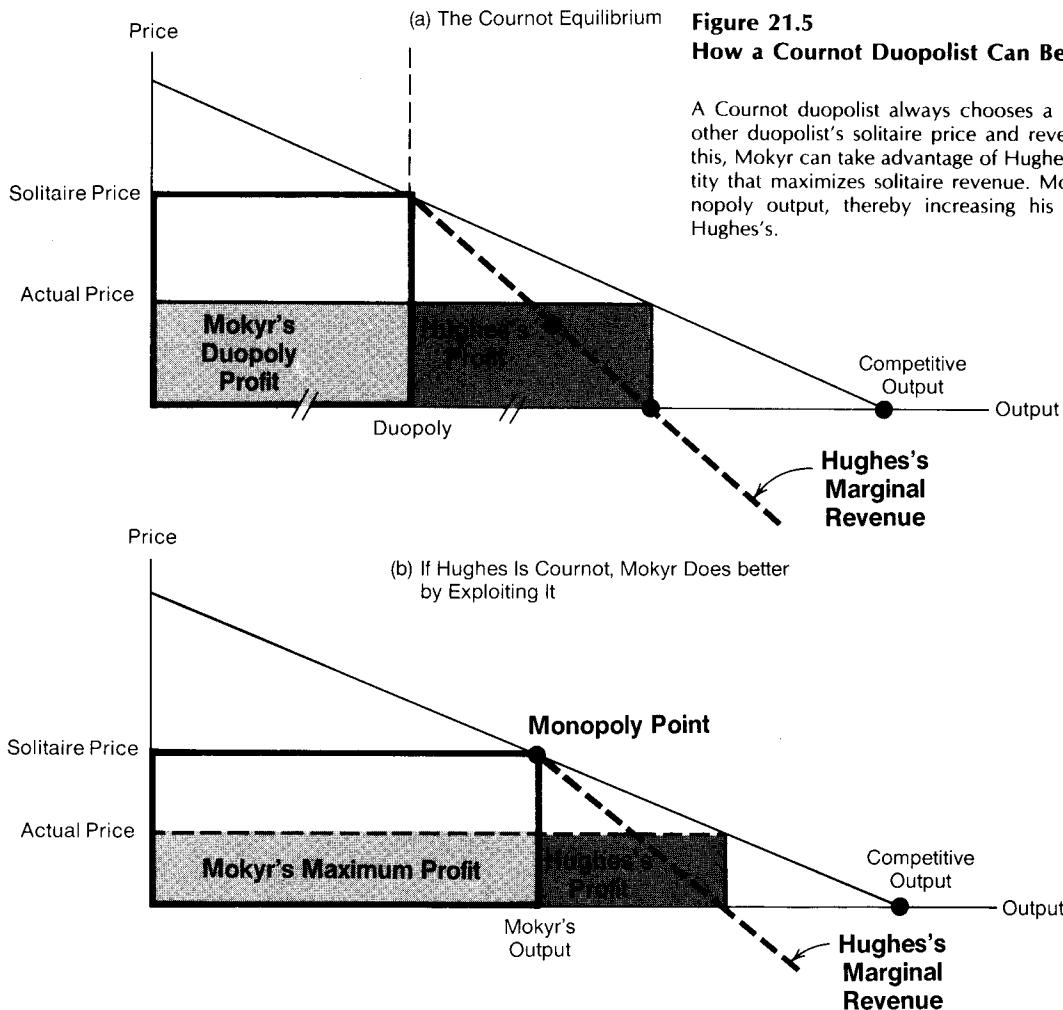


Figure 21.5
How a Cournot Duopolist Can Be Outsmarted

A Cournot duopolist always chooses a quantity that cuts the other duopolist's solitaire price and revenue in half. Knowing this, Mokyr can take advantage of Hughes by choosing a quantity that maximizes solitaire revenue. Mokyr chooses the monopoly output, thereby increasing his profits and reducing Hughes's.

Q: If you could predict when the economy will turn around (moving from boom to bust or bust to boom) and could convince some bankers that your prediction was true, then you could make a very large amount of money by investing in ways that used the prediction. *True or false:* Therefore it must be false that you can predict when the economy will turn around.

A: According to the principle of outsmarting, an analysis is unfinished if it leaves open an opportunity for one

of the parties to better her position. Well, you as the economic seer are one of the parties. Therefore, true. As was remarked in Chapter 14, which arrived at the same point, you should be suspicious of government spokespeople or financial journalists who predict what the future will bring, if their predictions would allow someone to make money. And you should be suspicious of "solutions" to the oligopoly problem.⁴

Summary The Cournot solution, then, is irrational. A Cournot duopolist can always do better by not acting like one, and by taking advantage of the simplemindedness of his competitor who goes on acting like one. But his competitor, too, is no fool. He too will recognize that his competitor does *not* keep selling the same quantity regardless of the quantity sold by his competitor. To assume otherwise is to violate the principle of outsmarting, which is a version of the principle of entry. The economic analyst should not suppose he is a better businessman than his object of study, or that the businessman is easily outsmarted.

EXERCISES FOR SECTION 21.2

1. In the game of odds and evens the two players stick out simultaneously any number of fingers, calling out "odd" or "even" as they do so. If the sum of the two sets of fingers is odd and Carla Oakes said it would be, she wins the stake. If the sum is even, and her opponent Elyce Rotella said so, then Carla loses. If both are wrong, nothing happens. Edgar Allan Poe argued in one of his short stories that an observant person could always win such a game, by guessing how many steps the opponent would take. First step: "Let's see: She put out an even number before; so I'll put out even, too, and call out even." Second step: "No, on second thought, she's smarter than that. She'll know I'm thinking the First Step, and will put out odd, calling out odd. So I'll put out odd, getting an even sum, and call out even." And so forth. Poe argued that one could guess how far into the hall of mirrors a person would go by knowing how intelligent they were. What do you think of Poe's argument?

2. Go back to Exercise 1 in Section 21.1, which calculated the equilibrium for the duopolists Hughes and Mokyr. Recall that in Cournot equilibrium they were each selling 333.3 bottles for a total of 667. Each made a profit of \$1110.

a. Using Figure 21.5 as a guide, what profit could Mokyr make by outsmarting Hughes? How much does he gain over continuing as a naive Cournot duopolist?

b. To check this result for consistency, find out what Hughes would produce when Mokyr outsmarts him (he goes on assuming that Mokyr's output is unalterable, and acts so to speak like a monopolist on the portion of the demand curve left over). Add the Mokyr and Hughes output. What is the market price? Therefore, what is Mokyr's revenue? What is Hughes's revenue?

⁴ The first formal statement of the point in the context of prediction was John F. Muth, "Rational Expectations and the Theory of Price Movements," *Econometrica* 29 (July 1961): 315–335, reprinted in Arnold Zellner, ed., *Readings in Economic Statistics and Econometrics* (Boston: Little, Brown, 1968). Muth's article is difficult. It has spawned, with a considerable lag, a large literature in macroeconomics.

21.3 Cartels and Game Theory as Solutions to the Oligopoly Problem

What to Read For

What are the advantages of a successful cartel and the disadvantages of its collapse? What is the *prisoner's dilemma*? How does it suggest cartels will work out? What is a *payoff matrix*? What is the *Cocktail Party Problem*? How can the *payoff matrix* be altered by punishment to bring about cooperation? Will a small group able to communicate be able to avoid the *prisoner's dilemma*? What is game theory? Does it solve the oligopoly problem?

Monopoly Itself Is an Economic Good

Useful as it is for beginning all manner of economic analyses, the Cournot and related assumptions do not lead to useful solutions to the original problem, namely: How will a group of firms behave that are neither perfectly numerous nor perfectly monopolistic? The Cournot assumption makes them out to be perfectly foolish. Perhaps it will prove more fruitful to imagine the results of assuming them to be perfectly wise.⁵

What is the best that a wise group of competing taxi firms can do? Obviously the best is to stop competing with each other for customers. That is, if there were no costs of collusion, the best that the taxi firms could do would be to collude to form a perfect taxi monopoly, dividing up the spoils among Checker, Yellow, and so forth in some agreeable way. The reason that collusion does not always happen is that it is expensive to enforce. Collusion among oligopolists, in other words, has a price, and the oligopolists will buy much or little of it depending on whether the price is high or low. An agreement to collude must be enforced, and police officers (literal or figurative) cost money. Industrial spies, lawyers, accountants, economists, and other worthy sorts might be employed to detect cheating on the agreement, but secret price cutting is usually possible to some degree, especially if the agreement is itself illegal—as collusive agreements are under the common law.

The reasons that resources must be spent to defend perfect collusion are apparent from two observations. The first is that a successful cheater on the agreement does better for himself than if he had cooperated with the agreement. A successful thief does better by cheating on the social agreement not to steal than if he were law abiding. Likewise, successful violators of the agreement by all taxi firms to charge a high price on a low volume does better than if he kept his promise. The second observation, however, is that if all members of the agreement in fact cheat, then the agreement is a dead letter. The thief (now living in a world of thieves) and the cheating taxi firm (now living in a competitive industry) are worse off than they would have been if they and their colleagues had cooperated. In a world of thieves and price cutters it is foolish to continue to obey the old laws against stealing and price cutting.

⁵ Or so it seemed to George Stigler, who accomplished this reorientation of the analysis in his "A Theory of Oligopoly," *Journal of Political Economy* 72 (February 1964): 44–61, reprinted in his *The Organization of Industry* (Homewood, Ill.: Irwin, 1968), to which subsequent reference is made.

The Prisoner's Dilemma

The situation is known as the *prisoner's dilemma*, from the following tale. Michelle McAlpin and M. Morris, two members of a secret conspiracy against the government of the Soviet Union, are captured without warning one night and are placed in separate cells, unable to communicate with each other. The police offer McAlpin a deal. If she will "defect" from her pledge of comradeship with Morris and testify against him, she will be set free. If at the same time, however, Morris also defects from the pledge, both will be given heavy sentences at hard labor on the damning evidence that each gives of the other's nefarious activities. If McAlpin refuses to defect—that is, continues to cooperate with the pledge—and Morris defects, then McAlpin will be shot: She does not turn state's evidence and will therefore not make amends for her crimes against the state. If McAlpin "cooperates" and Morris does too, neither testifying against the other, then both are given light sentences, for even under the somewhat elastic constitutional limits on the Soviet state it needs at least some evidence to convict for heavy sentences.

The only way that you or the prisoners can keep matters straight is to draw up a little two-by-two table of options (see Figure 21.6). McAlpin's choice to defect or cooperate with Morris is given on the left, as the two rows; Morris's choice is given at the top as the two columns. The results of the interaction for McAlpin (upper left in each box) and Morris are shown.

Looking at this table in her dreary prison cell, McAlpin faces a most distressing

Figure 21.6
In the Prisoner's Dilemma It is Best to Defect Under Any Assumption About What the Other Party Will Do: Table of Payoffs to McAlpin and Morris

A prisoner's dilemma is a matrix of payoffs from cooperation and defection that make it always desirable to defect. McAlpin's payoff is the upper left of each box, Morris's the lower right. No matter what Morris does, McAlpin will do best by choosing her Defects option. For example, if Morris defects, then McAlpin avoids death by choosing to defect as well.

	Morris	
Cooperates	Cooperates	Defects
	Light Sentence Morris Gets Light Sentence	Death Morris goes Free
McAlpin	Free	Heavy Sentence Morris Gets Heavy Sentence
Defects	Morris Dies	

dilemma. No matter what her former comrade Morris does, McAlpin is selfishly better off if she defects, that is, if she turns state's evidence. Look at the table again. If Morris cooperates, McAlpin is presented with the Morris Cooperates column. Since McAlpin obviously prefers being Free to getting a Light Sentence, she will (if selfishly motivated) choose to defect. On the other hand, if Morris defects, McAlpin is presented with the Morris Defects column. Since she prefers a Heavy Sentence to Death, she will again choose to defect. No matter what the other person does, then, a selfishly motivated McAlpin will defect. But the police offer the same deal to Morris, and he too has an incentive to defect no matter what the other does (his payoffs appear in the lower right-hand part of each rectangle). Both therefore defect. Both convict each other. Both end up with heavy prison sentences. By contrast, had they been able to keep faith, they would have received light sentences. The lower right-hand outcome is their equilibrium *even though it is by no means the best for them*.

Cartels Pose the Prisoner's Dilemma to Their Members

Although better than Death, the equilibrium is of course worse than the Light Sentence that would result if both kept mum. If they were both devoted unselfishly to the cause and could each therefore be sure that the other would not defect, then both would cooperate, and the prisoner's dilemma would be avoided. It would also be avoided if they could communicate during the negotiations or could make legally binding contracts to cooperate or could punish the other in a subsequent play of the game for defecting on this play.

All these have analogies in the case of a taxi cartel. If executives of the taxi companies simply believed that it was rotten behavior to cheat on the price agreement, they would be in effect devoted unselfishly to the cause. The cause would be stable. So too if they could take cheaters to court, as any cartel could in fact do under German law in the late nineteenth century and as taxi cartels in most American cities can do nowadays, or if they could punish defectors by fines or violence. The cartel would again be stable.

The prisoner's dilemma highlights the forces pulling the cartel in the other direction, toward collapse. The analogy of the political prisoners to the taxi cartel is exact.

Q: The marginal cost curve of a typical member of a taxi cartel of many firms is drawn in Figure 21.7, together with the single firm's share of the market demand curve and the corresponding marginal revenue. Notice the four named dots.

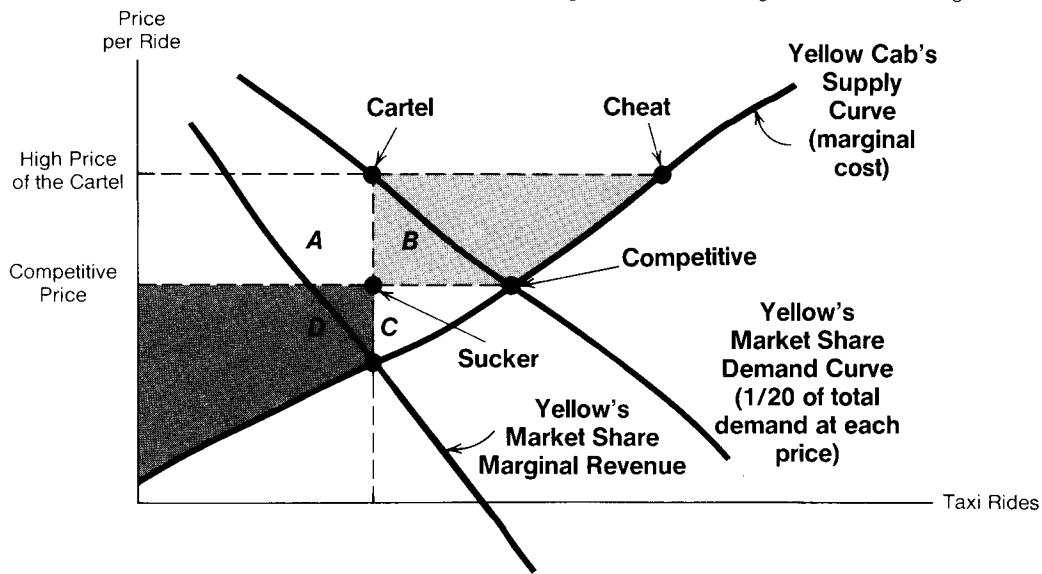
1. What profit does Yellow Cab make if it and all other companies cooperate in the cartel to set price and quantity at the Cartel point?
2. What profit does Yellow make if it is able to cheat, supposing that the other companies continue to cooperate with the cartel?
3. What profit does it make if everyone including Yellow cheats?
4. What profit does it make if everyone cheats but Yellow does not?

A: 1. Yellow gets the area $A + D$ in profit, namely, the area that results from maximizing industry profit, setting output so that marginal cost equals marginal revenue.

2. Yellow gets the High Price of the cartel but does not suffer the low cartel quantity if it moves to point Cheat. By shading its price a little, or by putting more than the allotted number of cabs on the street, it can do better than at Cartel. It gets the whole area $A + B + C + D$ in profit. In other words, the triangular area $B + C$ is *the incentive to cheat*. Notice that it will be larger the larger is the elasticity of the marginal cost curve. That is, if it is very easy (cheap) for Yellow to expand its output, the incentive to do so will be great.
3. If everyone cheats on the cartel, the equilibrium

Figure 21.7
Any Cartel Faces a Prisoner's Dilemma

The cheater in a cartel agreement can get the High Price of the cartel yet not pay the cost in sacrificed output: The cheater can produce all the way out to Cheat, making much more profit (namely, areas *B* and *C* in addition to *A* and *D*). But when every person is in the cartel thinks this way, as each will, the cartel is not effective, the price falls (since the market is glutted with output in excess of the output that would maintain the high cartel price), and each firm ends up at the Competitive point. Only a sucker would continue to honor the agreement once the agreement was no longer effective.



is of course Competitive, namely, the point of horizontal summation of the supply curves. The corresponding profit is *D* + *C*. Since *C* is less than *A*, each firm does worse under competition than under monopoly.

4. If everyone cheats except Yellow, Yellow arrives at point Sucker: It loyally maintains the low cartel quan-

ity but gets for its trouble only the Competitive Price (or only slightly above the Competitive Price, considering that its decision to keep a low output affects only part of the industry's output). The profit is only area *D*.

Now, to complete the analogy with the prisoner's dilemma.

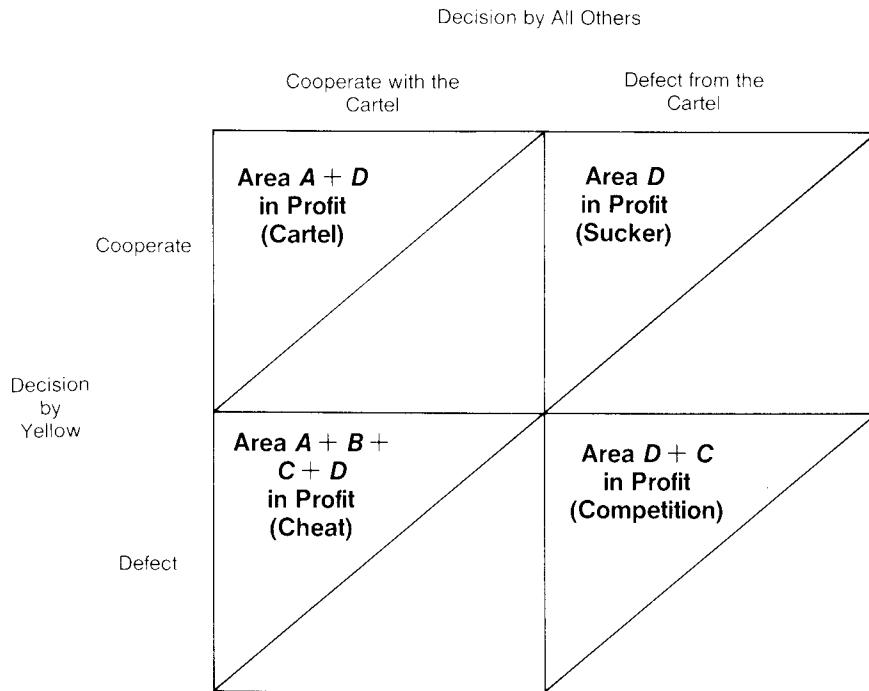
Q: Fill in the two-by-two table of results from Yellow's perspective for a game of cooperation or defection played between Yellow and all other taxi companies. Discuss the correspondence of each outcome to those in the prisoner's dilemma.

A: The table is given in Figure 21.8. Compare the table with the earlier diagram of the cartel and the earlier

table of the prisoner's dilemma. The Cartel cell corresponds to the Light Sentence, Cheat to Free, Sucker to Death, and Heavy Sentence to Competition. The point is that the incentive to cheat lures the members of a cartel into a heavy sentence of competition.

Figure 21.8
The Table of Payoffs to Yellow Cab Is a Prisoner's Dilemma

As with the prisoner's dilemma, the prospect of selfish gain tempts the firm to cheat on the cartel, regardless of whether the other members of the cartel decide to cheat or not. The result is to move the industry into the lower right-hand corner, that is, to universal defection. The Heavy Sentence of the prisoner's dilemma is, in the case of a collapsed cartel, the profits of Competition, lower than the profits obtainable from a successful cartel.



The Prisoner's Dilemma Applies to the State

The idea of the prisoner's dilemma is rich in applications. The most important application is to the leading question in political philosophy from Aristotle to the present, namely, the nature of government. One economic analysis of a government such as the federal government of the United States, for example, is that it is a cartel, which can achieve benefits for its members that they would not be able to achieve individually. Like any cartel, however, its members have an incentive to cheat. That is, each citizen has a selfish incentive to take the benefits of citizenship, such as the provision of lighthouses, public roads, and, above all, protection from the attacks of other governments, but to pay none of the taxes. If taxes were not compulsory, most governments would collapse. If it is thought that the services of the government are worth the trouble, therefore, taxes must be compulsory, and we would even wish them to be so. We

would voluntarily agree to subject ourselves to the government's power to fine or imprison nonpayers. The prison overcomes the prisoner's dilemma.⁶

A large group lacking the power to imprison or in other ways discipline defectors faces great difficulties in collective action.

Q: You and 50 of your most intimate friends have joined together to buy a building in which to live communally. But as a poet once said, it takes a heap o' livin to make a house a home. Yes (answered another poet) and a heap o' cookin', washin', cleanin', repairin', and above all, payin'. You hold a meetin' of the group on how to accomplish these things. "Let's not be authoritarian about it," says Elyce Rotella at the meeting. "I think we can depend on everyone's public spirit and good sense to get these things done." What is your reply to Elyce?

Another example is the following.

Q: "Bare-faced covetousness was the moving spirit of civilization from its first dawn to the present day, wealth, and again wealth, and for the third time wealth; wealth, not of society, but of the puny individual was its only and final aim." In particular, "the bourgeoisie . . . has left remaining no other nexus between man and man than naked self-interest." *True or false:* Even if one believed these assertions by Karl Marx and Friedrich Engels to be true, it would not follow from them that the bourgeoisie or any other class would be motivated by covetousness or other self-interest to behave in its *class* interest, because the cooperation required for *class* action suffers from the prisoner's dilemma.

No bourgeois motivated solely by rational covetousness will bother to contribute his efforts to the common good. The point is identical to the one on the irrationality of voting made earlier: No single vote matters for the outcome, and therefore no person motivated solely by the desire to change the outcome will bother to vote. Likewise, no single supplier bothers to manipulate the amounts she supplies to change the price she gets, because the output of no

⁶ One alternative to the view that the government is a cartel that its citizens have joined voluntarily is that the government is a band of robbers into whose clutches its citizens have fallen. The analysis reduces then to the economics of robbery instead of the economics of free exchange. On April 15 in the United States this view is surprisingly popular. On other dates it is taken seriously by anarchists (not bomb throwers, but *anarchos*, from the Greek, meaning "without a leader," that is, the notion that we can do quite well, thank you, without politicians endowed with special powers of compulsion).

⁷ Mancur Olson, *The Logic of Collective Action: Public Goods and the Theory of Groups* (Cambridge, Mass.: Harvard University Press, 1965), pp. 104–105, the Marx-Engels quotations appear on p. 104. Olsen's book is brilliant and readable.

A: "The problem, Elyce, is a classic prisoner's dilemma. Some people will be less public spirited than others and will take a free ride on the efforts of others. They will defect, getting their share of the benefits from the commune without doing their share of the work. With no rules to punish slackers, it will be rational for all to slack off. And in this tendency to collapse in the absence of rules, our commune will be following the lead of many a utopian community in history."

single small suppliers matters for the market price. Perfect competition is a prisoner's dilemma into which, fortunately for society as a whole, many firms are led, as though by an invisible hand.

The Prisoner's Dilemma Applies to Many Large Groups

Q: People at a cocktail party want to be heard over the din of other people who are talking. Everyone generally ends up shouting. Show that, if they cooperated they could avoid shouting, but that the shouting is the equilibrium of a prisoner's dilemma.

A: The choices I alone have are to talk normally or to talk loudly. The results of my choice and the choices

The prisoner's dilemma—the incentive to free ride on the efforts of others and the collapse of cooperation that such free riding brings—is pervasive. A clear if trivial example is the Cocktail Party Problem.

of the group around me can be expressed in terms of the distinctness with which I am heard by the person to whom I am speaking relative to background noise. The results appear in Figure 21.9. No matter what others do I will shout, because in each case I raise my audibility. But everyone makes the same decision, with the result that none of us is heard very well and all of us are hoarse leaving the party.

Still another case is the annoying anxiety when you have to choose one of several lines to join at a college cashier's office, a grocery store, a bank, or an airline terminal: Have I picked the fastest line? The line you join always seems to be the slowest. In a very few colleges, a few grocery stores, some banks, and quite a few airline terminals, everyone is put into one common line, the person at the head of it being always the next to be served. The advantage of the system is that the variability of the waiting time is cut while the average time is unchanged, for the following reason. The average time from arrival to service will in fact be the same whether there is a line in front of every server or a common line for everyone. In the many-line arrangement, the student or shopper or whoever with a complex and lengthy case to be served will choke up one line, significantly delaying the people behind her but not the people in other lines. In the single-line arrangement, however, the delay will be spread in a small amount over all the people waiting. With risk-averse people it will always be better to spread the cost of the unusually lengthy case over all the people than to burden more heavily the few who happen to have the misfortune

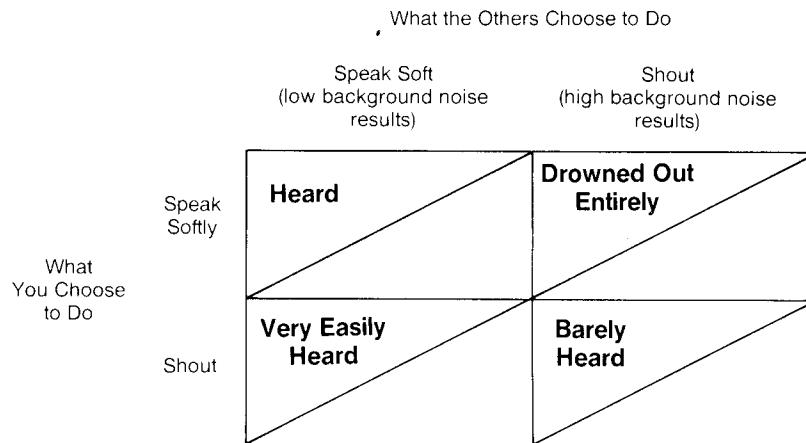


Figure 21.9
Why You Shout at Large Cocktail Parties

Again, if you defect from the social agreement to speak softly, you will be better off (will be heard more easily) no matter what the rest of the party does. But everyone faces the same incentives, and the result is universal shouting.

to join a line at cage 12 or desk 5. Therefore, the one common line will always be best (ignoring administrative costs): same average, lower variability.

T or F: That such single common lines feeding into many servers never arise without compulsion (such as ropes and guards) shows that all people are, contrary to economics, irrational.

A: The single common line without compulsion is the cooperative solution to a prisoner's dilemma. Anyone not at the head of the common line has an incentive

to defect, jumping to the first unoccupied server, and even to join a server's line if it is relatively short. Because of the defections the cooperator who waits her turn in the common line never gets served. The common line breaks down or, rather, never forms in the first place. That is, the assertion is false. That a common line does not form shows that people are indeed rational, too rational for their own good.

**Solutions to the
Prisoner's Dilemma
in Large Groups:
Punishment**

Any solution to the prisoner's dilemma requires some system of punishment for defection. The guards in the bank can enforce cooperation with the single line by threatening embarrassment, emotional distress, or, ultimately, physical violence against those who do not follow the rule. The Internal Revenue Service can enforce cooperation with paying taxes for public goods by threatening fines or, ultimately, physical violence against those who do not pay. A cartel such as the American Medical Association can enforce cooperation with cartel policies that are thinly disguised as medical ethics by threatening to take away a doctor's license or her right to use a hospital. In some cases the cartel authority can offer rewards for cooperation rather than penalties for defection: The *Journal of the American Medical Association* comes only to members, as do other private goods such as malpractice insurance. Clearly, this amounts to the same thing as punishment, the threat being to cancel one's subscription or one's insurance if one does not cooperate.⁸ The purpose of either device is to alter the payoff table facing any member.

Q: A union is being organized in the Lazonick Cotton Mill. Without the union the wages are \$5 an hour, with full cooperation among the workers they would be \$7 an hour. To be effective the union must collect \$0.50 an hour in dues, leaving the cooperating (that is, dues-paying) worker in a unionized mill with \$6.50 an hour.

1. Write down the payoff table facing Gary Walton, an oppressed worker in Lazonick's mill, under the possibilities of cooperation or defection by him and by all others.

2. Suppose that the union can punish Walton for not paying his dues, by ruining his social life or his knees. Show the change a certain dollar value of ruined

social life produces in Walton's payoff table. What now is the equilibrium of the game?

A: The answers to the two parts can be put into one table, with the penalty for defection worth, say, \$2 per hour (see Figure 21.10). Without the penalty for defection the *underlined* figures are the payoffs, and Walton has the usual incentive to defect regardless of what all others do. But the penalty of \$2 per hour changes the payoffs in the bottom row, changing his optimal strategy from defection to cooperation. For example, the payoff if he defects and others cooperate is reduced from \$7 to \$5 an hour, making defection a bad idea.

**Solutions to the
Prisoner's Dilemma
for Small Groups:
Shame and Self-
interest**

Making defection a bad idea is the social glue of any group. For a very small group, however, there is often no need for explicit, formal penalties for defection. The shame of being caught shirking is enough. In fact, the smaller the group, the larger the chance of being caught and therefore being shamed or punished. One secret price cutter in a cartel of 50 firms that are limiting steel output is

⁸ Ibid., Chapter VI, "The 'By-Product' and 'Special Interest' Theories."

Figure 21.10**You Gotta Go Down and Join the Union: The Punishment-Reward Solution to the Prisoner's Dilemma**

The ability to impose penalties for defection changes the matrix of payoffs that a single person faces and can change it enough to make it worth his or her while to cooperate rather than defect. On such arrangements are trade unions, political states, and many other social arrangements built.

The Choice of All Others
(results in wages per hour)

		Cooperate	Defect
		\$7.00 per hour – \$0.50 in dues = <u>\$6.50</u> per hour	\$5.00 per hour – \$0.50 in dues = <u>\$4.50</u> per hour
Walton's Choice	Cooperate (pay dues)	\$7.00 per hour – \$2.00 penalty = <u>\$5.00</u> per hour net	\$5.00 per hour – \$2.00 penalty = <u>\$3.00</u> per hour net
	Defect (do not pay dues)		

difficult to catch; one in a cartel of a few firms raising airfares between Boston and Washington is easy to catch. The one cartel will be unstable, the other stable. On such considerations of ease of detecting cheaters one can build a theory of oligopoly.⁹

If one single firm is so large a supplier to a market that the fate of the cartel hangs on its participation, then the crime of cheating brings its own punishment. No shame or fines or execution imposed by other members is necessary. The cartel is self-enforcing. The best example is the position of Saudi Arabia within the international oil cartel. Saudi Arabia produces such a large share of the world's supply of oil that its isolated decision of how much to produce significantly influences the world price of oil. If it produces little oil, the world supply is small and the price is high. If it produces much, the world supply is large and the price is low. That is, its share is so large that it faces by itself a downward-

⁹ Stigler, *The Organization of Industry*, Chapter 5, pp. 39–63.

sloping demand curve. By comparison the other participants, such as Venezuela, Nigeria, and Iraq, are tiny: They face the price Saudi Arabia arranges.

Q: What, then, does the payoff matrix to Saudi Arabia look like, interpreting “cooperation” as “restraining one’s production of oil” and “defection” as “letting one’s production rip, seizing the opportunity offered by the high price on a large instead of a restrained production”?

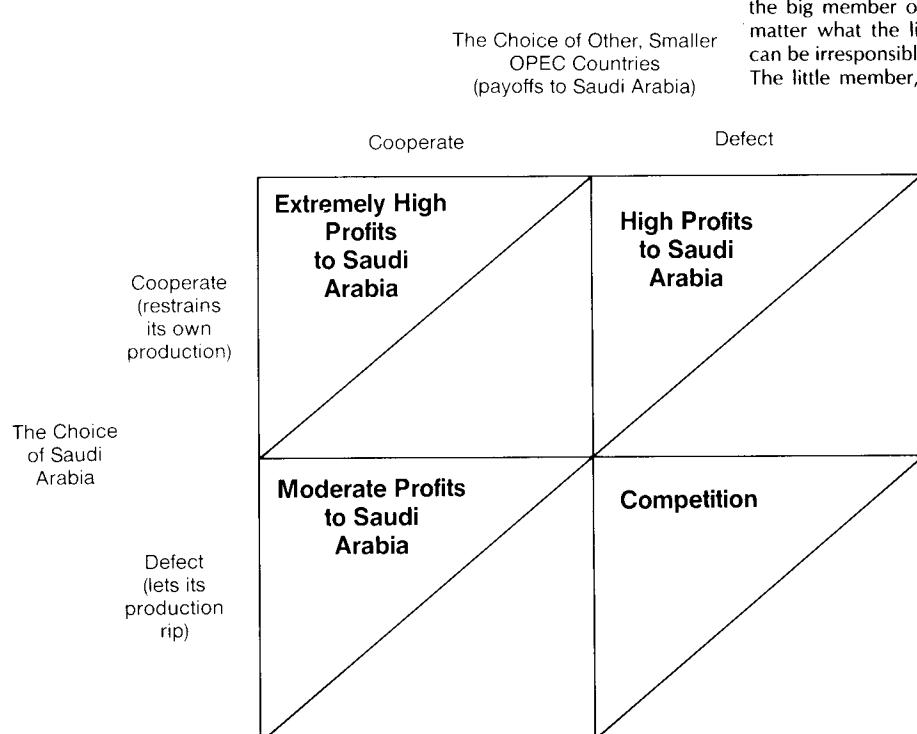
A: The payoff matrix is so to speak the opposite of the prisoner’s dilemma (see Figure 21.11). No matter what

the other countries do, Saudi Arabia finds that it is in its own interest to cooperate, that is, restrain its own production. So big is Saudi Arabia that the oil cartel’s high price stands or falls as Saudi Arabia cooperates or defects.

The situation is a common one. The United States is so important in world affairs that various collective undertakings—the United Nations, for example, or NATO—would fall apart if the United States did not support them. It is therefore in the self-interest of the United States to support them, even on unfavorable terms. A small country such as Sweden or France has the luxury

Figure 21.11
Little Members of a Cartel Exploit the Big Members

Because the cartel will collapse without its participation, the big member of a cartel must abide by its rules no matter what the little members do. The little member can be irresponsible without suffering the consequences. The little member, therefore, exploits the big member.



of refusing to pay for collective undertakings in the sure knowledge that the United States will pay anyway. As Mancur Olson put it,

Once a smaller member has the amount of the collective good he gets free from the largest member, he [sometimes] has more than he would have purchased for himself, and has no incentive to obtain any of the collective good at his own expense. In small groups with common interests there is accordingly *a surprising tendency for the "exploitation" of the great by the small.*¹⁰

The Theory of Games

The prisoner's dilemma is a special case of a general *theory of games*.¹¹ The theory discusses the situation of intelligent people facing other intelligent people in games of tic-tac-toe, poker, chess, cartels, exchanges, collective bargaining, price wars, business mergers, politics, extortions, kidnappings, and wars. Invented in 1928 by the famous mathematician John von Neumann and brought to the attention of economists in 1944 by von Neumann and Oskar Morgenstern in their astonishing *Theory of Games and Economic Behavior*, the theory gave early promise of solving the problem of bargaining among small groups, such as oligopolists. Like monopolistic competition and other approaches to the problem, it has not fulfilled the promise.

Cultivated for its own sake as a metaphor of social life, however, it must be judged a great intellectual success. At present its main use in general economics is in fact metaphorical. To say that the formation of a cartel is "just like" a prisoner's dilemma game or that a nuclear arms race between the Soviet Union and the United States is "just like" a two-person negative-sum game is to state the essence of the situations with persuasive elegance.

The very notion that we are "playing games" with other people is enlightening. So, too, are the notions of the negative, positive, or zero sumness of the game. In a zero-sum game my loss is your gain, as in the neighborhood poker game or in the distributing of the gains from trade as viewed by medieval towns. Gary's winnings are John's losses; what Venice gained in the trade with the East in the Middle Ages Genoa lost, or so the Venetians and Genoese believed. It is apparent that any constant sum of spoils (or negative spoils, damages) will give the same results. That is, as long as Venice and Genoa are fighting over a fixed pie, it does not matter whether or not the total size of the pie is called zero (Venice gets what Genoa loses) or 100 (Venice gets 80, Genoa 20, both gaining over zero). The alternative to such a constant-sum game is a variable sum, in which the size of the pie to be divided does vary with how the players act. The sum of the payoffs from the arms race, for example, depends on which solutions are chosen. If the bombs are not used, each country loses only the cost of making the bombs; if they are used, each loses its entire population. A more cheerful example, and the focus of much attention in the theory, is exchange. Exchange is "positive sum"; that is, both parties gain or, at worst, do

¹⁰ Olson, *The Logic of Collective Action*, p. 35 (italics in original).

¹¹ A brief idea of the theory can be gotten from William J. Baumol, *Economic Theory and Operations Analysis*, 4th ed. (Englewood Cliffs, N.J.: Prentice-Hall, 1977), Chapter 18. A more leisurely and thorough review is Morton D. Davis, *Game Theory: A Nontechnical Introduction* (New York: Basic Books, 1970). The original book of 1944 by von Neumann and Morgenstern mentioned in the text following (Princeton, N.J.: Princeton University Press, 1947) is highly technical on the whole, but also highly readable in places.

not lose. If two parties to bargaining over the exchange of steel pipes for natural gas fail to reach any agreement because they disagree over the price, the mutual benefit does not materialize. That is, the size of the pie in total varies with the bargaining strategies of the parties; the game is variable-sum.

As one might expect, the variability introduces complications of threat and bluff that are not present in the simpler case. As was noted, the theory does not literally "solve" the problem of bargaining games. Game theory merely provides the economist with a rich harvest of metaphor: coalitions, the core (mutual benefit), imputations (the prices agreed to), side payments (bribes), maximum strategy (avoiding the worst that people or nature can do to you), saddle points (when such avoidance implies the same strategy for both players), mixed strategies (flipping coins to keep one's behavior from being predictable), and other wonders. It does not, alas, solve the oligopoly problem.

Summary

The theory of games is the ultimate response to the principle of outsmarting. In reaction to the unattractiveness of supposing oligopolists to be outsmarted easily, economists have developed in it a virtual theory of outsmarting. The goal of a set of competing oligopolists is ultimately to eliminate competition. They play a game of cooperation and defection that may or may not have monopoly as its outcome. The approach is more attractive to economists than are Cournotesque approaches because, as George Stigler put it,

A satisfactory theory of oligopoly cannot begin with assumptions concerning the way in which each firm views its interdependence with its rivals. If we adhere to the traditional theory of profit-maximizing enterprises, then behavior is no longer something to be assumed but rather something to be deduced.¹²

As usual, however, the game theoretic approach to the theory of monopoly is more useful in applications than in reaching a satisfactory theory of oligopoly. The uses of the prisoner's dilemma alone justify the journey: The prisoner's dilemma is the very model of the social problem. Like a cocktail party or a line at a bank, a cartel is a little society facing the problem of defection from a mutually advantageous arrangement. The single member of an oil cartel ordered to cut back its output for the common good, like the single subject of a government ordered to pay taxes for the common good, has an incentive to cheat, to receive the benefit of a high cartel price or a wealthy society without paying the price in lower output or higher taxes.

The society must either find ways to punish free riders or accept collapse. The history of cartels, unions, and not a few governments is written in the algebra of the prisoner's dilemma.

EXERCISES FOR SECTION 21.3

1. a. Give Death, Free, Light Sentence, and Heavy Sentence values in dollars to McAlpin and Morris in such a way that the resulting matrix is a prisoner's dilemma. (*Hint:* Look at Figure 21.10.)
b. If Light Sentence is called \$a, Free \$b, Death \$c, and Heavy Sentence \$d, what relations must hold among a, b, c, and d for the matrix to be a prisoner's dilemma?

¹² Stigler, *The Organization of Industry*, p. 390.

2. In the style of Figure 21.10, impose a single fine for defection that changes the money values of the defection row enough to eliminate the prisoner's dilemma. What is the equilibrium now? Explain.

3. Describe the prisoner's dilemma in the following situations:

a. Overfishing of the Grand Banks fishing ground by trawlers from many countries, to the point where the fish cannot breed and the Banks are ruined for all.

b. Failures to vote, to the point where presidents are elected by small minorities of the population.

PROBLEMS FOR SECTION 21.3

1. If Morris and McAlpin, mentioned in the text, faced the prospect of playing the game many times again after their first play (never mind how they would recover from the prison sentences imposed by the solution), would they behave in the same, uncooperative way? For example, if Morris defected in the first game, what would McAlpin do in the second and third game? Make the analogy with ordinary social games, such as giving the floor and one's attention in a conversation. If you do not permit others to speak, and yawn and look away when they finally do speak, what will happen to you in the next playing of the conversation game?

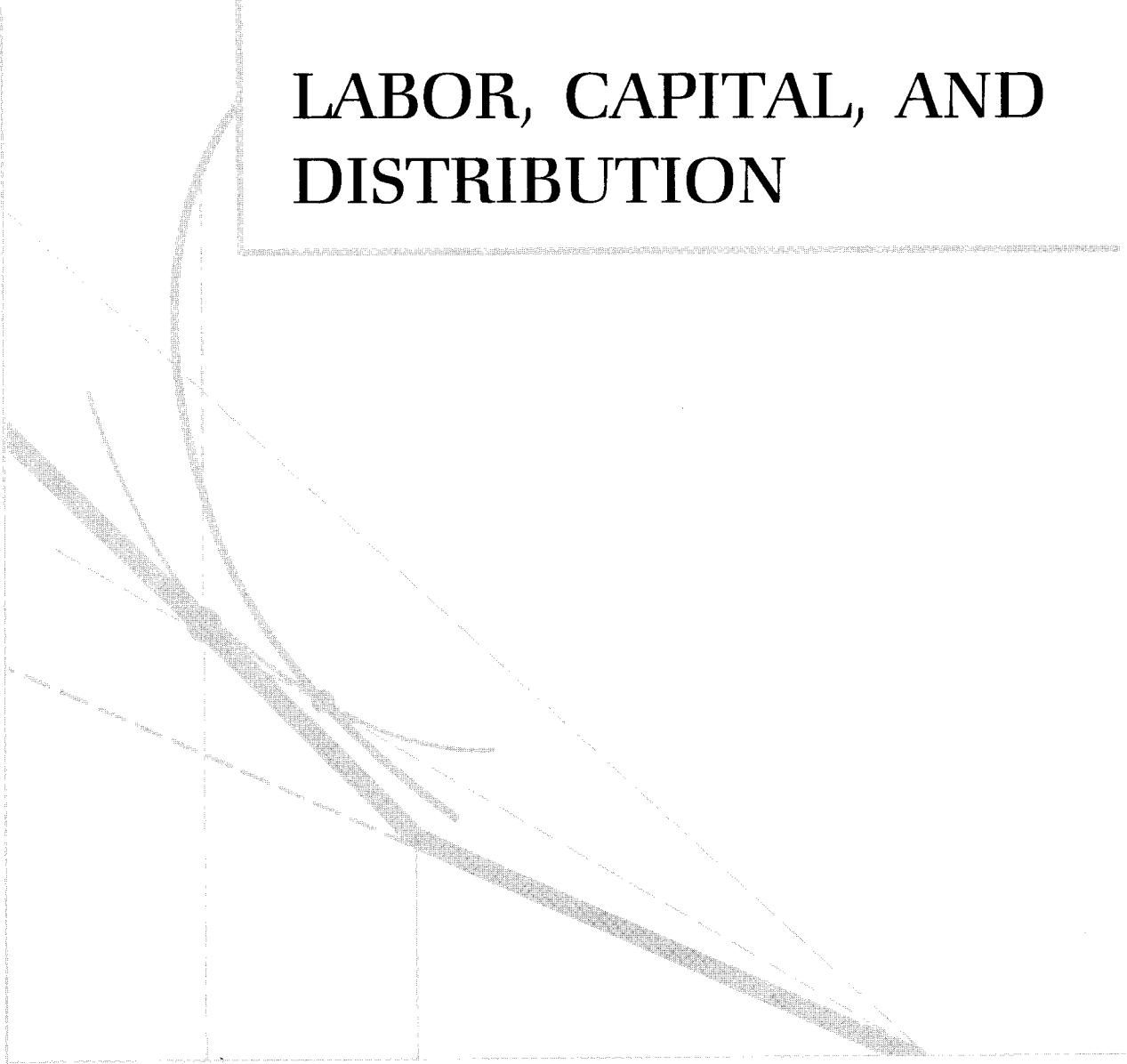
2. Certain social customs are "mere conventions," that is, good because people agree on them. Driving on the right-hand side of the road is a good example: We could equally well all drive on the left-hand side, and if we were British we would happily do so, so long as we all did it. What is the payoff matrix for thee and me for driving on the right side as against the left?

True or False

- 3. Since the money value of the hurt of the 10 million American consumers of automobiles from a protective tariff on automobiles is greater than the money value of the help to the four American producers of automobiles (or to 4 million stockholders represented by four companies), the proposal for a tariff will lose in the expensive political competition for votes in Congress.
- 4. A perfectly enforced cartel always earns more profits for the typical firm than does a competitive industry. That is, in terms of Figure 21.7, the area *A* is always bigger than area *C*.
- 5. That few people cast ballots, thereby endangering democracy for the gain of a half hour of leisure, is a prisoner's dilemma.
- 6. That a communal bowl of popcorn is eaten much faster than is a set of individual bowls is a prisoner's dilemma.
- 7. That a common property resource, such as the local public park, is overused and filthy from trash is a prisoner's dilemma.
- 8. The conventions of diplomacy whereby, for example, embassies are inviolable, can be broken with impunity by little countries (for example, Iran) but not by big countries (for example, the Soviet Union, the United States).

VI

LABOR, CAPITAL, AND DISTRIBUTION



22

Marginal Productivity and the Demand for Labor: The Fundamentals

22.1 Labor as a Commodity

What to Read For In what market do consumers earn the income they spend on goods? What is derived demand? Do firms hire labor for what the labor can do, or for who the laborer is? How does one portray racial discrimination in the labor market? Does the minimum wage make people better off?

Factors of Production Are Owned by Households

It is now time to close the economy—"close" meaning to finish the Wheel of Wealth. The Wheel, you will recall, arrayed consumers on one side and firms on the other, as on the left-hand side of Figure 22.1. The top flows are the markets of the goods that have been the explicit subject of the analysis so far: corn, natural gas, housing, books, college education, and so forth. The bottom flow is the subject of the next five chapters, the market for the services of factors of production such as labor, land, and capital.

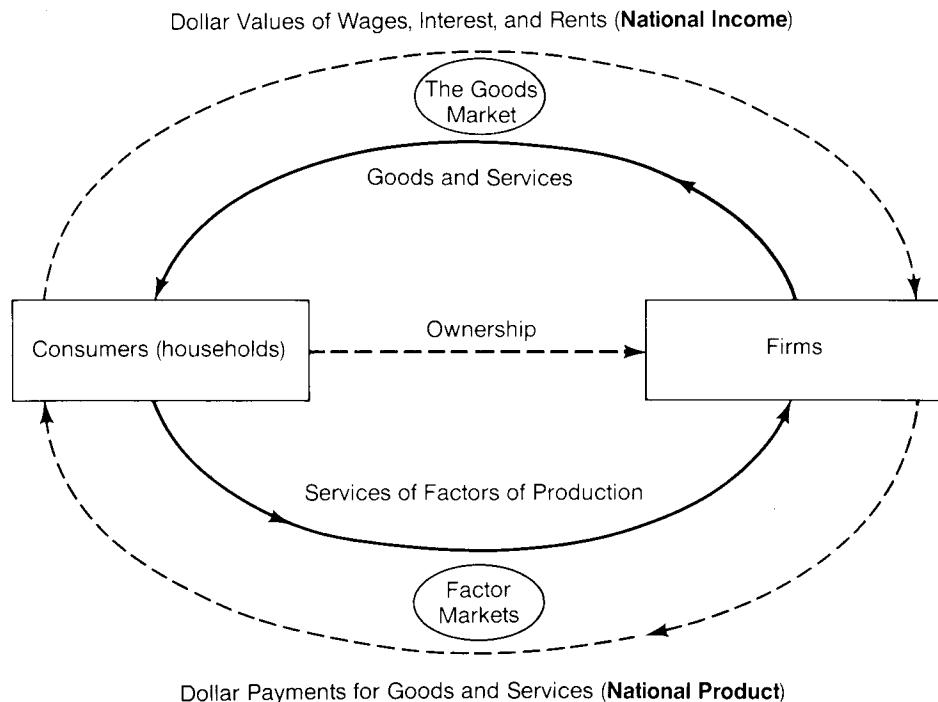
The factors of production are those basic inputs from which all commodities can be thought of as flowing. Look at your pencil. True, it is made of wood: no factor of production there. But look a little harder. Where did the wood come from? Well, it came ultimately from land, labor, and capital—the land of a forest combined with the labor of the lumberjack and the capital of the sawmill operator cooperated to produce a piece of wood, fashioned at later stages into a wooden pencil.

Analysis of this sort, in truth, finding in each commodity the "ultimate" factors of production that made it, is more persuasive than it has any right to be. After all, the sawmill's saw was itself produced by labor and land and capital; and the laborer is what he eats. Nothing is really "ultimate" or "basic." The Wheel of Wealth is indeed a wheel, with no beginning or end. Like salt or

*Alex Dv
of*

Figure 22.1
Factor Markets Complete the Economy

The solid lines represent the flow of goods and services from firms to consumers and the flow of factor services (labor, for instance) from consumers to firms. The dashed lines represent the money payments for these flows, the one being national product, the other national income. The next few chapters concern the lower, factor market.



sugar, however, the story of basic factors of production is not too bad if you do not take too much of it.

The spending on commodities that comes out of the households on one side of the Wheel must come into the households on the other side as income. That is, what factors of production earn is income, the income that is spent on commodities. The way in which the households earn their income, in other words, is to sell their services—services of their labor or their land or their capital—in a market in which they are sellers and the firms are buyers. The market is called the market for factors of production. It finishes the Wheel and, as was just said, closes the model. In a sense, then, by this step the analysis of the economy is finished.

The Law of One Price Applies to Labor Markets

The mere idea that the factors of production sell in a market is a boon to anyone attempting to measure the rewards to those factors. If the market is functioning well and the cost of moving from one part of the market to another is not very great, then the price of all factors of production of a particular sort

and quality must be the same. All land in a city must earn the same rent, once the special advantages of particularly favorable locations have been allowed for. All common laborers in the same country must earn the same wage. All machines of the same productivity must rent or sell for the same price. Factors of production, in other words, are to be treated like grains of wheat or tons of steel: as fungible.

T or F: The indexes of wages of building artisans in England over the past seven centuries are inadequate because they refer overwhelmingly to wages in southern England, especially in London and in a few cathedral towns.

A: More information is always better than less, to be sure. It would be nice to have wages from the North. But to suppose without further argument that a sample

based on the South is wrong is to suppose that England as a whole did not constitute, roughly, a single market for labor. Perhaps it did not. But that it did not needs to be shown, not merely assumed without evidence. And if it was a single market, little is gained by looking further into regional wages—except perhaps to confirm the very oneness of the market or to detect some persistent superiority of one region's wages over another's due to differences in the cost of living, say, or climate.

A New Idea: Derived Demand

Nothing so far is analytically new. The substantive novelty in the analysis of the demand for labor can be summarized in the phrase *derived demand*. The demand for labor is largely a derived demand, that is, a demand not *directly* by consumers but by firms who will then use the labor to satisfy the demands of consumers. In fact, the only reason for the firms to buy the labor is to satisfy the demands of consumers. If a certain laborer cannot produce for the firm enough consumer satisfaction to pay his wage, the firm simply does not hire the laborer. The demand for the laborer's services is derived from the demand for whatever the firm makes, such as popcorn, crackerjacks, or baseball.

Q: Pete Rose was a baseball player of high skill, a great pleaser of crowds, a star. *True or false:* Rose was, at \$800,000 a year, overpaid by the Philadelphia Phillies.

A: If someone pays Rose such a sum, that someone believes that Rose will earn the club an amount equal to the sum or more. The owner of the Phillies is not in the business for his health. The payment was mutually advantageous or it would not have been made. The owner was willing to pay and therefore did not "over-

pay." The only meaning of "overpay" is one that objected to Rose's being paid so much "merely" for hitting, throwing, and catching a baseball superbly well. But behind the willingness of the owner to pay Rose is the willingness of the fans to pay the owner. Would it make sense to object to 3 million separate deals between Pete Rose and his fans to put on an exhibition of his skills? No. The demand for Rose is a derived demand.

The point applies to all businesses and to "underpaying" as well as to "overpaying." Geoffrey Hellman wrote for the *New Yorker* magazine for a long time and had incessant quarrels with its editor, Harold Ross, about how little Ross paid a man of Hellman's seniority. Ross insisted that he paid what each piece of writing was worth:

You say that you have been here eighteen years and are not treated better than a good writer a couple of years out of college would be, so far as pay for individual articles is concerned. . . . My firm viewpoint is that we ought to pay what a piece is worth, regardless of age, race, color, creed, financial status or any other consideration. I don't know

how, in an enterprise of this sort, one in my position can take into consideration anything beyond the actual value of the things.¹

The point is that *employers are not to be viewed as having tastes for employees* in the way that consumers have tastes for apples or oranges. The steelworker makes steel that Bethlehem Steel turns around and sells to automakers. It is not the steelworker that makes the owners of the company happy directly but the indirect profit from the worker's work.

There will be cases that violate the point, of course. An important case in which the hired employee himself does directly enter the utility function of the owner, instead of merely as an instrument for making money, is the case of discrimination in hiring, in favor of native-born citizens, say, or against blacks.

Q: If employers, to a varying degree, have a distaste for hiring black workers rather than white workers at the same wage, then a rise in the proportion of blacks to whites in a labor market will be accompanied by a fall in the relative wages of blacks.

A: As the proportion rises, the blacks must face more and more discriminatory employers. All blacks must

be paid the same wages (assuming that blacks and whites are identical except for color), or else the low-wage person will undersell the high-wage one. So the black wage is determined by the wage differential that just compensates the most discriminatory employer hiring any black for hiring "the" marginal one. Therefore, true.²

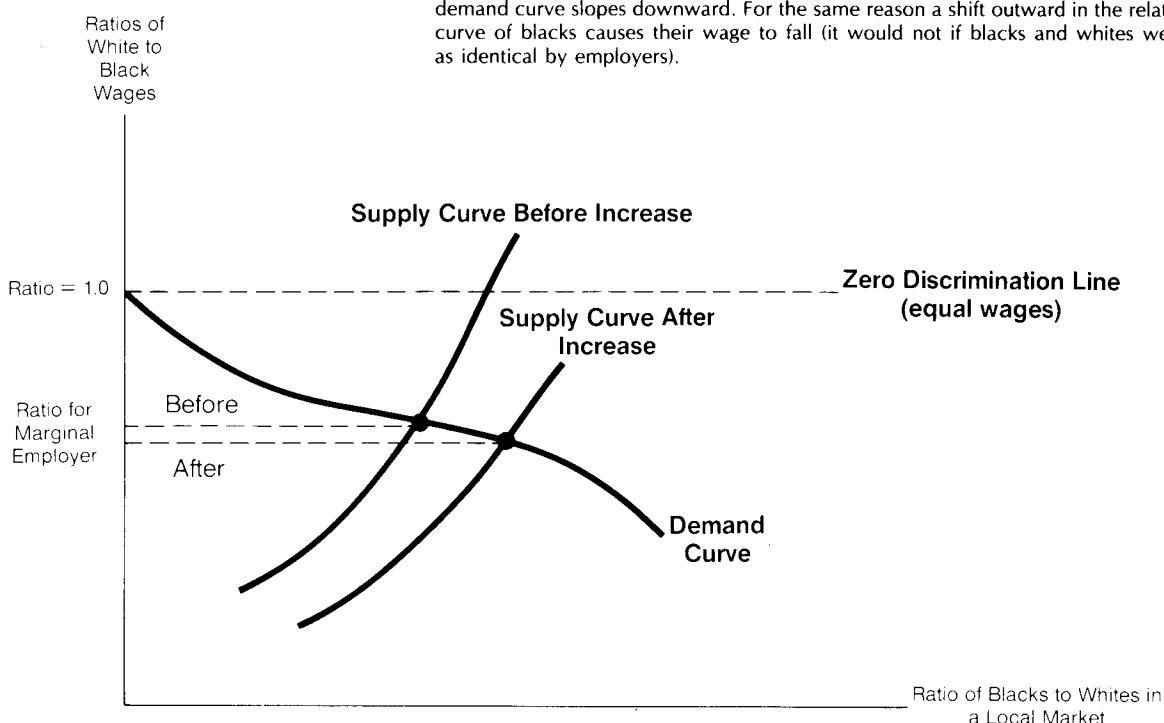
The diagram is in Figure 22.2. Notice that the number of firms with a given degree of discrimination is assumed constant—invariant, that is, to the wage paid. A perfectly functioning market would increase the elasticity of the curve, whatever its level, by driving out firms that discriminated a lot until all firms had the lowest discrimination (not necessarily zero) available among the class of potential employers. The point is that a firm that did not hire black workers—that indulged its taste—would pay more for labor than would a firm that took advantage of the lower wage for blacks. This applies across industries—relatively nondiscriminatory industries would have lower costs than they would in the absence of discrimination elsewhere. It is said that the radical shifts in the racial composition of the work force in the New York garment trades is a result of this mechanism.

As useful as the exception is for analyzing some situations, it must be realized that it is indeed an exception. Most employers are cold-blooded about hiring, within narrow limits. That does not mean they are fair in the ordinary sense of the word, merely that they are cold-bloodedly efficient, which has its own charm. If a secretary does not do his job well, he will lose it. If he is obviously incapable of performing the job, he will not get it in the first place. If he is competent but demands special treatment of an expensive sort, again he will not get it. The discipline of the market stands ready to punish an employer who is anything but calculating in this way. If he is the owner, he will be competed to bankruptcy by competitors who do watch the value they get from the people they hire. If he is a hired manager he himself will be fired if he makes a habit of indulging his preference for white Anglo-Saxon Protestants

¹ Quoted in Brendan Gill, *Here at the New Yorker* (New York: Random House, 1975), p. 360.

² See William Landes, "The Economics of Fair Employment Laws," *Journal of Political Economy* 76 (July–August 1968): 507–552, especially pp. 509–510; and Gary Becker, *The Economics of Discrimination*, 2nd ed. (Chicago: University of Chicago Press, 1971), pp. 5–85.

Figure 22.2
The Economics of Discrimination



as cooks when all the competition is using another sort of person that is cheaper to hire. In short, it is a premise of the theory of the firm that employers normally do not indulge their own tastes. They hire to satisfy their *customers'* tastes, the better to enrich themselves.

The Law of Demand Applies to Labor as a Commodity

Although employers do not normally have tastes in employees, they do normally exhibit the behavior that consumers (with tastes) exhibit in the face of varying prices, namely, downward-sloping demand. The precise reasons for the parallelism between the theory of the demand for households and the theory of demand for firms is taken up later. The common sense of the matter is persuasive. At a high price for secretaries an insurance company will ration out the secretaries to its executives in small numbers. At a low price every suboffice of the company will have two or three. There are not, in other words, fixed and limited "slots,"

and jobs are not "created." The jobs to be done by the insurance company are unlimited. It could always use another person to organize the files just a little bit better, to handle claims just a little bit faster, and (above all) to collect premiums just a little more promptly. But doing these things a "little bit" more will not be worthwhile if the price of the person to do them is high. Only the most important secretarial tasks should get done if the secretary's price is high. For this reason it will be rational for the company to be stingy with secretarial time when it is expensive, and generous with secretarial time when it is cheap. That is, the company has a downward-sloping demand curve for secretaries. And likewise for other people hired.

By the usual argument, if the demand curve for each company slopes downward, the demand curve for an entire industry probably slopes downward. And indeed as will be shown presently, the demand curve slopes downward for the entire economy as well.

Uses of the Demand for Labor

That the demand for labor is a curve, not a point, is a tremendously important fact. It undermines, for example, all manner of manpower studies, projections, "needs," "supplies," and other products of the official imagination eager to find in its vision of the future a reason to act or not to act. "Demand for Nurses Outruns Supply" screams the headline. Why, the economist replies, do the wages of nurses not increase? "Shortage of Engineers Developing" screams another. But will not the consequent rise in wages provide the remedy? It is simply wrong to speak in terms of "the" supply or "the" demand of nurses or engineers as though each were a number, such as 1 million. The demand (and as Chapter 25 will show, the supply also) is dependent on the wage, the wage itself being determined inside the very market being discussed.

Q: Before the founding of the state of Israel, the administrators of the British Mandate in Palestine prohibited more than a certain number of Jews from immigrating. The number was lower than the number who wished to immigrate, with terrible consequences. One reason given for restricting immigration was that more than a small number would flood the labor market and result in unemployment. That is, the British believed that there were just so many slots in the economy: To pour more people into the given slots would be irresponsible and foolish. Comment.

A: The belief is another example of the tragedy of Palestine, in this case a tragedy complicated by mistaken

economics. If jobs were slots, the British would have been right. But jobs are not slots. Jobs exist because at the going wage employers wish to hire people in that number. If there are more people, then by the usual workings of supply and demand the wage will fall to induce employers to take up more of the workers. If more immigrants came the wage would fall. Perhaps this is what the British truly feared: impoverishment of the existing population. In fact, Israel came into existence, Jews poured into Israel in large numbers, and unemployment turned out to be small and diminishing.

An example closer to home, if your home is the United States, is the minimum wage. According to the law most employers are not permitted to pay anyone less than some amount per hour.

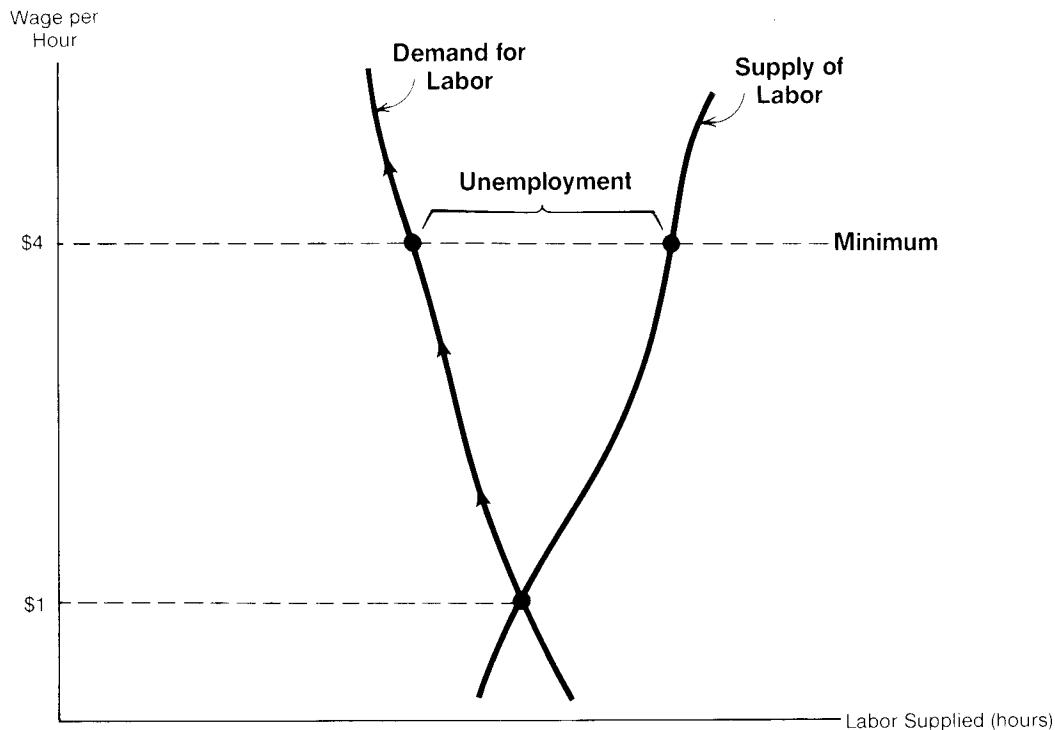
Q: Use supply and demand curves to begin an analysis of the effects of the minimum wage.

A: Simply treat the labor market like any other, such as wheat. If the price of wheat is supported by law, then there will be "unemployed" wheat, that is, a surplus of wheat produced over what is consumed. Likewise with labor. If the wage is supported by law, then there will be a surplus of labor, called "unemployment." The hour that is worth \$1 will not be used if an hour must by law be paid \$4. As usual in economic arguments, the point is made clearer by going to an extreme. Suppose that the minimum wage were \$50 an hour. In that case only doctors would have jobs. The rest of us would be prevented by the state from

engaging in an exchange with our employer. I may be willing to work for \$6 an hour; my employer may be willing to pay me the \$6 for the hour, but the state forbids the deal. The unemployed victim of such interference may be forgiven for wondering about the state's claim to be "protecting" the very worker who loses the job. In any event, a worker whose value to the employer is less than the minimum wage per hour simply does not get employed. If jobs were fixed slots, the demand curve for workers would be inelastic, and the low-value worker would be hired. But if jobs are *not* fixed slots (and they are not), the minimum wage forces the company up its demand curve and the low-value worker is not hired. The result is portrayed in Figure 22.3.

Figure 22.3
The Minimum Wage Causes Unemployment

At a wage allowed to reach \$1 an hour the supply of labor equals the demand. If the wage is held at \$4 an hour, however, the quantity supplied is higher and the quantity demanded is lower, leaving Unemployment.



The usual example of this effect is unemployment among teenagers. Present company excepted, teenagers are on the whole less reliable, prompt, responsible, strong, and skilled than adult workers. They are therefore, by the logic of derived demand, less useful to, say, a manufacturing company. They are not worthless, but worth less. A minimum wage would therefore be expected to cause disproportionate unemployment among teenagers. It does.

COMMENT

The assertion that the minimum wage causes unemployment and especially that it causes it among teenagers is controversial. It is fair to say that most economists believe the assertion.³ But some do not believe it, on various grounds. One line of counterargument is that the minimum wage encourages businesses to improve the machinery, buildings, materials, and other things workers work with to justify the higher wage paid. It is argued that an unskilled worker pushing a broom is not as valuable as the same worker pushing a \$1500 automatic broom-mop-waxer. There is an element of confusion in the argument, for it must be admitted that if it were good for the economy to invest in such automatic equipment before the minimum wage the economy might well have done so already. And if it were bad for the economy it is strange to argue that the investment thus induced artificially by the minimum wage is a good thing. Another and more persuasive line of argument, which will be taken up again in Section 25.2, is that working conditions will adjust to offset the higher wages. A slow janitor at \$2 an hour is no better bargain to the company than a fast one at, say, \$5. The company that hires the janitor will be willing to pay the higher price if it can specify that the janitors rush around at top speed. A faster pace of work or a greater degree of self-supervision or a higher standard of precision might all tend to compensate for the higher wage paid. This line of argument, however, has the same fault as the first. True as it may be, it does not necessarily justify the minimum wage. The mix of wages and conditions that existed before the state intervened in exchange presumably had some desirable feature, or else it would already have been bargained away. In the end the argument in favor of the minimum wage must come down to a simple distaste for the result of exchange in the absence of intervention. The feeling is that we simply should not tolerate anyone in a job so undignified that it was worth only \$2 an hour. Better that such people be supported by the rest of us, or even starve, than that they be required to work at such a job.

A less controversial example of the same point, familiar from Chapter 15, is the following.

T or F: That employers are required by law to pay half of the cost of social security taxes is irrelevant to the real burden of the tax.

A: The employers have a downward-sloping demand for labor. The demand price is what they are willing to pay for an additional hour of work in view of what

³ In J. R. Kearn, Clayne L. Pope, Gordon C. Whiting, and Larry T. Wimmer, "A Confusion of Economists," *American Economic Review* 69 (May 1979): 28-37, it was found that 68% of economists "generally agreed" with the proposition that "a minimum wage increases unemployment among young and unskilled workers," 22% agreed with some provisions (see the text following), and only 10% flatly disagreed.

X
Correct
this:
conditions

the hour can earn for them in profits. But if their share of the tax is, say, 5% of the wages they pay, they will get 5% less profit, and their willingness to pay falls by 5%. The 5% imposed "on" employers results in a fall in the demand for labor by employers as a group. But the fall in demand results in a fall in the wage, as the intersection of supply and demand moves back down along the supply curve. Fewer workers are employed at a lower wage than before the tax was imposed

"on" the employers. In short, true. The workers bear some of the burden. The 50:50 split of the tax between workers and employers is irrelevant for answering the question of who really bears the burden. What are relevant, as is always the case in such problems of supply and demand, are the elasticities of supply and demand and the total size of the tax, regardless of how it is apportioned legally among participants in the market.

Summary The leading point is that firms have demand curves. That such an extension of the law of demand should have so many applications will come as no surprise. On a formal level, the law of demand for firms serves to close the economy, giving households somewhere to earn the income that they spend in other markets on the products of firms. At a more substantive level, the result is that all the machinery of demand and supply curves applies to labor markets. The real surprises come from the proposition that firms buy labor in order to use it, not to admire it. That is, unlike the demand of households, the demand of firms for what they demand is "derived." The violation of this proposition gives a way of analyzing discrimination in the labor market. Its fulfilment gives a way of looking at all manner of ersatz economics that speaks of jobs as slots, such as labor "requirements" and the belief in the power of acts of Congress raising the wage to benefit working people. What remains is to understand the derivation of the firm's demand for labor more deeply.

PROBLEMS FOR SECTION 22.1

- 1. One interpretation of the arrangements in early nineteenth-century England for aiding the poor (the so-called "old Poor Law") is that they applied money from taxing farmers to subsidizing the employment of poor people in the hire of the same farmers. The chief objections to the arrangement were two. First, it made the workers less dependent on the farmer for their income and more dependent on the government (in this case the local government of the village), thus undermining the respect of workers for their employers. Second, it reduced the effort of the workers, by enriching them. Decide which if any of these two propositions is correct by analyzing the old Poor Law in terms of the supply and demand for agricultural labor. Assume that the *demand* curve was unaffected by the tax (an argument can be made that it was not).
 2. Private schools in Chicago often complain that the large increases in the wages of public school teachers increase the wages the private schools have to pay their teachers.
 - a. One conceivable method of hiring public school teachers in Chicago would be to announce a wage and to hire all qualified teachers who applied. If this were the method, would the complaint be true or false?
 - b. One interpretation of the actual method of hiring public school teachers is that the demand curve of the public schools is perfectly inelastic (they "need" N thousand), entry to employment is limited to N thousand (all being members of the teacher's union), the wage is held above the market clearing price, and increases in wages are granted in response to the political power and strike threats of the union. Under this interpretation is the complaint of the private schools true or false?
 - c. An alternative interpretation is that the union chooses a wage to demand this year,

the public school system accepts the wage, but decides that at the higher wage it needs fewer teachers (average class sizes rise, teachers of calculus and football are dropped, etc.). Now is the complaint true?

3. Bumper sticker: "30 Hours Work for 40 Hours Pay Means More Jobs." Comment.

True or False

4. In Japan, workers cannot be fired once they have been hired, and therefore in Japan a minimum wage law would not cause unemployment.

5. Effective minimum wages in trucking increase the demand for truck-driving schools, which are good substitutes for on-the-job training.

22.2 Marginal Productivity as the Demand for Labor by the Firm

What to Read For

Does the firm choose its level of output or its level of input? What is the total product of labor curve? The total revenue product of labor? The marginal revenue product of labor? The demand curve for labor? Does the amount a firm is willing to pay for labor depend on the total, average, or marginal product of labor? How does marginal productivity give a theory of the distribution of income among factors of production? What is the analogy with consumer's surplus? How does marginal productivity show the desirability of free immigration?

The Production Function Tells How Much Input Is Required for Profit-Maximizing Output

A deeper understanding of the demand for labor requires another use of that many-purpose idea, the production function. A lumber firm produces lumber to sell. The output of lumber, says the production function, depends on the inputs of forests, tractors, and lumberjacks. Since the output depends on the input, the firm can be viewed either as choosing output (then looking around for the inputs to produce it) or as choosing inputs (then looking around for a place to sell the output). Clearly, it does not matter which end of the stick one picks up. The following argument picks up the stick at the input end, asking how the firm decides how much of various inputs to demand in order to maximize profit. The problem is like the earlier way of looking at the firm, as deciding how much *output* to produce in order to maximize profit. One view suppresses some details about the firm's decision of how much to produce; the other suppresses some details about the firm's decision of how much labor and so forth to hire. The underlying problem—maximize profits—is the same.

The lumber firm employs forestland, tractor capital, and lumberjack labor. It has three decisions to make that amount to the "single" decision of how much lumber to produce and how (that is, with what recipe of inputs) to make it in order to make as much money as possible. Clearly it will make all three decisions about hiring the three factors of production in the "best" way it can. Equally clearly, these three decisions are connected. It simplifies the argument, however, to ignore for the moment that the firm has three decisions and to concentrate on the hiring of one input, which may be thought of as lumberjack

labor or as inputs in general. In other words, take for the moment the inputs of other factors of production—forestland and tractor capital—as fixed and given to the firm.

In such a case the output depends only on the amount of lumberjack input. That is, one can draw a diagram with output of wood on the vertical axis and inputs of lumberjack hours on the horizontal. The diagram, which appears in the top panel of Figure 22.4, is the same diagram as in Chapter 12 on the cost curves of a firm, but is here placed on its other side. Instead of relating output to the inputs necessary (and to costs), it relates the inputs to the output produced (and to revenue). An input of 400,000 hours of lumberjacks produces 20,000 thousand board-feet of lumber. The curve has *diminishing returns* to employment of the lumberjacks: Notice how it becomes flat eventually. The employment of still more lumberjacks—after a great many have been hired to work on one stretch of forest with a given set of tractors and power saws—bears little fruit in additional lumber. And notice the upward curvature at low levels of employment, indicating that at first each additional lumberjack bears great fruit in additional lumber.

The second panel merely multiplies the top one by the going price of lumber. The lumbering firm is supposed to be perfectly competitive in the market in which it sells lumber. That is, it is supposed to have no influence over the price it faces, which is to say that the price is given. Naturally, multiplying the total product curve by a price that does not vary is going to yield a curve with the same characteristic S shape as the total product curve has. The resulting curve is one of total revenue, because it is quantity produced multiplied by the price at which it is sold.

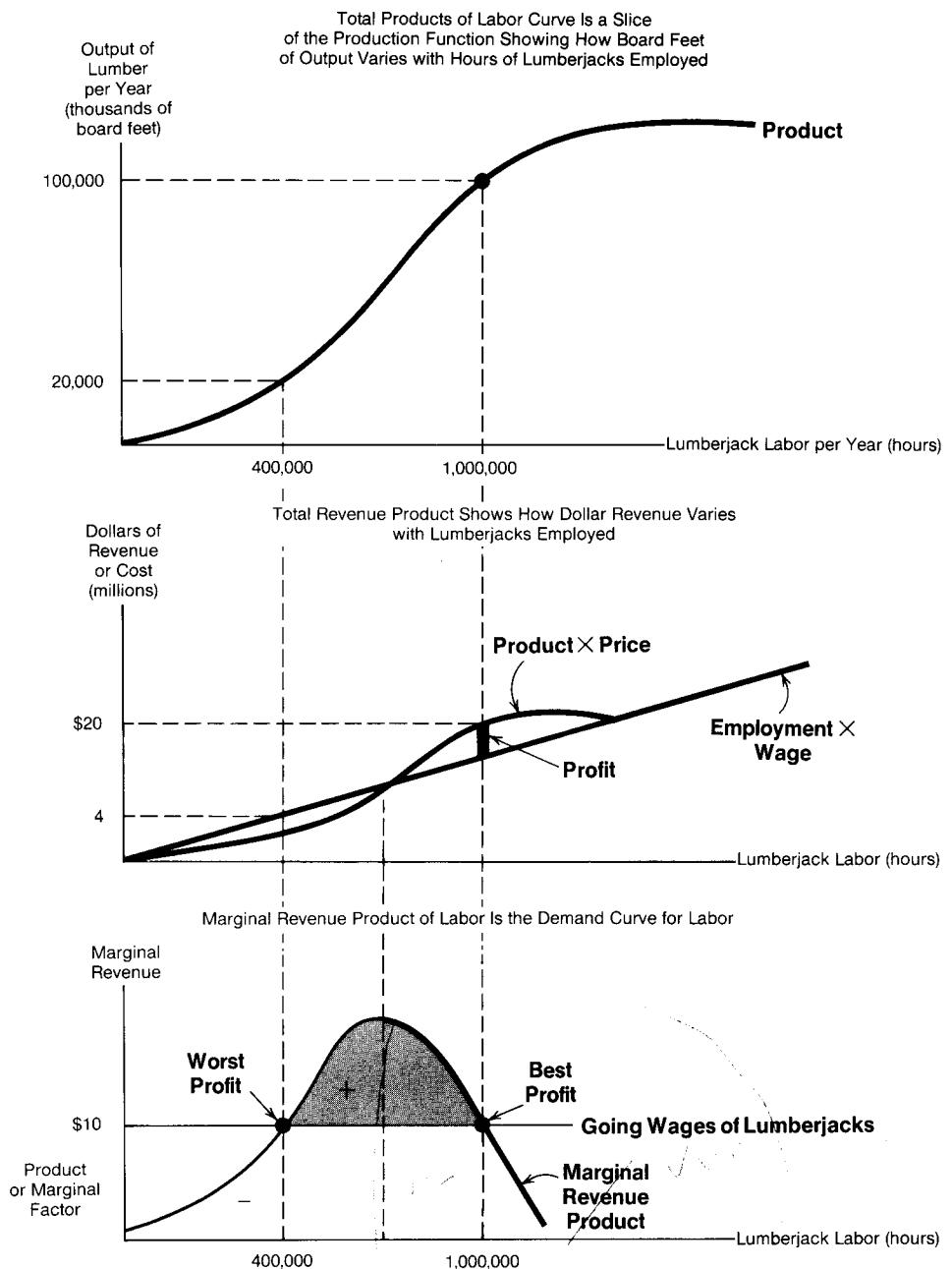
One can then draw on the logic of marginal curves to infer what the curve in the bottom panel must look like. It is the *marginal revenue product of labor*. The total revenue product is the middle panel, that is, the revenue produced from various different levels of employment of lumberjacks. The marginal revenue product—the change in dollar revenue with a one-unit increase in labor—is simply a plot of the *slopes* of the total revenue product. It too fits with the S shape of the total product curve: It rises at first, reaches a maximum, then falls.

The emphasized portion of the curve of marginal revenue product of labor is the downward-sloping portion. *It is the firm's demand curve for labor*, under the assumption ruling for the moment that other factors are fixed in amounts. The reason it is the demand curve may be seen by introducing a going wage for lumberjacks, say, \$10 per hour and asking how much labor the firm will hire. The answer will be the demand curve, for that is what a demand curve is: the amount of *X* demanded at various different prices of *X*. The marginal revenue product is evidently the *marginal benefit* to the firm from hiring an additional hour of labor (or at least it is if the firm does not, in the style of a racial discriminator, get pleasure or pain beyond profit from hiring the hour of labor). The wage, on the other hand, is the *marginal cost*. The rule of rational life suggests that the firm will do as well as it can by choosing the amount of hiring that brings marginal benefit and marginal cost into equality, namely, at the point Best Profit, hiring 1,000,000 hours of lumberjack labor to produce 100,000 thousand board-feet of lumber per year. The point *is* one of best profit. The area under the curve of marginal revenue product is the total revenue, because the curve is marginal to the total revenue curve. The area under the

Figure 22.4

The Demand Curve for Labor Comes from the Production Function and the Going Price of Lumber

The top panel is a slice of the production function. The middle panel multiplies the slice by the price of lumber to get the total revenue product. The bottom panel is the slope of it to be set against the marginal cost of labor (the going wage) to maximize profit.



going wage is the total cost, and the rest of the costs are by assumption fixed. It is obvious that the area between the marginal revenue and marginal cost reaches a maximum at Best Profit, the area marked with a plus offsetting the area marked with a minus. Any point on the rising portion of the marginal revenue product curve, such as Worst Profit, would be irrational. Therefore, the demand curve is the falling portion.

The Equation of Marginal Productivity

In the simplest case in which the firm is a perfect competitor in both the lumber market and in the lumberjack market, the algebra corresponding to the diagrams takes on a particularly simple and revealing form. Call the amount that an extra lumberjack hour produces the *marginal physical product of labor*, or MPP_L for short. Call the (unchanging) price that the firm gets for each unit of the extra amount of lumber P . Call the wage, which is the cost of the extra hour, w . Then the profit the firm makes on the additional hour will be the price times the marginal physical product (often called the "value of the marginal product") minus the wage: Hiring another hour causes the output and therefore the revenue and profit to go up by something, but it also causes the cost of production to go up by something. The "somethings" are the terms in the expression $(P \times MPP_L) - w$. The expression measures the increments to profit from hiring another hour. Evidently, the firm will make the most profit when no more profit is to be made by pushing hiring further. That is, profit will be at a maximum when $(P \times MPP_L) - w$ equals zero, or in other words when $P \times MPP_L = w$. When the wage equals the value of the marginal product, the firm stops hiring more labor.

The theory is called the *theory of marginal productivity*. It asserts that firms are willing to pay at most the dollar value of the marginal increment to product that the hiring of an additional lumberjack hour produces. Inputs are valued, in other words, not for their total product—not for how much lumber would be lost if all the hours of lumberjacks quit work—but for their marginal product—the lumber that would be lost if *one* alone quit work.

Since one lumberjack hour is fungible with all others, all must receive this same marginal product. As the expression goes, the hourly wage of lumberjacks is *determined at the margin*. In this respect as in many others the theory of the demand for inputs is identical to the theory of the demand for products by consumers: The value of the marginal utility (the marginal valuation) determines the price at which a consumer will buy; the value of the marginal product determines the price at which a firm will buy.

The argument is innocuous, even boring, until one realizes that alternative theories of value underlie the sentiments of the man in the street.

T or F: Garbage collectors should be paid \$100,000 each, because the city would collapse in a stinking heap if the garbage collectors did not collect.

A: The total product, or the product taken away if all labor is taken away, does not determine economic value. It may determine moral value. That is, one may believe that such an argument implies that garbage collectors ought to be paid a lot, on moral grounds. But it does not determine observed economic value in a freely func-

tioning marketplace. The marginal product of one extra garbage collector determines economic value, and this is at present well below \$100,000. In short, false.

T or F: The price paid for water is no indication of its true value in use because the water makes the production of additional wealth possible. Thus a farmer may pay \$8 for water for one acre of his land, yet the value of the crop grown on that land might be \$100.

A: The \$92 pays the other factors of production necessary to make the irrigation have more use than merely an expensive humidifying system. The marginal value

product of the water is still \$8, its marginal contribution. Therefore, false.

In a sense the firm is indifferent between hiring or not hiring the last worker or the last gallon of water or the last input of anything. The last worker is paid exactly what he or she contributes to revenue.

Q: Joe Namath (a famous football player) retires from the New York Jets (a famous football team) and begins to play for the Shreveport Steamer (another football team, *not* famous). The stock in the Shreveport Steamer rises in value and the stock in the New York Jets falls. Joe Namath was and is earning less than his marginal revenue product.

A: If he was earning his marginal revenue product, owners of either team (the stockholders) would be indifferent between keeping Namath or letting him go, because Namath's contribution to their profit would be zero (*he* would be receiving the contribution he made to the net revenues of the teams). Therefore, true.

Marginal Productivity Is a Theory of Distribution

In the manner of all marginal curves, the area under the curve of marginal product is the total product. The total revenue of the firm must be divided up in some way between labor and the other factors of production. If a lumber firm accords with the theory of marginal productivity, the division will be accomplished as shown in Figure 22.5.

The heavily outlined areas is the total product. The rectangular area marked wL_e is just that: the wage multiplied by the amount of labor in equilibrium—here, \$10,000,000. Since it is the income to labor, the shaded triangular area must be the income to any other factors of production, such as land or capital. In other words, *the theory of marginal productivity implies a particular distribution of the total product between labor and other factors*. Think about that: The theory tells about the distribution of income, about who is rich and who is poor.

It Is a Theory of Profit

The triangular area is equivalent to consumer's surplus, the consumer in this case being the firm and the good bought being the labor hired. The other factors of production, such as land and capital, can be lumped together as fixed and owned factors, the reward to which the firm is trying to maximize. With such an interpretation the triangular area becomes the profit. The interpretation is natural: Just as consumer's surplus gives a measure of the total gains from trade to the consumer, so too profit (which is an analogous area) gives a measure of the total gains from trade *to the firm*.

In view of the analogy, it is not surprising that the case in favor of free trade among firms involves the same diagrams as the case for free trade among nations. Big firms such as Sears or General Motors, for instance, are often faced with the question of whether to buy an input (such as advertising or computing services) from the market (which is similar to importing it) or from one of their own subsidiaries (which is similar to getting it domestically). Since the subsidiaries earn profit for the mother firm, one might suppose that it would be desirable always to buy internally. But to suppose so would be to make the same mistake as in the false argument for keeping our business to ourselves as against trading with the Japanese. It is desirable for Sears to buy advertising from its own advertising division up to the point at which the rising marginal

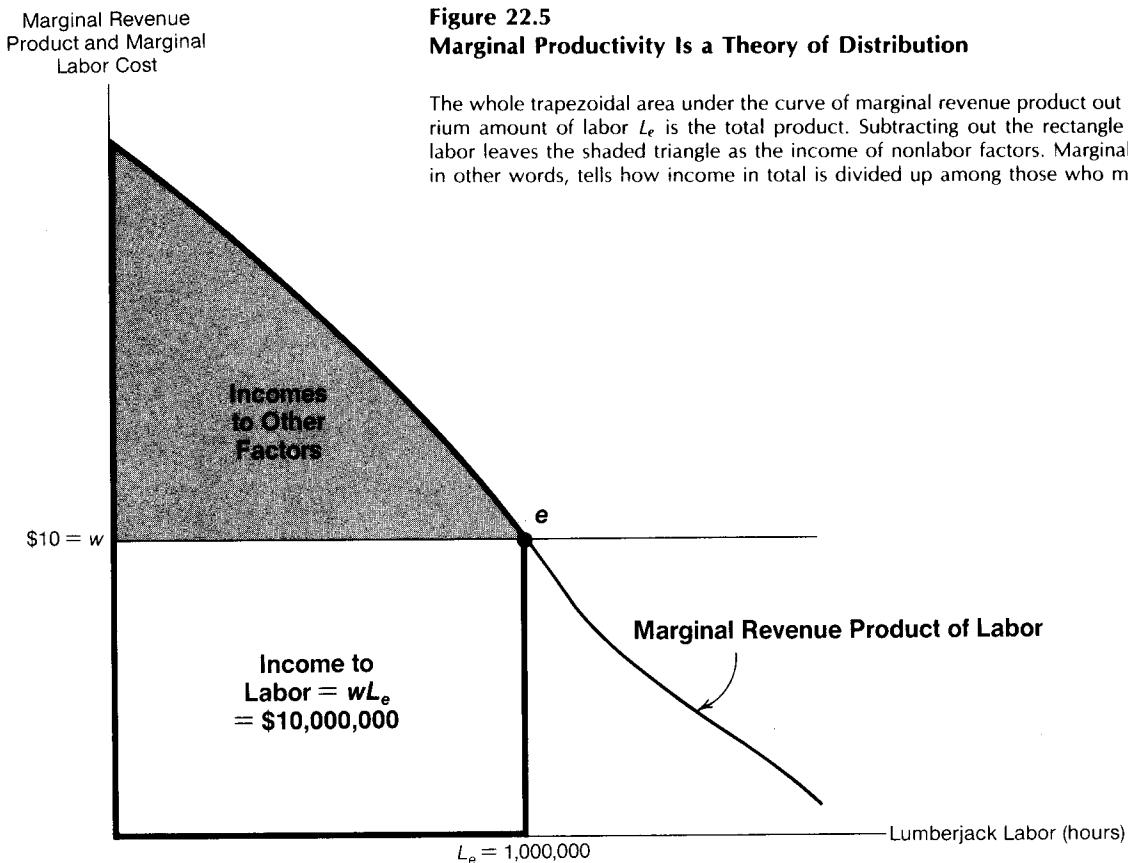


Figure 22.5
Marginal Productivity Is a Theory of Distribution

The whole trapezoidal area under the curve of marginal revenue product out to the equilibrium amount of labor L_e is the total product. Subtracting out the rectangle of income to labor leaves the shaded triangle as the income of nonlabor factors. Marginal productivity, in other words, tells how income in total is divided up among those who make it.

cost of the advertising reaches the price of advertising in the open market. Any further advertising, however, should be purchased in the open market, not taken from the now more expensive internal source.

Q: Draw the demand curve for advertising by Sears and the rising marginal cost for the Sears advertising subsidiary as more advertising is taken by Sears. Draw the market price of advertising (a horizontal line to Sears, assumed to be a small buyer of advertising). At what point should Sears start using outside advertising firms? At this point, what is the "profit" earned by Sears (that is, the income earned by all factors other than advertising)? Make sure that you include the profit earned by Sears's own advertising division (viewing it as supplier's surplus). What would Sears's equilibrium amount of advertising and of "profit" be if Sears had a policy of making as much money as possible, but never using outside advertising? What would they be if Sears had

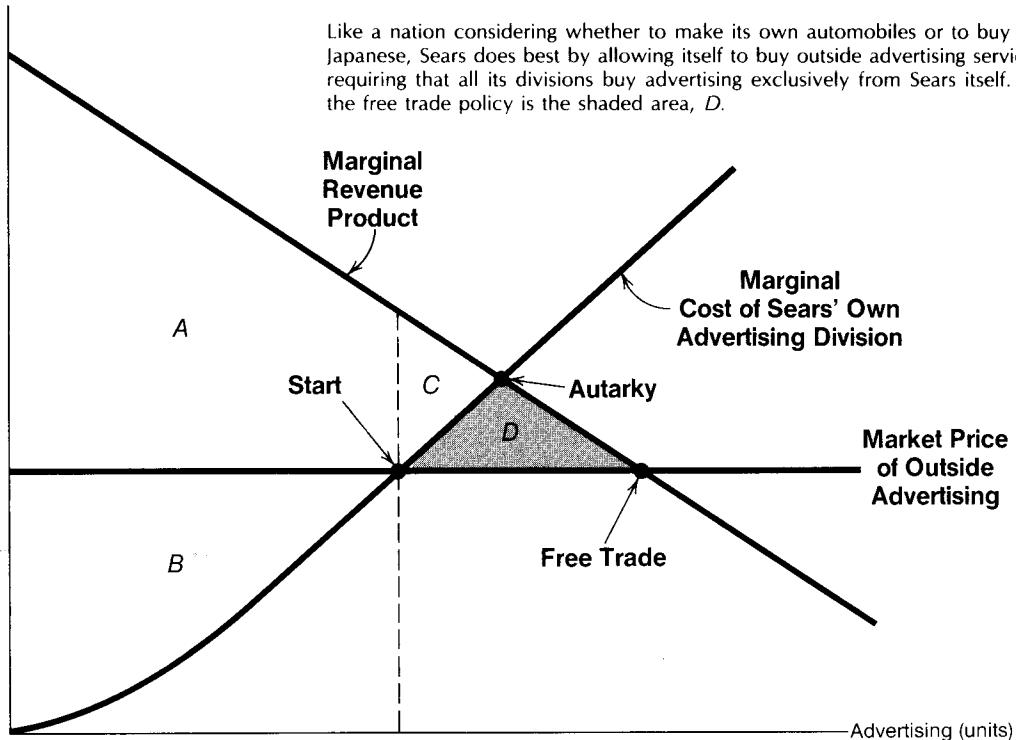
a policy of making as much money as possible, but permitting itself to buy outside advertising?

A: Sears starts using outside advertising at the point Start in Figure 22.6. The profit at that point is the area *A*, which is the consumer's surplus (so to speak) accruing to Sears as a buyer of advertising (from itself), plus the area *B*, which is the producer's surplus accruing to Sears as a seller of advertising (to itself). Sears's equilibrium point under the incorrect policy of forbidding outside advertising would be at the point Autarky, with profits *A* + *B* + *C*. Sears could clearly do better by buying foreign advertising. The better equilibrium would be the point Free Trade, with more advertising

Marginal Revenue
Product and Marginal
Cost of Advertising

Figure 22.6
The Maximization of Profit Entails Seeking out the Lowest-Cost Supplier

Like a nation considering whether to make its own automobiles or to buy them from the Japanese, Sears does best by allowing itself to buy outside advertising services rather than requiring that all its divisions buy advertising exclusively from Sears itself. The gain from the free trade policy is the shaded area, D .



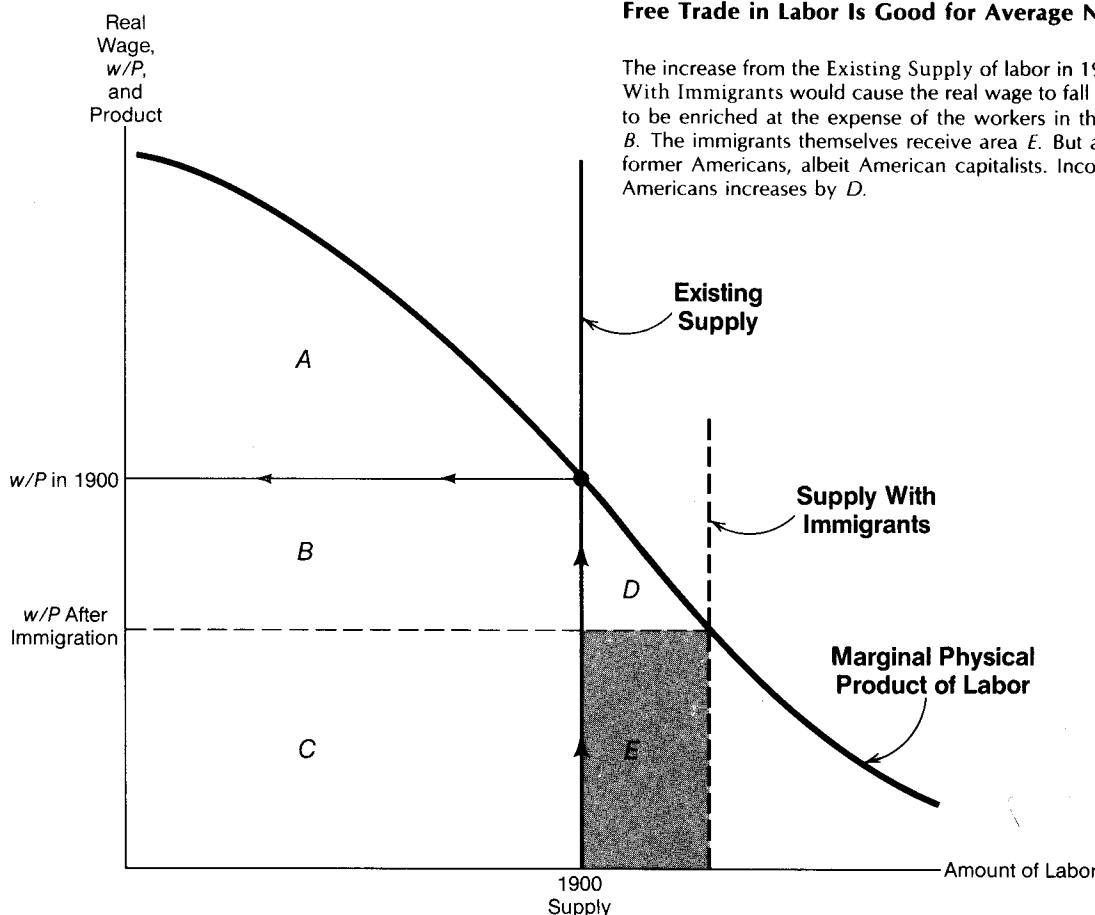
bought and more profits (in the amount of the shaded area, D). Notice that the question and the problem is the same as an earlier one on the cost curves of American Motors in making or buying crankshafts. The pres-

ent problem concerns the maximization of product; the earlier problem concerned the minimization of cost. They are two sides of the same coin.

And It Is a Theory of National Income Distribution

The total product to be distributed to different factors can be the whole national product as well. If other factors such as land and capital are taken as fixed in amount, at least in the short run, then the marginal product of the labor force of 1900 in producing American national product in 1900 determines the prevailing wage. Look at Figure 22.7, ignoring the dashed lines for a moment.

For economywide problems it usually makes sense to express the equilibrium condition (and the diagram corresponding to it) as $w/P = MPP_L$ instead of $w = P \times MPP_L$. In contrast to the situation of a single firm, neither the money wage nor the money price of national product is given from outside the nation. Only their ratio is determined by marginal productivity. The ratio w/P is the dollar wage per dollars of price—it is the purchasing power of the wage, or the real wage. Notice too that the units work out correctly: w/P is dollars per person divided by dollars per ton of product, which reduces to physical tons per person, in keeping with the marginal physical product to which it is equal. The existing supply of labor in 1900 leads to the real wage in 1900.



**Figure 22.7
Free Trade in Labor Is Good for Average National Income**

The increase from the Existing Supply of labor in 1900 to the Supply With Immigrants would cause the real wage to fall and the capitalists to be enriched at the expense of the workers in the amount of area *B*. The immigrants themselves receive area *E*. But area *D* accrues to former Americans, albeit American capitalists. Income of the former Americans increases by *D*.

The diagram can be used to think about the effects of America's policy in 1900 of free immigration. Between 1900 and 1910, 13.4 million people—among them your great grandparents—immigrated to the United States, a 15% increase in the population.

- Q:** 1. In Figure 22.7 identify the income earned by the Existing Supply of labor in 1900.
 2. What is the whole national income? What is the income earned by nonlabor factors, such as capital and land?
 3. Notice the Supply with Immigrants. How much do the immigrant laborers alone earn?
 4. How does the coming of the immigrants affect the income earned by the laborers who were in America before immigration?
 5. How does the coming of the immigrants affect

the income earned by American capitalists and landlords?

6. What is the income earned after immigration by *all* former Americans (*excluding what the immigrants themselves earned*)—laborers, capitalists, and landlords together? Does immigration on the whole hurt or help nonimmigrant Americans?

- A:** 1. The income earned by labor in 1900 in area *B* + *C*.

2. The whole national income is $A + B + C$, which is to say that A is earned by nonlabor factors.
3. The immigrants earn the lowered marginal product, getting in total the shaded area E for their trouble in immigrating.
4. The coming of the immigrants lowers the marginal product of labor, effectively transferring the area B from former American workers to former American capitalists and landlords. It should be pointed out here that real wages did not in fact fall from 1900 to 1914, but rose sharply. Other things were happening, however. The thought experiment asks what the effect of immigration in isolation was.

5. To the delight of capitalists and landlords, their income grows by area B and by D (D being their earnings as capitalists and landlords on the hiring of immigrants).

6. In other words, the income of former Americans, once $A + B + C$, is after immigration $A + B + C + D$. It has risen by D . Immigration hurts the factor with which it competes (unskilled labor in the case in question), helps the other factors, *and helps the other factors more*, leaving a *net gain* to the whole of income.

Summary

The demand for labor by the profit-maximizing firm is related to the production function of the firm. To be precise, the demand curve is the curve of the marginal revenue product of labor, derived from the total product curve. The theory is called marginal productivity and is closely analogous to the theory of marginal utility in consumption. In both theories, the amounts bought are determined by marginal, not total or average, valuation. The price of labor determines through the marginal revenue product curve the amount purchased by a firm; the amount of labor determines through the marginal physical product curve the real wage of labor. Both versions of the argument are widely applicable, to questions of the desirability of buying advertising from other firms, for example, or the desirability of buying citizens from other countries.

EXERCISES FOR SECTION 22.2

1. At a certain level of land, capital, equipment, and so forth the yearly output of lumber for the Olmstead Lumber Company varies with hours of lumberjacks hired as follows:

Yearly Hours of Lumberjacks	Yearly Output of Lumber (thousands of board-feet)
400,000	20,000
500,000	27,000
600,000	40,000
700,000	61,000
800,000	79,000
899,985	93,999.925
900,000	94,000
999,999	99,999.95
1,000,000	100,000
1,100,000	104,000

Suppose that the price of lumber is \$200 per 1000 board-feet. What are the total revenue products at each number of hours? (Watch it: It's 20,000 *thousands* of board-feet; the units are chosen to fit the way lumber is actually quoted.) What are the marginal products? (*Hint*: Reduce them to marginal revenue products per hour by dividing by the 100,000-

hour interval; thus, MRP at 400,000 hours is the increment that comes from the next 100,000 hours—namely, \$5,600,000 minus \$4,000,000, or \$1,600,000—divided by 100,000 hours, or \$16 an hour.)

2. In Exercise 1, what are some points on the demand curve for labor? How much labor is hired if the going wage is \$10? \$30?

3. At \$10, how much do lumberjacks earn? How much does the firm as a whole earn? How much is earned by nonlumberjack inputs (for instance, the owners of the firm)?

4. At \$15 an hour how much do lumberjacks earn? How much does the firm as a whole earn? How much is earned by the nonlumberjack inputs? What happens to the share of labor in total costs if lumberjacks get \$15 rather than \$10 an hour?

PROBLEMS FOR SECTION 22.2

- 1. It is said that in poor countries, especially in agriculture, the population is so large that labor is surplus, that people work at conventionally defined jobs, but that the marginal worker does not in fact add anything to income or at best adds something less than what he or she is paid. In 1918–1919 India was struck by famine and, worse, by a flu epidemic that killed 5 million people. If the strong form of the surplus labor hypothesis is true (the marginal worker adds nothing), what would have happened to agricultural output? If the weak form is true (the marginal worker adds something, but less than the wage), what would have happened? In fact, agricultural output fell roughly by the amount of the wages paid to the workers killed. What do you conclude about the plausibility of the surplus labor hypothesis?
- 2. Suppose that the production function in some industry is a Leontief or *fixed coefficient* function, namely, $Q = \min(\alpha L, \beta K)$, where L is the amount of labor, K the amount of capital, and α and β constants. For some given amount of K , what does the marginal product of labor look like?

True or False

- 3. If the total product curve did not bend downward (exhibiting diminishing marginal returns to additional inputs of labor, given other factors), then firms would specialize in hiring labor.
- 4. Since workers are paid their marginal product, shooting half of them would not affect the income of the rest of society.



Marginal Productivity in Theory and in Use

23.1 Many Inputs, Constant Returns to Scale, and the Fundamental Theorem

What to Read For How does a firm decide how much of several factors of production to hire? What is the *problem of the exhaustion of the product*? What geometrical device generates a constant-returns-to-scale production function? What is the *Cobb-Douglas production function*? Does the Cobb-Douglas production function always exhibit constant returns to scale? How can a competitive firm have constant returns to scale? What is the Fundamental Theorem of Marginal Productivity?

Deciding on Many Inputs at Once Is a Problem in Simultaneous Equations

The point at which labor's marginal revenue product and labor's supply curve cross, then, determines how much labor a lumbering firm wants to hire. But of course the same is true of how much advertising or land or coal or machinery a firm wants to hire. One can view the firm as looking at the curves for each input and making a decision for each. The curves for each input are related. The marginal product of lumberjack 100, for example, is obviously higher in a lumbering operation covering 10,000 rather than 10 acres, or having 12 caterpillar tractors rather than 1. The decisions to hire lumberjacks, forestland, and tractors should therefore be made all at the same time. The analysis so far can be easily extended to three factors at the same time. The statement that "the marginal revenue product for labor is related to the amounts of land and capital available" amounts to saying that land and capital as well as labor *appear in the algebraic expression for labor's value of marginal product*.

The algebraic expression for labor's value of marginal product could take many

forms, depending on the shape of the production function connecting output with the inputs of factors. The most general way of putting the matter is simply to note that MPP is some function of L , T (in French, *terre*, meaning land), and K . The marginal conditions are three:

1. $P \times MPP_L(L, T, K) = w$
2. $P \times MPP_T(L, T, K) = r$
3. $P \times MPP_K(L, T, K) = iP_K$

in which each of the subscripts serves to specify a different mathematical form (in other words, MPP_L and MPP_T may be different algebraic expressions, the one containing a squared T , for example, the other not). To each of these equations corresponds a marginal productivity *diagram* in which a value of marginal product curve is set against a (flat) marginal factor cost curve. The w is the wage, r the rent (per acre, say), and iP_K is the annual cost of repaying the load at $i\%$ per year on the P_K dollars necessary to buy a machine. The procedure of moving out to the point of equality of the value of marginal product of labor and its price is equivalent to *solving the three equations simultaneously*.

The Exhaustion of the Product

The extension of the argument to more than one factor brings up a problem that long bothered economists, namely, *the problem of the exhaustion of the product*. The problem is that there does not appear to be any reason why satisfaction of the three equations by a firm would result in spending all and only the money that came in. A simplified farm, say, with two factors, land and labor, producing wheat and paying each acre and each hour at its marginal product in wheat might over- or undercommit the money from wheat produced. The wheat crop per year might be 10,000 bushels, but the claims for rents and wages might be higher, 15,000 bushels, or the claims might be 7000, leaving 3000 unclaimed. A theory of distribution that left such large amounts undistributed would not be much of a theory.

Production Functions That Overexhaust the Product

It is no trick at all to draw or to write down in algebraic form a production function for which rewarding factors by marginal products does *not* exactly exhaust the product. For example, suppose that wheat output, Q , were equal to TL , where T is the number of acres of land employed and L is the number of hours of labor employed. The equation appears to be perfectly ordinary and acceptable. Let us see, however, whether the output is just exhausted if labor and land are rewarded by marginal productivity.

How much would output increase if one more hour of labor came to be employed? Well, with the production function being TL , the rate of increase of output is evidently just T , the amount of acres: One can view the equation $Q = TL$ as a *proportional* equation in L . For example, if 100 acres are in use, the coefficient T is 100. From $Q = 100L$ it is clear that each additional hour produces an output of 100. The marginal product of each hour in this example is therefore whatever T happens to be, which means that all the hours (L in amount) will earn in total TL if they are paid their original product. But wait a minute: TL is the *whole* output! There is *nothing* left over to reward land (which indeed by the same argument itself demands the whole output as its reward). The production function fails.

Not every conceivable production function fails, however. For example, if $Q = T + L$, then rewarding land and labor by marginal product exhausts the

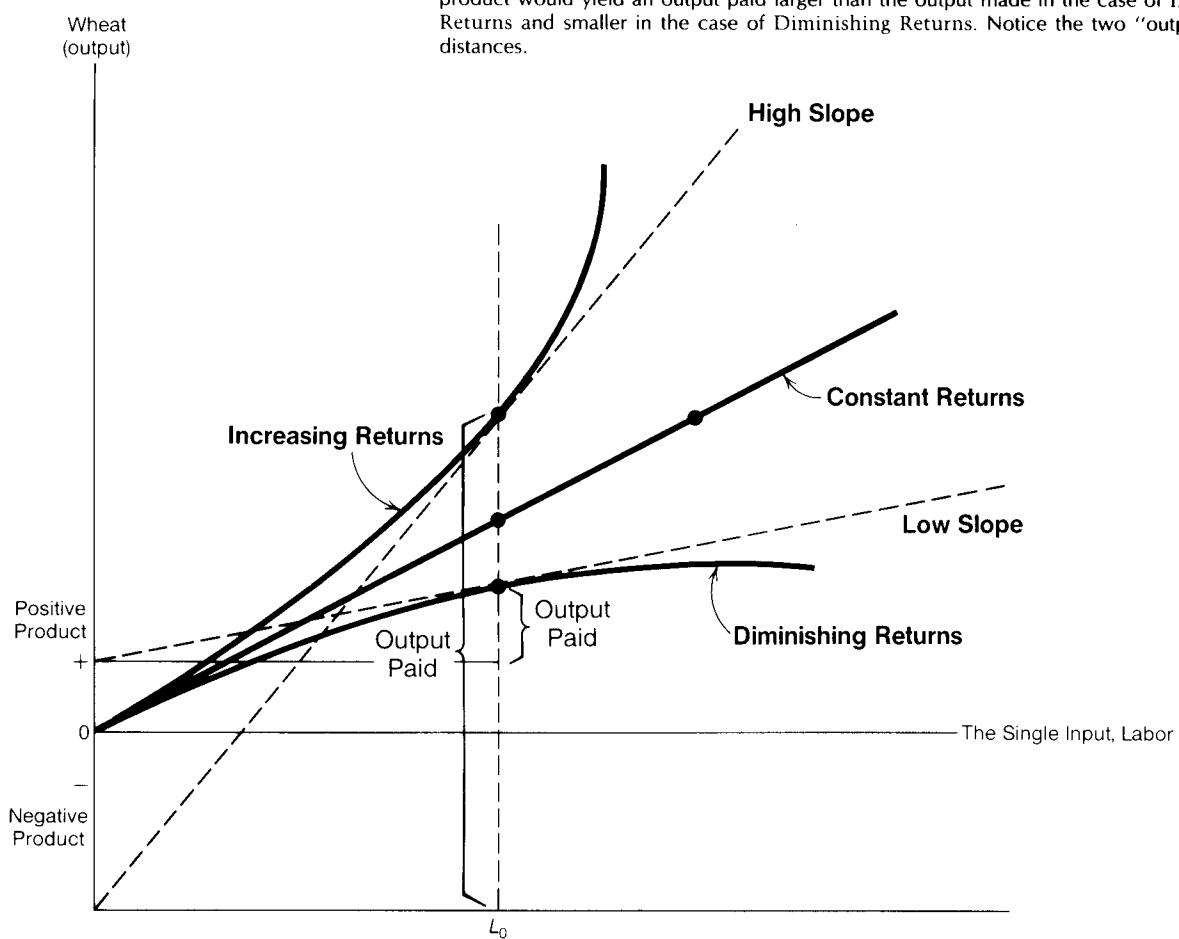
product, neither more nor less. Output increases in this equation by 1 when L increases by one unit, since 1 (and not T as before) is the coefficient in front of L . The marginal product that each hour is paid will therefore be 1. In consequence the total payment to labor will be $1 \times L$ and to land $1 \times T$. Adding them up does indeed just equal the whole output, $Q = L + T$.

Constant Returns to Scale Prevents Over- or Underexhaustion

The production functions that do *not* fail, it turns out, have one feature alone in common: They all exhibit constant returns to scale. Constant returns to scale, you will recall, is the condition that when all inputs double the resulting output also exactly doubles—not triples or rises by 50%, but doubles. In the special case of an output (wheat) produced by one input alone, it is easy to see (in Figure 23.1) why constant returns would imply that the production

Figure 23.1
Constant Returns Exactly Exhaust the Product: The Case of One Input

Since the marginal product is the slope of the total product curve, paying labor its marginal product would yield an output paid larger than the output made in the case of Increasing Returns and smaller in the case of Diminishing Returns. Notice the two "output paid" distances.



function for wheat was a straight line through the origin, not increasing or diminishing returns.

Were the input at the level L_0 , the Low Slope (that is, the marginal product) along Diminishing Returns would imply that less than all the output would be paid out; along Increasing Returns the High Slope would imply that more than all the output would be paid out. Only for Constant Returns is the output paid out equal to the output produced.

Constant Returns in the Two-Input Case

In the more general case of two inputs, the idea of constant returns is captured geometrically by the following restriction. Imagine a production function rising out of the corner on the floor of your room. Look at the corner. Take the right edge of the room near the floor to be the Labor Axis, the left the Land Axis, and the corner of walls rising up to the ceiling the third axis, measuring Output. Output is zero, therefore, when either the amount of labor or land is zero. The hill of production thus imagined could have any shape. But suppose that it has a shape that could be generated by sticking a rod through the zero point (the floor corner of the room) and rotating it through this zero point around the room in any way whatsoever. Every place through which the rod passed (one point on the rod fixed at the origin always) would be the surface of the hill.

The resulting surface exhibits constant returns to scale. This can be shown quite easily. Imagine the rod. Moving along any one position of the rod—going, say, from the floor corner to where your left thumbnail is now—the inputs of labor and land are increasing in some identical proportion (straight lines through the origin are lines of proportion). At the same time output is increasing by the same proportion. Therefore, as was to be shown, constant returns holds.

If you start the rod lying along the right floor edge of the room, rotating it in a smooth arc up toward the ceiling and then down to the left edge, the production function will have all the nice qualities known as “neoclassical.” Output will be zero, for example, when any one of the inputs is zero (because the rod is on the floor—at zero output—when it is on either edge). Because the arc is smooth, the function will everywhere have “well-behaved” slopes (marginal products). Because the arc rises and then falls instead of snaking or jerking through the air, the isoquants, or contour lines on the hill of production, will have a normal shape, convex to the origin. Finally, to repeat, since the surface of the hill is made up of rays through the origin, the production function exhibits constant returns to scale.

It is not instantly obvious why constant returns as expressed in this geometric way causes the product to be exhausted by rewards of marginal products. But it does suggest very strongly that the many-input case is analogous to the single-input case (in which it *is* obvious why the product is exactly exhausted: look at Figure 23.1 again). A moving rod is merely a generalization of a single ray through the origin.

The Algebra of Constant Returns

The corresponding algebra is necessary for later applications, and in fact exhaustion of the product is a theorem in this algebra. The theorem requires calculus (given in the appendix to this section), but the simple algebra at least makes precisely clear what is being asserted by exhaustion. A production function that is supposed to exhibit constant returns should have output doubling when

all inputs do. For instance, $Q = L + T$ does: If L and T are replaced by $2L$ and $2T$, the result is $2L + 2T = 2(L + T) = 2Q$, and as required, output also doubles.

Q: Does $Q = LT$ exhibit constant returns?

A: Replacing L and T by $2L$ and $2T$ leads to $(2L)(2T) = 4(LT) = 4Q$, not $2Q$ as required. Therefore, the

function exhibits increasing, not constant, returns to scale. When L and T double, output Q more than doubles.

More generally, if some function F in $Q = F(L, T)$ exhibits constant returns, then multiplying all the inputs by some number λ will result in a rise in Q in the same proportion, namely, by the proportion λ . That is, $F(\lambda L, \lambda T) = \lambda^1 Q$. The exponent of 1 on λ is meant to emphasize that such functions are *homogeneous of degree 1*, which is mathematical jargon for constant returns to scale. An exponent of 2 instead of 1 (that is, squaring of the proportion of change) would indicate that the function exhibited *increasing* returns to scale, being homogeneous of a degree greater than 1 (namely, 2). A rough way of determining the degree of homogeneity is to treat all the different variables as being the same (say, all xs), multiply them as indicated by the function, and examine the exponent on the resulting variable. You can see that to do this you have to remember the high school algebra of adding exponents.

Q: Is $Q = LT$ homogeneous of degree 1?

A: If L and T were the same the equation would be L^2 (or T^2). The exponent 2 is not 1. Therefore, false.

The Cobb-Douglas Production Function

The most famous example of a degree 1 homogeneous function is the *Cobb-Douglas production function*, named after an Amherst mathematics professor Cobb and a University of Chicago economics professor (later U.S. senator) Paul Douglas, who attempted in 1934 to fit it to the facts of the American economy 1919–1922.¹ The result was an equation such as $Q = AL^{3/4}K^{1/4}$, where A is a constant representing the height of the production function (that is, efficiency in the usual sense), L is the amount of labor, and K the amount of capital.

Q: Is the function homogeneous of degree 1?

A: If L and K were made identical, the equation would be $Q = AL^{3/4+1/4} = AL^1$, the exponent being 1 as required. Therefore, yes. The Cobbs-Douglas function with exponents summing to 1 is homogeneous of degree 1.

Q: Does the more general version of the Cobbs-Douglas

production function, $Q = AL^\alpha K^\beta$, in which $\alpha + \beta$ do not necessarily sum to 1, always exhibit constant returns to scale?

A: The equation with K and L made identical would be $Q = AL^{\alpha+\beta}$, in which $\alpha + \beta$ need not be 1. Therefore, no, the function does not always exhibit constant returns to scale.

¹ Paul H. Douglas, *The Theory of Wages* (New York: Macmillan 1934), especially Chapter V.

Why Constant Returns and Marginal Productivity Exhaust the Product	<p>The exhaustion of the product is assured by <i>Euler's theorem on homogeneous functions</i>, which essentially asserts that exhaustion is true for functions constructed of rays through the origin, that is, homogeneous of degree 1. A proof of Euler's theorem using calculus, which involves some mathematics important for later application, is given in the appendix.</p>
How Can a Competitive Firm or a Varied Industry Have Constant Returns?	<p>In short, if the production function exhibits constant returns to scale, payment by marginal product just exhausts the product. But something is wrong. A firm that experienced constant returns, supposing that it faced elastic supplies of factors (that is, fixed prices at which it could hire factors), would have a flat, not rising, marginal cost curve. And, as was shown in an earlier chapter on the firm, if a price-taking firm has a flat marginal cost curve, it will expand indefinitely when the price of what it produces is greater than marginal cost. Or it will contract to zero output when the price is less than cost; or it will be indifferent among outputs when the price is equal. None of these is very pleasing.</p> <p>The response is something of a shell game, but it stifles doubt. Suppose that all firms are identical but do not have production functions with constant returns everywhere. In full competitive equilibrium, however, each firm finds itself at the bottom of its average cost curve. If demand grows, another firm enters, keeping the supply price exactly at the minimum point of average cost. In other words, the <i>whole industry</i> has a cost curve that looks as if it came out of a constant returns production function. Indeed, at the precise point of competitive equilibrium the average costs of each firm just exhaust its revenues (economic profits, or rents, are zero). This suggests the convenient fiction of thinking of <i>each firm</i> as being at the constant returns point of its individual production function. If the firms were not identical, some would be earning rents in equilibrium. But such a deviation from a flat supply curve can be eliminated by assigning a name (<i>entrepreneurship</i>, say) to whatever it is about the firms that allows them to earn more profits. Their profits become, then, payments to the factor entrepreneurship and <i>voila!</i> their underlying production functions are <i>made</i> to be identical, and to exhibit constant returns. In other words, the argument moves from zero profits to constant returns, not in the other direction. Since profits are zero in competitive equilibrium, the product is just exhausted, and one can speak of "the" production function of either the firm or the industry as exhibiting constant returns.</p>
The Fundamental Theorem of Marginal Productivity	<p>Such flirting with tautology makes possible a useful tool, <i>the Fundamental Theorem of Marginal Productivity</i>: The share of costs actually paid to labor (or to capital or to whatever) in a competitive industry experiencing constant returns to scale is the elasticity of output with respect to labor (or capital or whatever). Look at that statement again. It says that the number that is the share of cost will <i>equal</i> the percentage increase in output per change of input. A little input—number 2 pencils, say—will not much affect output if it doubles. The proof goes as follows. For a competitive steel industry, say, the "marginal value product" of labor is simply the marginal physical product of labor multiplied by the going price of steel, since each company as a competitor believes that it will get the price (and not some marginal revenue lower than the price) if it expands output by one unit. According to the theory of marginal productiv-</p>

ity, labor is hired up to the point where its wage, w , equals the marginal revenue product of labor, or, in this case, $w = P \times MPP_L$. The real wage, w/P , is therefore set equal to the rate of change of steel output with respect to labor inputs. Write out the equation in this form:

$$\frac{w}{P} = \frac{\Delta Q}{\Delta L} \quad (= MPP_L)$$

Now multiply both sides by the workers used per ton of output (the inverse of output per worker):

$$\frac{L}{Q} \left(\frac{w}{P} \right) = \frac{L}{Q} \left(\frac{\Delta Q}{\Delta L} \right) = \frac{\Delta Q/Q}{\Delta L/L}$$

Notice the rewriting of the right-hand side. It is the elasticity of output with respect to the input labor. The left-hand side is the share of labor in total costs. So, the theorem is true:

$$\begin{aligned} \text{Share of labor} &= \frac{wL}{PQ} = \frac{\Delta Q/Q}{\Delta L/L} = E_{Q-L} \\ &= \text{elasticity of output with respect to labor} \end{aligned}$$

Uses of the Fundamental Theorem Consider the following:

T or F: If the labor force in steel increased 10% and the share of labor in steel costs were 25%, output would increase 10%.

A: It would increase by the increase in the labor force multiplied by the elasticity of output with respect to labor. If steel is a constant returns industry, the compa-

nies can pay workers (and other inputs) the real value of their marginal product and survive. Therefore, by the fundamental theorem of marginal productivity the elasticity is 0.25 (one-fourth) and the resulting increase is one-fourth of 10%, $(10\%)(0.25) = 2.5\%$, not 10%. Therefore, false.

The reason that constant returns to scale is necessary for the Fundamental Theorem is that without it the payment of factors by marginal product will over- or undercommit the revenues available to pay costs. In such a case the factors cannot really be paid their marginal products. The shares will have to be determined some other way. The shares resulting from this other way will not bear any relation to marginal product and therefore will not permit the algebra that led to the theorem.

The Fundamental Theorem is applicable to the output of an entire nation:

Q: It is often alleged that the United States is rich because of its land, rich in coal, oil, iron ore, and nutrients for plants. The share of land in American national income is well below 10%. How much would income fall if the riches of the land fell by 50%?

A: If the nation's production function can be thought

of as exhibiting constant returns to scale, then the conditions of the Fundamental Theorem apply. Therefore, the fall would be $(50\%)(0.10) = 5\%$ at most. You should be suspicious of arguments that natural resources explain much of the differences in the wealth of nations.

Summary The extention of marginal productivity to more than one factor of production is trivial. Instead of solving one equation, the firm solves—crudely, no doubt, and largely by rule of thumb—a set of simultaneous equations. But paying its marginal product to each factor does not always exactly use up the product, an embarrassing oversight in a theory of distribution. The one condition of the production function that assures exact exhaustion of the product is constant returns to scale, which amounts to the condition that the function be generated by pivoting a rod fixed at the origin. Constant returns can be reconciled with the diminishing returns necessary for equilibrium in the size of the firm by fixing attention on the long-run equilibrium at the minimum point of average cost and by introducing a background factor of production, entrepreneurship, units of which serve to define the firm. By these devices we are permitted to speak of an industry's production function exhibiting constant returns even if a firm's production function does not. And we are permitted to use the Fundamental Theorem of Marginal Productivity, namely, that an input's share in costs is the elasticity of output with respect to the input. An input with a small share is unimportant.

APPENDIX TO SECTION 23.1: The Proof of Euler's Theorem Using Calculus

Euler's theorem asserts that, for a first-degree homogeneous function, $K(MP_K) + L(MP_L) = Q$. The following proof calculates the marginal products, and from them shows that it is indeed true.

If F is homogeneous of degree 1 in K and L , then by definition $F(\lambda K, \lambda L) = \lambda Q$. Suppose you took λ to be $1/L$ (well, why not?). The result would be $F(K/L, L/L) = Q/L$, which says that the output per person, Q/L , depends only on the ratio K/L , not on the absolute size of K or L separately. The result makes sense geometrically, since output per person is the slope relative to the L axis of a rod from the origin out to the point in question, and along any of the rods used to generate the production function the slope will of course not change, since a rod is a straight line no matter how you look at it. The whole output can therefore be written in a back-door manner as follows: $L(Q/L) = L[F(K/L, 1)] = Q$. Take the partial derivative of the middle expression for Q with respect to K alone to find the marginal product of K . The result is $\partial Q/\partial K = L[\partial F/\partial K](1/L)$, by the function-of-a-function rule [K/L being the inside function, $F()$ being the outside]. The L 's cancel, leaving $MP_K = \partial Q/\partial K = \partial F/\partial K$. That is, the marginal product of capital, as one might expect, is just the derivative of F .

In this way of putting it, the F was a function of K/L alone, not K and L by themselves. So too, then, is its derivative. In consequence, for a production function homogeneous of degree 1, the marginal product of capital, $\partial Q/\partial K$, can be expressed as a function of K/L alone, not merely K and L separately. Homogeneity of degree 1 allows one to draw a marginal product curve as a function of K/L alone, regardless of the absolute scale of the output.

What, then, is the other marginal product, MP_L ? It is the partial derivative with respect to L of the total output expressed as $L[F(K/L, 1)]$, namely (by the derivative of a product rule),

$$\begin{aligned} MP_L &= \frac{\partial}{\partial L} \left[L \times F\left(\frac{K}{L}, 1\right) \right] = F\left(\frac{K}{L}, 1\right) + L F_K \left(-\frac{K}{L^2}\right) \\ &= F\left(\frac{K}{L}, 1\right) - F_K \left(\frac{K}{L}\right) \end{aligned}$$

Notice that the derivative of F with respect to K appears in this because only the first, K , variable in F can change (the second, L , variable being equal to 1 all the time). The equation says that the marginal product of labor is output per person, $F(K/L, 1)$, minus the payment to capital, $F_K K$, per person (divided by L). Clearly this equation by itself is nearly Euler's theorem. In fact, if you multiply both sides by L , recognizing that F_K is indeed (by the earlier step) the marginal product of capital, the result is

$$L(MP_L) = L \left[F\left(\frac{K}{L}, 1\right)\right] - K(MP_K)$$

or, rearranging and noticing that L times the output per person is simply output,

$$Q = L(MP_L) + K(MP_K)$$

as required.

EXERCISES FOR SECTION 23.1

- 1.** Here are the curves of the value of the marginal product for the Shlomowitz Sugar Mill of Australia in 1900, with the amount of labor in man-years called L and the amount of machinery in mill years called K :

$$(\$P) \cdot MPP_L = 2000 - 25L + 1000K$$

$$(\$P) \cdot MPP_K = 100,000 - 150,000K + 1000L$$

The wage of L is \$500 per year, the rental of a sugar mill is \$50,000 per year.

- a. Write down the equilibrium conditions, that is, the two equations that determine how much L and K Shlomowitz will employ.

- b. Why can't you determine the amount of labor, L , that Shlomowitz will want to buy simply by looking at the first equilibrium condition (the wage and labor's value of marginal product by itself)? What is the way to get around the problem?

- c. Solve the two equations simultaneously for L and K . (*Hint:* The solutions are nice, simple, whole numbers.)

- 2.** (For students who know calculus) What are the *marginal revenue products* (the rate of change of total revenue) of L and K if the total revenue depends on L and K like this:

$$\text{Total revenue} = 200 + 2000L + 100,000K + 1000KL - 12.5L^2 - 75,000K^2$$

- 3.** Which of the following are constant-return production functions:

- a. $Q = KL$
- b. $Q = K^{3/4}L^{1/4}$
- c. $Q = \alpha K + \beta L$
- d. $Q = \alpha K^2 + \beta L^2$
- e. $Q = \alpha K + \beta K^2 + \gamma L$

- 4.** Develop arguments about which of the following are characterized by constant returns to scale. (*Hint:* Imagine how output would change if the inputs were doubled.)

- a. The popcorn stand industry in the United States.
- b. The popcorn stand industry at the corner of Washington and College in Iowa City.
- c. The American space program.
- d. Electric power from one coal-fired plant.
- e. Your university.

5. The share of number 2 pencils in the budget of Mitch Tuba Theatricals, Inc., is 0.005, the share of labor is 0.90, and the share of the tuba rental is the rest.
- If Mitch decided to double his use of number 2 pencils, how much would his output increase? How much if he doubled his labor (time) spent playing?
 - If the price of tuba rental doubled, how much would the price he charged have to rise? How much if his wage in alternative employment fell 10.56%?

PROBLEMS FOR SECTION 23.1

- Does the *constant elasticity of substitution production function*, $Q = (aK^{-\epsilon} + bL^{-\epsilon})^{-1/\epsilon}$, in which a , b , and ϵ are constants, exhibit constant returns to scale?
- The differences around 1900 between the successes of German industry and the failures of British industry are often explained on grounds of differences in the quantity of entrepreneurship. The differences cannot have been very great, for only 30 years before it was British industry that was supposed to be superior in this regard. Suppose it is believed that in 1900 German entrepreneurs were effectively a third more numerous as British. Suppose, as appears to be the case, that the share of costs that could conceivably be attributed to entrepreneurship—as distinct from routine labor, capital, or land—was only 10%. What is the greatest difference in output that can be explained by differences in entrepreneurship?
- The elasticity of *output* with respect to the *quantity of an input* is the share of the input in cost. What is the elasticity of *cost* with respect to the *price* of the input? This is another example of duality.
- In view of Question 3, if $Q = AL^\alpha K^{1-\alpha}$ is the Cobb-Douglas production function, what is the Cobb-Douglas *cost* function, which relates cost, c , to the prices of L (the wage, w) and of K (the rental rate, r)? (Notice that output does not appear in it: It exhibits constant returns to scale).

True or False

- In view of the fundamental theorem of marginal productivity, the coefficient α in the Cobb-Douglas production function, $Q = AL^\alpha K^{1-\alpha}$, is the share of labor in costs.
- If the nation's output is produced by a Cobb-Douglas production function, then increases in the real wage of labor from increases in machinery per worker will not increase labor's *share* in national income.
- The share of labor in the total costs of a nonprofit institution like a university is *not* the elasticity of output with respect to labor, because universities do not hire factors up to the point where marginal value equals marginal cost.

23.2 Changes in the Production Function

What to Read For

What is a rising production function? A unit isoquant? Is a country with higher output per man always on a higher production function? What is total factor productivity change? The residual? How can it be used to find out how much invention and war affected the American economy? Can productivity be measured by prices? Would there be any productivity change if new knowledge were the result of methodical investment? Is a country with relatively

expensive labor more likely to save labor? Why is knowledge itself difficult to analyze economically?

The Problem of Detecting Different Production Functions

The most fruitful use of the ideas of marginal productivity and the production function is in thinking about when they change. Our ancestors struggled with oxen and wooden ploughs to make wheat. We make ten bushels with the effort they expended on one. Steel before the 1850s was a luxury used in small quantities for sword blades. After the 1850s the cost of steel fell enough to make ploughs with it, and then ships and bridges. A computer that filled a building in 1950 can now be held in the palm of your hand and can be made for a thousandth of the cost. These are all cases of *rising production functions*, of getting more output for the same input of sweat and equipment and land.

It is often unclear whether two points are merely on different points of the same production function or truly on different production functions. For example, output per person was higher in America than in Britain around 1860. But it is an error to infer that American techniques were necessarily superior, because there are other factors of production besides labor to be brought into the accounting.

The Solution: The Unit Isoquant

A diagram with many uses will make the point. With constant returns to scale, the production function can be represented fully by *one* contour line, the contour line of a single unit of output (the *unit isoquant*), because 100 or 10,000 units are produced by scaling up the production of 1 unit. If 10 laborers and 1 machine are used to produce 1 unit of national income, then 10,000 laborers and 1000 machines are used to produce 1000 units (if production takes place without diminishing or increasing returns to scale). For example, taking American national income to be produced with labor and land in 1860, the American unit isoquant might be as pictured in Figure 23.2 (ignore the dashed lines for a moment).

With its abundant land relative to population in 1860, America might take up a position such as American Point in the diagram, using the indicated (Small) amount of labor and the Large amount of land to produce one unit of income. It will make the argument more definite if you think of income as bushels of wheat. Whether income or wheat, America takes up just one of many possible positions on the Unit Isoquant. The position it chooses is the one that its endowment of land and labor picks out. The slope of the tangent line at the position it chooses is then the American price of labor relative to land. Alternatively, the American Point can be thought of as the position that cost-minimizing industries are led to by the price of labor relative to land. At Bad Point for America the same output would be produced, since the point is on the same isoquant as the American Point, *but at a higher cost*, since a budget line (with the correct slope) would be farther out than at the American Point.

As portrayed in the diagram, the British Point is on the same production function. True, the British use radically different proportions of people and land to make wheat or housing or ships. But according to the diagram they have the same book of blueprints as the Americans. The production function, represented here by the shape of the unit isoquant, is a picture of what *can* be done—what is in the book of blueprints—not what *is* done. A “can” becomes an “is” only when firms are faced with the factor prices picking out one blueprint. To recall an earlier discussion, there is more than one way to skin a cat.

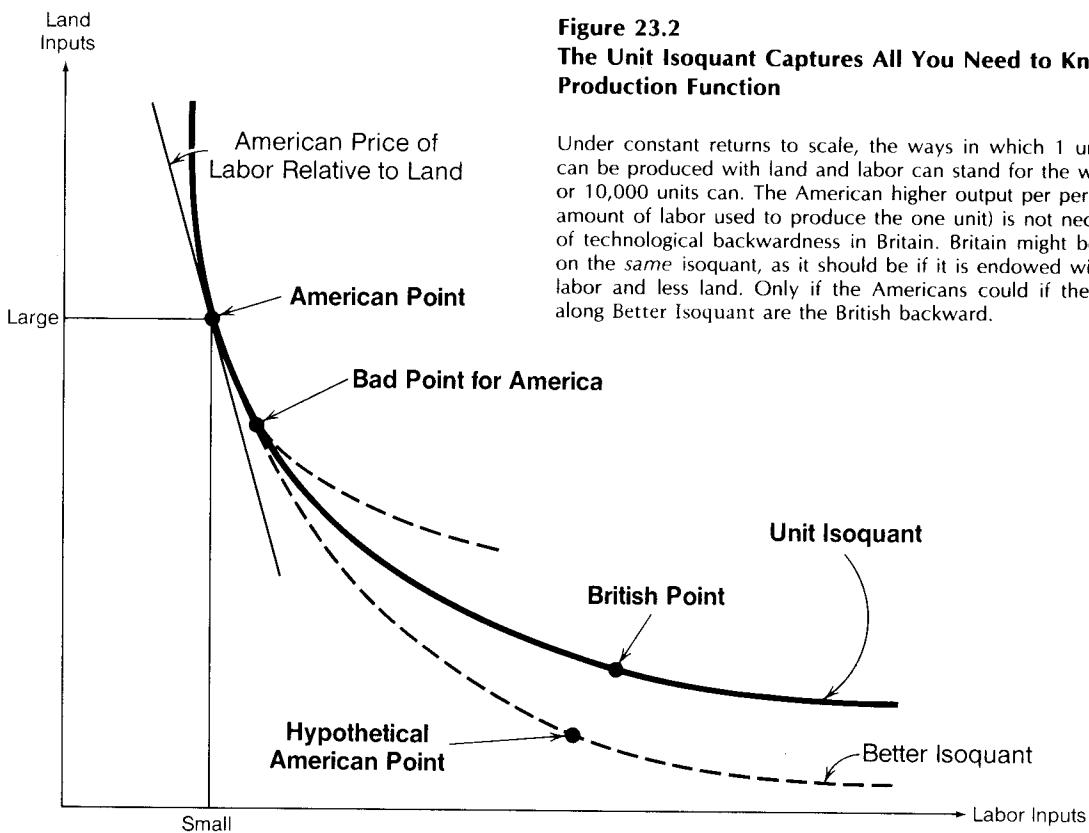


Figure 23.2
The Unit Isoquant Captures All You Need to Know About a Production Function

Under constant returns to scale, the ways in which 1 unit of the product can be produced with land and labor can stand for the ways in which 100 or 10,000 units can. The American higher output per person (that is, lower amount of labor used to produce the one unit) is not necessarily indicative of technological backwardness in Britain. Britain might be at British Point on the *same* isoquant, as it should be if it is endowed with relatively more labor and less land. Only if the Americans could if they wished operate along Better Isoquant are the British backward.

Observe, however, that American output per laborer in the diagram is much higher than the British, because the American use of labor to produce the one unit of output is much smaller. British output per acre, of course, is much higher than is the American, on the same argument. America used labor sparingly and land generously, Britain the reverse. It is quite possible, in other words, that the British higher output per acre offset the lower output per laborer. Both countries are in such a case doing as well as they can. They are skinning cats differently, because they have different endowments of factors and therefore pick different blueprints from the book—not because they have access to different books.

The alternative is the dashed Better Isoquant. It is better because it requires lower amounts of land and labor to produce the same amount, namely, one unit. If the American production function can be shown to have been this dashed line, then the British were in 1860 definitely inferior. The higher output per acre does *not* make up for the lower output per laborer. If Americans with their superior book of blueprints faced British factor prices, the Americans would not, the argument says, go to the British Point; they would go to the Hypothetical American Point on the Better Isoquant.

How Much Better? The question of whether American technology was superior to British technology in 1860 reduces, then, to the question of how much more output America could have produced with the resources used at the British Point. In other words, such questions reduce to the question of how much higher one production function is compared with another. Think back to the hill of production coming out of the corner of your room. If the production function rises, each combination of, say, labor and capital marked off as a point on the floor produces more output measured by the height of the surface above the floor. The question is how to measure the change in the height of the surface.

**Generalizing
Average Product to
All Inputs**

The answer is to generalize the notion of output per person or output per acre into output per unit of all inputs. This is a move from partial productivity to *total productivity*. Nothing is lost by focusing on the rate of change (or difference) in productivity instead of on its level. A comparison of the level of productivity in steel making and in coal mining is meaningless because the steel and coal are in different units. Is a ton of coal more or less than a ton of steel? But a comparison of the rates of change of productivity is meaningful because rates of change are unitless. To reintroduce a well-worn piece of algebra from earlier chapters, then, the measure of productivity change is the rate of change of output per unit of input, that is, signifying rates of change with asterisks, $(Q/I)^*$, or, equivalently, $Q^* - I^*$: the rate of change of output minus the rate of change of input.

The economics of the measure starts with the observation that under constant returns to scale a 10% rise in all inputs to wheat growing will result in an exactly 10% increase in wheat. Suppose that the output of wheat *actually* grows 15%. Then the production function has risen 5%. The 10% rise in inputs explains only part of the rise in output, namely, two thirds of 15%, leaving 5% as the "rise in the production function" or "total factor productivity change" or, to be modest, "the residual," or, to be honest, "a measure of our ignorance." We do not know what caused the additional 5% increase. We call it productivity change.

**Using the
Fundamental
Theorem to Derive a
Measure**

The remaining question is how to give economic life to the notion of "the rate of change of all inputs." The answer uses the Fundamental Theorem of Marginal Productivity. Crudely, a factor's share in cost is its importance among all inputs, and so its change should be weighted by its share. Formally, the theorem says that in competitive equilibrium the share of labor, S_L , is the elasticity of output with respect to labor input. If the elasticity is ϵ_L and the rate of change of labor in producing wheat is L^* , then $\epsilon_L L^* = S_L L^*$ will be the contribution of the rise in labor input alone to explaining the actual rise in output. The same holds for land. The whole rise in input is the sum of the contributions of labor and land. So $S_L L^* + S_T T^* = I^*$, the rate of change of all inputs. If Q^* is the actual increase in output, then $S_L L^* + S_T T^*$ is the amount of that increase accounted for by the theory of marginal productivity. The rate of change of output minus the rate of change of input, $Q^* - (S_L L^* + S_T T^*)$, is the residual measure of how much wheat productivity has risen.

**An Application to
American Growth,
1800–1967**

The idea is among the richest in economics. Here, for instance, are some statistics of American economic growth.²

Growth Rate in Percent per Year in:				
	Output	Land and Capital	Labor	Share of Labor
1800–1905	4.00	4.00	3.30	0.55
1905–1967	3.25	2.20	0.71	0.55

Q: What was the rate of total factor productivity change in each period? Comment on the relative importance of capital accumulation and productivity change in explaining American growth per worker in each period.

A: Applying the formula, in the 1800–1905 period $Q^* = 4.00$ and $S_{T,K}T^* + S_{L,L}^* = 0.45(4.00) + 0.55(3.30) = 3.62$, leaving a residual of $4.00 - 3.62 =$ about 0.4% per year. Only a small part of growth was accounted for by invention, better education, and so forth: Most

was a matter of piling up land and machines. By contrast, in the later period the residual was $3.25 - 0.45(2.20) - 0.55(0.71) = 1.87\%$ per year, over four times its nineteenth-century value and a large share of growth in total. Not without reason, then, do historians speak of the first industrial revolution in terms of capital accumulation and the second industrial revolution in terms of science, education, and high technology.

**Another Derivation,
with an Application
to the Civil War**

The measure of productivity change is sometimes called “ A^* ,” for reasons worth knowing. In 1957 Robert Solow showed that the measure is exact, or at any rate can be made exact to any desired degree, when the true production function before and after the change can be written as $AF(K, L)$.³ The term A is the height of the function and $F(K, L)$ is some production function homogeneous of degree 1 in capital and labor. The point is that A stands outside F and simply multiplies it. No one factor is favored by the productivity change. The change is, as it is put, *Hicks neutral*. For instance, the Cobb–Douglas version would be $Q = AK^\alpha L^{1-\alpha}$. Although it is not necessary to do so, the constancy of the shares (K 's share is α) makes it convenient to think of productivity calculations as following the adventures of A in a Cobb–Douglas function.

² Inferred from Moses Abramovitz and Paul A. David, “Reinterpreting Economic Growth: Parables and Realties,” *American Economic Review* 62 (May 1973): 428–439, especially p. 431.

³ Robert M. Solow, “Technical Change and the Aggregate Production Function,” *Review of Economics and Statistics* 39 (August 1957): 312–320; and Moses Abramovitz, “Resources and Output Trends in the United States Since 1870,” *American Economic Review* 46 (May 1956): 5–23.

Q: After the South had been defeated in the American Civil War and the slaves had been freed, agricultural output in the South fell by about a third. The supply of labor (formerly slave labor) also fell by about a third, as did other important factors of production except land. Land, of course, did not fall. *True or false:* By translating these facts into the equation for the rate of change of A depending on the rate of change of land and of labor (labor standing for all nonland factors), one can assert that there was definitely a fall in the agricultural production function in the South, perhaps because the war damaged the South or because slavery, though morally repugnant, was productive.

A: Write down the formula for A^* , the rate of rise—or in this case possibly a fall—of the production function

and see what the givens of the question say about it: $A^* = Q^* - S_T T^* - S_L L^*$. The Q^* and L^* were both -33% ; the T^* was zero. So the formula reduces to $A^* = -33\% - S_L(-33\%) = (1 - S_L)(-33\%)$. That is, there was definitely a fall in the production function, because the 33% fall is multiplied by a positive number. If one minus the share of “labor” (that is, all nonland inputs) were, say, 20% , the fall would be $(0.20)(33\%) = 6.6\%$. In short, true. In a sense the result is obvious. Output fell by a third, but inputs on the whole did not fall by a third, because the average of a zero fall in land and a one-third fall in nonland inputs must be less than a third. Output, therefore, fell more than input. Productivity regressed, telling something about the changes wrought by the war.

The Price Dual Measure of Productivity Change

Productivity affects the ratio of *prices*, too. Because it reduces the input required per unit of output, a rise in productivity will reduce the competitive price of the output relative to the prices of inputs. To put it another way, the measure of productivity change using quantities of inputs and outputs is a generalized *average physical product*, the rise in Q/I , where I is all inputs; a measure using prices is a generalized *marginal* physical product, the rise in P_I/P_Q (which in equilibrium is equal to the marginal physical product of I). To put it still another way, a cheapening of a product not explicable by a cheapening of its inputs is attributable to productivity change. The measure using prices is dual to the primal using quantities.⁴

The proof is an exercise in the algebra of rates of change. In competitive equilibrium revenues equal costs:

$$PQ = wL + rK$$

Take the rate of change of both sides:

$$P^* + Q^* = \frac{wL}{PQ}(w^* + L^*) + \frac{rK}{PQ}(r^* + K^*)$$

Sorting out into price and quantity terms,

$$Q^* - S_L L^* - S_K K^* = -P^* + S_L w^* + S_K r^*$$

If constant returns to scale holds, then the left-hand side is a measure of productivity change. *But then so too is the right-hand (price) side.* The price

⁴ The dual measure was invented, before the primal was, by an economic historian, G. T. Jones, in his *Increasing Returns* (Cambridge: Cambridge University Press, 1933). Since no one outside of British economic history read this book, the idea had to be reinvented by Dale Jorgenson, in “The Embodiment Hypothesis,” *Journal of Political Economy* 74 (February 1966), p. 3n.; and by Jorgenson and Zvi Griliches in “The Explanation of Productivity Change,” *Review of Economic Studies* 34 (July 1967): 249–283.

measure is not merely an approximation to the quantity measure. If the data are collected consistently (so that L and Q are consistent with S_L , w , and P), the two are identical.

An Application to Cotton Textiles, 1815–1859

It is often more natural to use the price measure. A piece of cotton cloth that sold in England in 1815 for 28 shillings was selling 44 years later for 5 shillings.

Q: From 1815 to 1859 the price of cotton cloth fell at 3.5% per year, the price of raw cotton fell at 2.5% per year, and wages fell at 0.33% per year (1815–1859 was in fact a period of general deflation).⁵ Labor and raw cotton were each about half of the costs of cotton cloth.

1. What was the rate of rise of productivity?

2. A piece of cotton cloth cost about 28 shillings in 1815 and about 5 shillings in 1859. Had productivity change not occurred in cotton textiles, what would have been the price in 1859?

A: 1. The rate of productivity rise is $-P^* + S_c P_c^* + S_L w^*$
 $= -(-3.5\%) + 0.50(-2.5\%) + 0.50(-0.33\%) = 2.08\%$ per year. Two percent a year is very rapid productivity

growth, most of it attributable in this period to the introduction of the power loom (a sum growing at 2% doubles in 35 years).⁵

2. The rate of fall of the price of cloth would have been, looking at the calculation in (1), $0.50(-2.5\%) - 0.50(-0.33\%) = 1.42\%$ per year. You can look up an interest rate of 1.42% per year in tables or calculate it on your computer to find that a sum growing at 1.42% grows to 1.86 times its initial level in 1859–1815 = 44 years. This is to say that the price would fall to $1/1.86$ times its initial level. The price in 1859 would have been $(0.54) (28 \text{ shillings}) = 15.1 \text{ shillings}$, three times its actual level of 5 shillings.

Why Productivity Change Occurs

The measurement of productivity change is not a theory of why it happens. The measurement is typical of much applied price theory in that it measures something (technological advance) by looking away from it, asking what output would be in the absence of technological change if marginal productivity were a complete theory of why output changed. Likewise, one can measure changes in taste by asking what bundles consumers would reveal they preferred on the hypothesis that tastes did not change; one can measure the amount of irrationality of business managers in some industry by asking what techniques a perfectly rational manager would choose and comparing the ideal with the real.

It Might Be Viewed as Not Occurring

An alternative intellectual strategy is to suppose that in fact marginal productivity is a complete theory and therefore that if a measure of productivity change is anything other than zero the measure is wrong. For if marginal productivity were the whole story of larger output, then all larger output would be attributable to larger input, not to some mysterious, noneconomic thing called "productivity change." The most radical expression of this view is the Griliches and Jorgenson work cited, which argues that invention itself is an economic activity expected to earn normal returns. True, the successful inventor of, say, the Bessemer process of steel making (by a remarkable coincidence named Sir Henry Bessemer) became rich, and the price of steel fell to a fraction of its earlier price. But others had attempted a similar process unsuccessfully. Their investments (yielding zero return) must be added into the balance with Bessemer's (yielding a high return).

⁵ See Roderick Floud and D. N. McCloskey, eds., *The Economic History of Britain Since 1700*, Vol. I (Cambridge: Cambridge University Press, 1981), p. 110.

The society as a whole did not experience productivity change, merely the normal return to investment in the search for new ideas. More generally, it can be argued that earlier investments in widespread literacy, in an orderly society of laws, in chemical education, and so forth contributed to Bessemer's success. On this view the cheapening of steel was not a free lunch dropping out of the sky, but the normal yield from investment.

The Cause of Biased Technological Change

The view is attractive because it is economic, accepting scarcity and thinking about it. But it is little more than a speculative guide to research. Most attempts to think about the economics of changes in the production function have a speculative air. For instance, it is commonly argued that labor scarcity relative to land will result in a labor-saving bias in improvements of productivity. In other words, the unit isoquant will move inward more toward the labor axis than toward the land axis. The argument appears plausible on the surface, but it has flaws.

Q: A maker of rifles in America in 1850 faced a higher price of labor relative to materials and machines than did a British maker. *True or false:* Clearly, the American would want to save labor dollars more than materials dollars.

A: A dollar saved is a dollar earned, no matter which dollar. If inventive activity is free, every dollar to be saved would be saved. As expressed, therefore, false.

The argument can be rescued by getting down to the scarcity involved. Suppose that, if the American maker devoted a month of tinkering to try to invent things, he could save two hours of labor per 100 rifles at 10 cents an hour. If he devoted a month to saving wood he could save on each gun an eighth of a board-foot of hard wood at \$10 per 1000 board-feet. The figures reduce to a labor savings of 20 cents per 100 rifles and a wood saving of 12.5 cents per 100 rifles. That is, at 10 cents an hour for labor and \$10 per 1000 board-feet of hardwood, and given the abilities of the inventive tinkerer, it would be best for him to spend the month on labor saving. A British tinkerer with similar gifts but facing a higher price of wood relative to labor might well choose to spend it on wood saving. In other words, one can posit a production possibility curve of reductions in labor and wood inputs caused by a month of tinkering, against which is set a budget line embodying the relative prices of labor and wood (see Figure 23.3).

The Problem of Spillovers

The amount of technological change, then, can be viewed as the return on investment and the direction of technological change as a result of relative factor prices. The trouble with applying the theory to America and Britain around 1860 is that technology even then was international. If a Yankee tinkerer developed a machine to save labor in rifle making the machine would be available even to labor-rich Britain. Since after it was invented it would cost nothing to apply, the British industry would exhibit the *same* labor-saving bias as the American. America, too, would use Britain's wood-saving inventions. There might be a worldwide bias in technological change or a worldwide response of invention to profit, but the alleged differences between countries in their technologies (as distinct from their choice of position along a given technology) are not so easily explained economically.

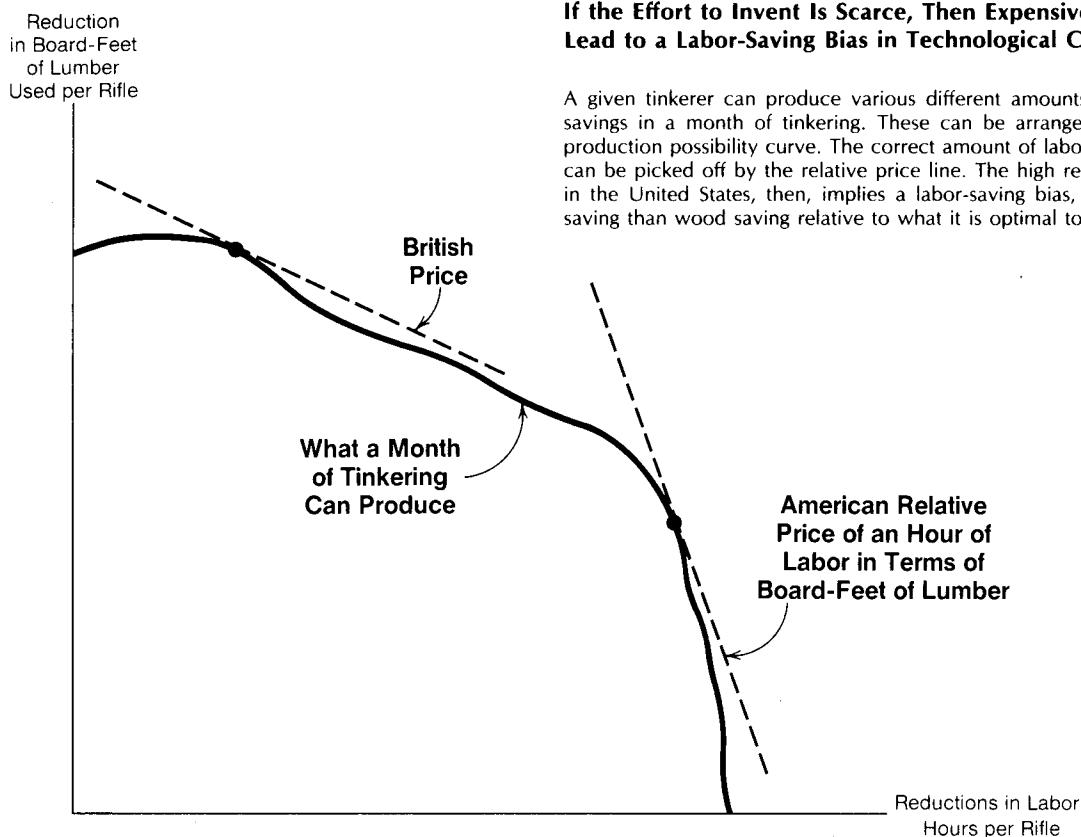


Figure 23.3

If the Effort to Invent Is Scarce, Then Expensive Labor Does Lead to a Labor-Saving Bias in Technological Change

A given tinkerer can produce various different amounts of labor or wood savings in a month of tinkering. These can be arranged in the form of a production possibility curve. The correct amount of labor and wood to save can be picked off by the relative price line. The high relative price of labor in the United States, then, implies a labor-saving bias, that is, more labor saving than wood saving relative to what it is optimal to do in Britain.

Knowledge Is Elusive as an Object of Economic Analysis

What makes the economics of technological change difficult is that it is the economics of knowledge. Though scarce, knowledge is expensive to keep private and like all externalities, therefore, is difficult to analyze. Knowledge spills over from American to British rifle makers and therefore is less likely to be produced in an optimal amount. Patents for inventions or copyrights for books attempt to internalize the externality. But they too run afoul of economic logic.

T or F: Since knowledge is costless once it is produced, the optimal price for the use of knowledge is not what the holder of the patent or copyright can extract; it is zero.

A: The use of knowledge does not have an opportunity

cost. My use of the idea of transistors does not reduce the opportunity for you to use it. Knowledge is like national defense. Not only is it difficult to exclude another person from enjoying the plays of Shakespeare or the protection of the Navy, but to exclude another is inefficient. Therefore, true.

The problem is, of course, that transistors, national defense, and Shakespeare's plays are—or were—costly to produce. Once they are produced, their use is costless, but they will not be produced if no one pays the cost. The case is similar to that of the bridge. The marginal cost, and therefore the appropriate

toll, is nearly zero, but the average cost is high. The social optimum requires that only *marginal* cost be charged for using the bridge of knowledge and that a charity or a state pay for its construction. For this reason universities depend on endowments and state funds to support their research. We would be sunk in barbarism if the charity of kings and the passion of scholars had not, quite irrationally, constructed knowledge for all to use.

Summary

The ideas of constant returns to scale and the Fundamental Theorem find use in measuring the rate at which the production function changes. The measure is a generalized average product, the change in output per unit of input. The rate of change of input is added into one number by factor shares, these being the elasticities of each factor. An identical measure, this time a generalized marginal product, can be derived using prices instead of quantities. Whether price dual or quantity primal, the result applies to all manner of issues, from productivity change in a single industry to the growth of a nation. To go beyond measurement to explanation, however, is difficult, because the economics of technological change is the economics of knowledge. Knowledge is the most typical public good, and it is a wondrous thing that public goods are produced at all.

EXERCISES FOR SECTION 23.2

1. The islands of Stultior and Sapientior have Cobb–Douglas production functions for island income that depend on inputs of labor, L , and land, T , so $Q = AL^{0.5}T^{0.5}$. But they do not have the same A . In 1990 the inputs and outputs of the islands were

For Stultior:

Output: 1,000,000 units of income
 Labor input: 100,000 man-days
 Land input: 30,000 acres

For Sapientior:

Output: 2,000,000 units of income
 Labor: 220,000 man-days
 Land: 40,000 acres

- a. Which island has the higher output per man? Which the higher output per acre?
- b. Which has the higher A ? What is it for each island? (*Hint:* Use the square root key on your calculator to figure out what A must be for each island, given Q , L , and T for each.)

2. One could view the Stultior/Sapientior issue as one of percentage differences. Use Stultior's levels as the base for the percentage changes. Sapientior gets 100% more output from 120% more labor ($220,000 \div 100,000 = 2.20$, which is an increase of 120%) and 33% more land. If Sapientior were Stultior in the future, what would be the percentage of total factor productivity change? (*Remember:* The shares of both inputs are 0.5.) Why is the figure somewhat above that of Exercise 1? (*Hint:* Try the other base.)

3. For 10 years after 1990 output in Stultior grows at 2.06% per year while land stays unchanged and labor input grows 1% per year. What is the annual rate of total factor productivity change? In 10 years how much will the production function have risen? If Sapientior meanwhile has not advanced, what is the result?

PROBLEMS FOR SECTION 23.2

- 1. During the 1870s, 1880s, and 1890s, Britain is said to have failed economically. In particular, British business managers are supposed to have been slow to adopt new technologies. The data of output, capital, and labor are as follows:

	GNP at 1900 Prices (millions)	Capital Stock (millions)	Labor Force (millions)
1870	1000	4320	13.06
1900	2040	6660	17.74

Assuming that the share of capital was 0.44 and that the share of labor was 0.52 (the rest, 0.04, was rent to land, but the rate of change of land can be assumed to be zero), what was total factor productivity change 1870–1900? If American productivity change in the same period was about 1.5% per year, how much of a British failure was there?

2. In 1348 (before the Black Death and subsequent plagues), the population of England and Wales was 4 million. By 1400 it had fallen to 2 million. Real wages had risen 40%, wages were always about half of national income, England and Wales were overwhelmingly agricultural countries, and no new land or capital was employed in agriculture over the half century. *True or false:* One can conclude from these facts that technological change was nil in England and Wales and that the fall in population by itself explains the rise in wages.

True or False

- 3. Without the assumption of constant returns to scale, the unit isoquant cannot be drawn.
- 4. In the diagram of the unit isoquant (Figure 23.2), the American point is revealed preferred at American prices to the British point, and likewise at British prices the British point is revealed preferred to the American.
- 5. The distance between two unit isoquants is subject to an index number ambiguity.
- 6. The rate of change of real wages is equal to the rate of change of output per person.
- 7. If productivity change occurred at the same rate in all industries, the relative prices of all goods would remain constant.
- 8. In the absence of well-enforced laws of patents internationally, a devotion to science in one country is irrational, at any rate if it is viewed merely as an investment in economic well-being.
- 9. If an invention (such as the cotton gin) raised the value of some asset (such as cotton land or slaves) and only the inventor knew about the invention before it was announced, then patents would be unnecessary to induce inventors to invent.

CHAPTER

24

Misallocation and Monopoly in Factor Markets

51,000,000
for 2 street
corner
in Haarlem

24.1 Good and Bad Allocation

What to Read For Is the value of the marginal product of labor higher in capital-intensive industries? How is the value of marginal product used to think about who benefits from technological improvements? Why is overfishing bad? Will fisheries that are not owned by someone be used correctly? In what way is sharecropping like a tax? What feature of the contract prevents it from having the efficiency effects of a tax?

Values of Marginal Product Are Equalized Across Industries

Marginal productivity is not a matter of engineering and production functions but one of sociology and markets. Thinking of it in engineering terms leads to error.

T or F: If the ratio of capital input to labor input is higher in the chemical industry than in the steel industry, then according to marginal productivity the wages of labor in chemicals will be higher than in steel.

A: It is meaningless to say that the marginal physical product would be "higher" in the capital-intensive industry, because chemicals and steel are different products. What matters is the *value* of the marginal product (the price per ton multiplied by the tons per hour), which is not higher in chemicals merely because chemicals choose capital-intensive methods. It might be

higher for an instant. But if it were, the higher wage would attract workers out of the steel industry and into chemicals until the values of the marginal products and therefore the wages became equal. If chemical workers do make more than steelworkers, it is because they are more valuable as workers in alternative occupations—better educated, say, or stronger—not because they work with more expensive machinery per person. The amount of machinery and land per person affects the marginal product of labor in the economy *as a whole*, but not in a single industry itself.

The principle applies generally. For instance, workers of a given quality in the computer industry, which did not exist before 1950 and has since then had faster productivity change than any other industry, earn the same amount (with some allowance for momentary disequilibrium) as do workers in the university industry, which has existed since 387 B.C. and has never experienced any productivity change at all. It is a good thing, too. If the value of marginal product were different in the two industries, then something could be gained by shifting workers from the high-wage to the low-wage industry. Marginal productivity operates within a market, bribing factors to flow to their highest reward.

Uses of the Equalization of Values of Marginal Products

In the usual way of selfishness in economics, the pursuit of high rewards in a market maximizes national income. Suppose, for example, that the village of Vilyatpur in northern India experiences the Green Revolution, that is, the coming of dwarf varieties of wheat that can hold more grain on their short stalks than a normal plant. The Revolution can be viewed as raising the curve of the value of marginal product for labor (and for land, capital, materials, and other things; but focus on labor).

Q: 1. If the number of laborers in Vilyatpur remains the same, what does the Green Revolution do to the wage and the village output? Draw a diagram of the value of the marginal product of labor before and after the Revolution. Identify the area of increase in output, assuming that labor remains the same (refer to Figure 24.1).

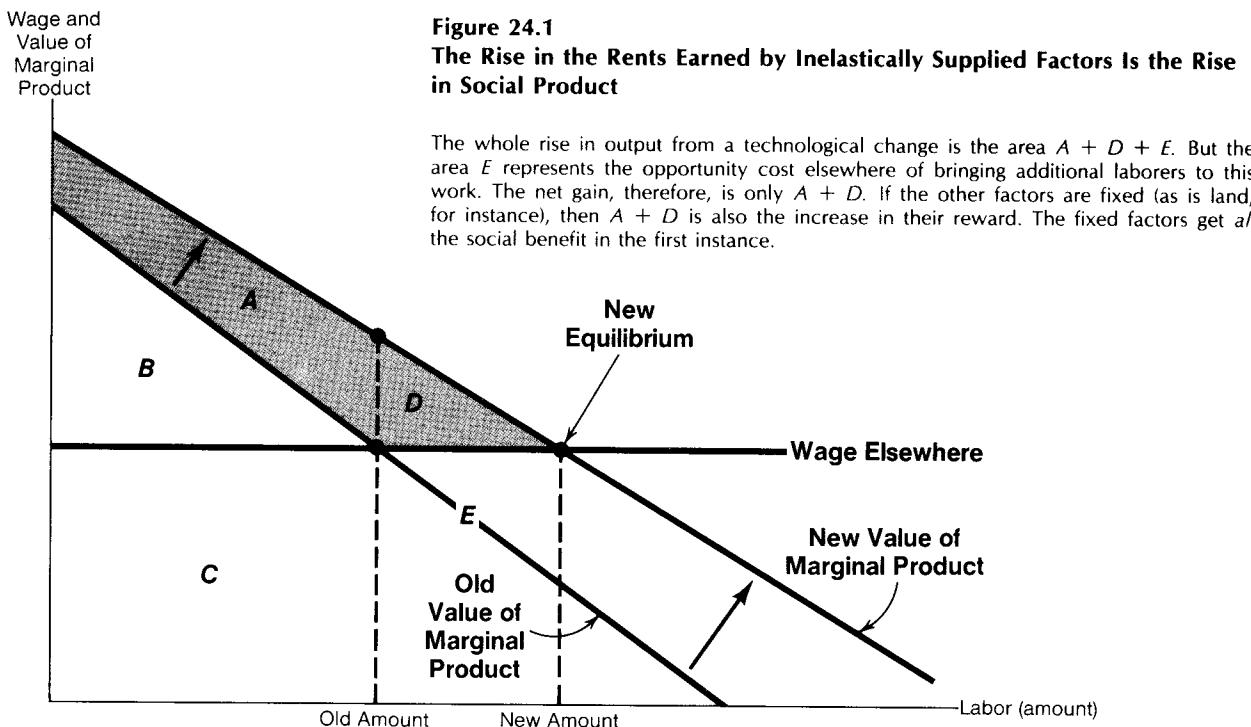
2. If Vilyatpur faces an elastic supply of labor at the old equilibrium wage and if labor flows into the village in response to the initial rise in wages, what is the whole additional output in the village? How much is earned by labor on the one hand and by all other factors on the other?

3. How is output elsewhere in the economy affected by the inflow of labor to Vilyatpur? What, therefore, is the *net* social gain from the Green Revolution coming to Vilyatpur and from allowing labor to flow into the village in response to it? Who gets all the gain—labor or the other, *inelastically supplied* factors, such as land?

A: 1. Clearly, the wage will rise, since the fixed labor is worth more. So too will output rise, productivity being higher. Labor's value of marginal product (which is the price of wheat multiplied by the marginal physical product of labor) rises because the marginal physical product has risen. At Old Amount of labor the output increases by the area *A*, since the output is now the area under the higher marginal product curve out to Old Amount.

2. The Wage Elsewhere is the elastic supply of labor. If labor flows into Vilyatpur, the new equilibrium will be at New Equilibrium. Output is now the area under the higher marginal product curve out to New Amount. That is, it is the fringe area *A + D + E*. Labor (including the new labor) earns the rectangle *C + E*, namely, the unchanged wage multiplied by the New Amount of labor. The nonlabor factors (call them land) earn what is left, the triangle *A + B + D*. Land earned *B* before; it earns *A + D* more than *B* after.

3. The new labor earned *E* outside of Vilyatpur, which means that its marginal product was *E*. That is, output in the rest of the economy went down by the amount *E* when the labor moved to Vilyatpur. The whole higher output was just found to be *A + D + E*. The *E* part of this must be subtracted to avoid double counting, for it is not a *net* gain. The net gain therefore is the shaded area *A + D*. *The area is the same as the rise in rents* (that is, in nonlabor incomes, supposing the nonlabor factors of production to be supplied inelastically, not elastically, to the village). *The inelastically supplied factor gets all the gain* because land is immobile, and therefore no more land can move to Vilyatpur to dissipate high rents. The result is a familiar one. Inelastic participants in a market pay a tax imposed on the market; or urban landlords, not tenants, win from cleaner air; likewise, inelastically supplied factors earn the benefit or suffer the hurts of a change in agriculture.



A Failure of Equalization Is Inefficient

Not allowing the labor to flow into Vilyatpur would prevent landlords from earning the benefit, but it would leave *national* income lower by the difference between the marginal valuation of labor in Vilyatpur and in the rest of the nation, namely, by the area D . The area is identical to the social loss from forbidding exchange in other contexts, such as housing or gasoline.

The general case in which Vilyatpur is not small in relation to the labor market it faces involves a diagram of many uses. Consider the allocation of the whole amount of labor in Iceland between two industries, fishing and all other industries.

Q: Draw a diagram with the whole amount of labor on the horizontal axis and the two curves of the value of marginal product facing each other, each with its origin at opposite ends of the line representing the amount of labor. What is the optimal allocation of labor? At the optimum, what is Icelandic national income? If for some reason too much labor were allocated to fishing, what would be the lost income?

A: Following the instructions dutifully, the diagram (ignore the dashed line) appears in Figure 24.2. The optimal allocation of labor is the allocation Best, in which

the values of marginal products in the two industries are equal. Icelandic national income is the income earned in fishing plus the income earned in all other industries. That is, income is the entire M-shaped area under the two marginal product curves. It is evident that any movement away from Best is going to result in less income. At Bad, for example, the shaded area is lost to income. Fishing is pressed so far that the value of marginal product in fishing is well below that earned in all other industries. A transfer of labor into all other industries would capture the higher product, but the transfer does not occur. There is overfishing.

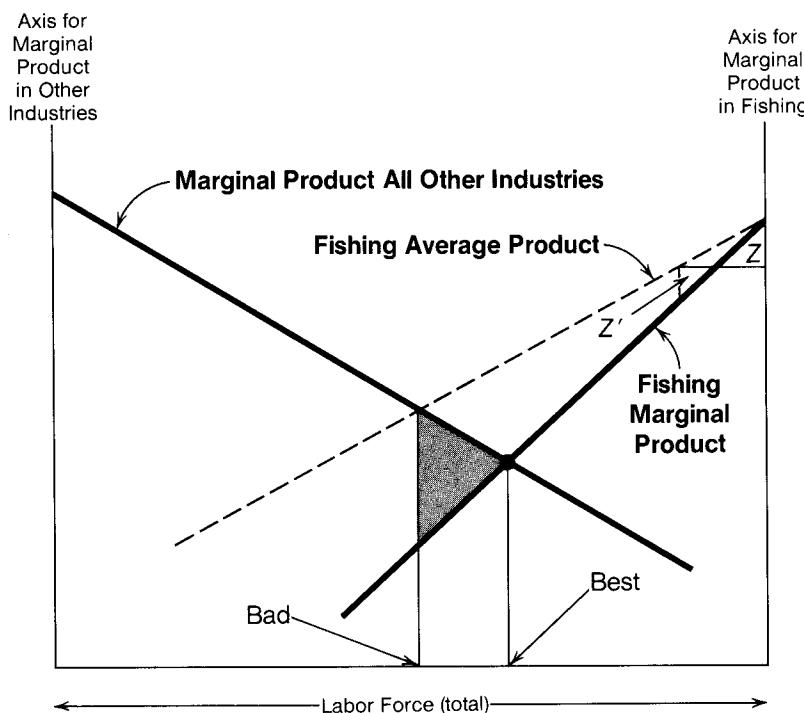


Figure 24.2
The Lack of Ownership of the Sea, or Some Other Distortion, Causes Labor to Be Misallocated and Income to Fall

The shaded area is the loss from a Bad allocation of the labor force relative to the Best allocation. One way in which the Bad allocation can come about, pictured here, is the overuse of the sea, the fisheries problem. The problem is that if the sea is not owned by someone no rents are charged for a scarce resource, and it is treated as free. Fishermen go to sea as long as the output per person (the average product of labor) is above the marginal product in other occupations. Too many go to sea, overfishing the fish and driving the marginal product down below its value elsewhere.

The Fisheries Problem Is the Failure to Charge Rent

One possible cause of overfishing is contained in the diagram. To be concrete, suppose that the other industry in Iceland is the farming of land. Icelanders either farm the land or farm the sea. But suppose further, as was the case before Iceland extended its territorial limits 200 miles into the ocean, that nobody owns the sea. No landlord can charge rent on the use of the sea, with the result that all the output goes to labor ("labor" here, as in the Vilyatpur problem, is meant to stand for all mobile factors). In other words, labor in farming the land earns only its marginal product, leaving a remainder to be earned by landlords, while labor in fishing earns its average product (output divided up among the people fishing), leaving no remainder. The fish in the sea are assumed here to be scarce (the average and marginal product of fishing does fall noticeably as Iceland fishes more). But because private property does not extend to them, the fish in the sea are not *treated* as though they are scarce. Unsurprisingly, misallocation is the result.

The result can be understood by looking closely at the diagram. The usual relation between average and marginal curves is portrayed up in the corner of the fishing curves by the equality of the areas Z and Z' . If labor is paid its average product, Q/L , all the product will go to labor—as is obvious from the algebra that $(Q/L)(L) = Q$. People (equipped with ships and nets) will keep entering fishing until the *average* product in fishing is equal to the marginal product in other industries. Such an allocation, however, is bad.

The case is known in economics as the *fisheries problem*, or the *common pool problem*.¹ The problem is that the fish are overfished if they are not owned and merely sit in a common pool for anyone who can catch them first. It has elements of the prisoner's dilemma. If Icelanders and other fishing nations could agree on quotas, the fish in the ocean would be treated as scarce. But each fisherman has an incentive to cheat on the agreement, with the result that all nations overfish.

Social Arrangements, Not Biology or Engineering, Cause the Problem

Notice that it is not the biology that makes it a problem. That overfishing can interfere with the efficiency of breeding among the fish, say, and in the extreme make them extinct is just one reason among several for the marginal product to decline with much fishing. But the problem of excessive "fishing" arises in exploiting minerals as well. The Ogallala Aquifer, for example (effectively a lake underlying much of the dry lands of Nebraska and neighboring states) is owned by no one. Anyone can take water out of it for the cost of drilling. As a consequence the water level is falling fast, it is said, and will soon finish irrigated agriculture in the Central Plains. That the water is "nonrenewable" is not to the point, however much the word excites the engineering mind. Resources are to be used, not "conserved for our grandchildren." But the lack of ownership leads to the water being used too fast.

Taxation, Too, Attenuates Ownership and Causes Misallocation

Any attenuation of ownership will cause a misallocation of resources. Taxation of wages in agriculture, for example, will reduce the incentive to apply labor to the land. The laborer will not own all the (marginal) fruits of his labor. The laborer will apply labor to the acre of land to the point at which the value of marginal product after taxes is equal to the wage in alternative, untaxed occupations before taxes. Too little labor will be applied to the acre.

Sharecropping as Analogous to Taxation

One must, however, step with care. Consider the following important problem. The landlord of an acre of land, Pamela Nickless, can profit from it in three alternative ways. She can hire Joe Reid as a laborer, supervising him in working the land directly; she can hire out the acre to Reid for a fixed amount per year, letting him keep for himself whatever he can grow in excess of the rent; or she can enter into a "sharecropping" agreement with Reid. Instead of getting wages or getting the excess over rent, under sharecropping Reid gets a share of the crop, the share being negotiated with the landlord. Suppose that the agreed share was 50:50. If Reid's seventieth hour applied to the cotton crop in picking time yields a marginal product of \$3.00 worth of cotton, then Reid would keep only \$1.50, giving the other half to Gray.

The situation looks much like a tax on labor in agriculture. Pamela's fifty percent share of the crop is like a tax on Reid. If the tax analogy is correct, then it follows that sharecropping will be a bad system, as bad as taxes. Like

¹ Anthony Scott, "The Fishery: The Objectives of Sole Ownership," *Journal of Political Economy* 63 (April 1955): 116–124. Compare the earlier discussion of Frank Knight's case of the Dan Ryan Expressway and the city street.

taxes, it will discourage Reid from working the cotton to the point where his full marginal product is equal to what he can earn per hour elsewhere.

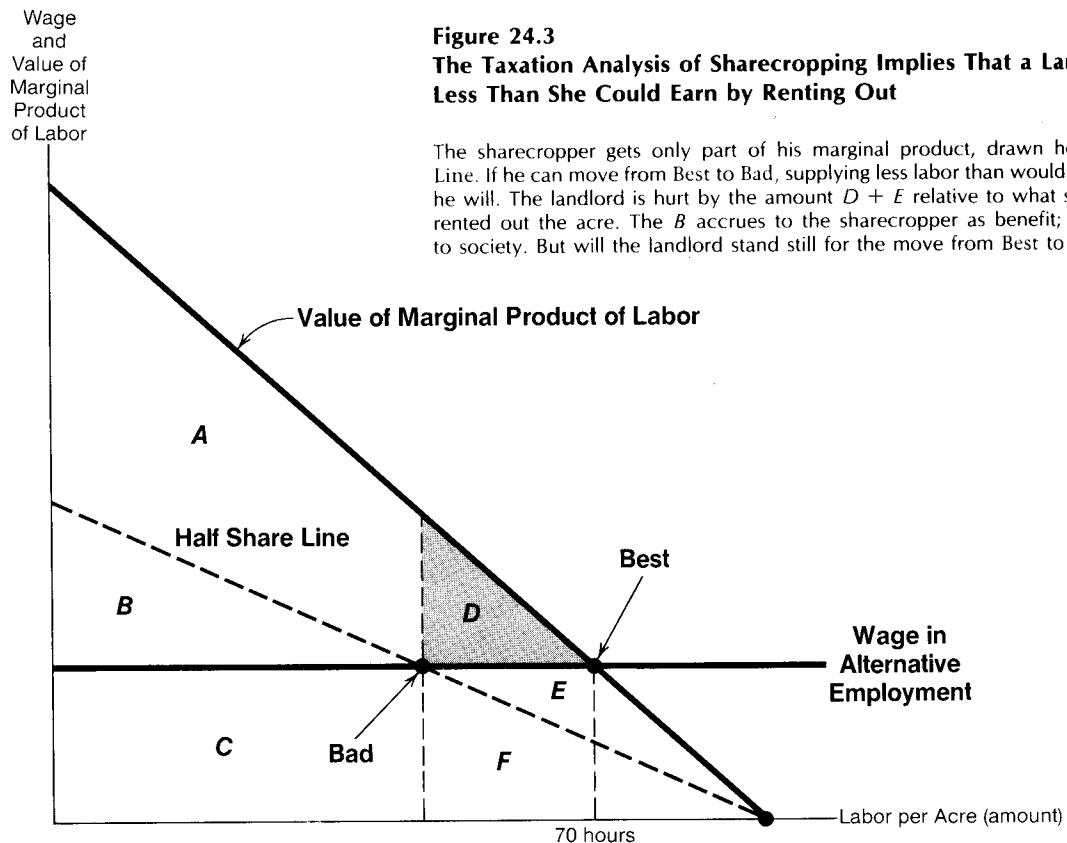
Q: Draw the diagram for a 50% tax imposed on a declining marginal product of labor on a single acre. What net proceeds does the taxed laborer face? If his wage in alternative occupations is given, what amount of labor will be supplied to the acre? What is the deadweight loss from the tax?

A: The tax cuts the curve of marginal product in half at every amount of labor (see Figure 24.3). The dashed line of after-tax marginal product (the Half Share Line) is simply pivoted around the intersection of the before-

tax marginal product on the horizontal axis (since half of zero is zero, the marginal product before and after the tax is the same only when the marginal product is zero). The taxed laborer faces the dashed Half Share Line. Reid will only supply labor out to Bad, leaving the shaded area *D* as the deadweight loss from the tax. The cotton will get too little time, the acre will get too little profit, and the society will get too little cotton. Many people find sharecropping distasteful, and the tax analogy seems to support the distaste with analytic rigor.

The Analogy Is False

Or so it seemed to many excellent economists for an embarrassingly long time. Looked at coolly, the facts were always a little disquieting. Sharecropping is used by lawyers and other business people to organize their own affairs. Strange. Sharecropping characterizes some rich agriculture, such as Iowa, as well as poor, such as India. Stranger. Sharecropped and rented land otherwise identical have



**Figure 24.3
The Taxation Analysis of Sharecropping Implies That a Landlord Will Earn Less Than She Could Earn by Renting Out**

The sharecropper gets only part of his marginal product, drawn here as a Half Share Line. If he can move from Best to Bad, supplying less labor than would be socially desirable, he will. The landlord is hurt by the amount *D* + *E* relative to what she could earn if she rented out the acre. The *B* accrues to the sharecropper as benefit; the *D* is simply lost to society. But will the landlord stand still for the move from Best to Bad?

identical outputs per acre, contrary to the implication of the tax model that output per acre would be lower under sharecropping. Stranger still. And sharecropping and renting (and owner occupation) coexist, sometimes with the same landlord renting one acre to Steven Cheung and letting out on shares another identical acre to Joe Reid. Strangest of all.

The diagram shows that the landlord will get the whole triangular area $A + B + D$ in rent if she rents out the acre at a fixed rent, giving the tenant no incentive to hold back his labor. Likewise, it claims that the landlord under sharecropping gets only the area A , her share of the whole output $A + B + C$. An acre under sharecropping earns less than its value in alternative employment (namely, in renting) by the difference, $B + D$. The hours of the sharecropper's labor, on the other hand, earn more than their value in alternative employment by the area B , since he can earn $B + C$ being a sharecropper part time and area $E + F$ being employed elsewhere for the rest of his time, which is in total more than the rectangle $C + E + F$ he earns if he rents. Something is wrong. All landlords would want to rent at a fixed rent and all laborers would want to enter a sharecropping agreement. The two systems could not coexist. But they do.

What is wrong with the tax analogy is that the landlord, unlike the tax collector, does not stand still for the withdrawal of Joe Reid's labor from the cotton field. She simply makes it a part of the sharecropping agreement that Reid must work out to Best, regardless of whether his share of the \$3 he makes putting in that hour equals his alternative wage. She is able to get away with such a requirement so long as Reid in following the requirement is even a cent better off than he would be in alternative employment as a renter with her, as a worker on someone else's acre, or as a worker outside of agriculture altogether. If renting and sharecropping coexist, the Share Line must swivel around until triangle B is equal to triangle E , that is, until what the landlord earns from sharecropping an acre, $A + D + E$, is equal to what she can earn from renting out an acre, $A + D + B$.

The upshot is that sharecropping and renting have *identical* outcomes. *It is not the case that sharecropping is always the inferior system*, or that one or the other party is exploited more under one system than under the other. As long as landlords can enforce the work clause of the agreement (and such clauses in fact exist in sharecropping agreements), the share will vary until both sides of the market are indifferent between the two. Why one is chosen over the other is unclear—perhaps one system has advantages of less riskiness or lower supervision costs than the other. But the gross differences that are implied by the tax analogy do not exist.²

² The modern position on sharecropping was first stated by Steven N. S. Cheung, in *The Theory of Share Tenancy* (Chicago: University of Chicago Press, 1969), especially pp. 42–55. It was stated more explicitly in J. D. Reid, Jr., "Sharecropping and Agricultural Uncertainty," *Economic Development and Cultural Change* 24 (April 1976): 549–576. In truth, much of the modern position is contained in the two pages of Alfred Marshall's treatment, *Principles of Economics*, 8th ed. (London: Macmillan, 1920), Book VI, Chapter X, Part 5, pp. 644–645: "His landlord has to spend much time and trouble, either of his own or of a paid agent, in keeping the tenant to his work. . . . If worked out thoroughly, it will result in the cultivation being carried just about as far and affording the landlord the same income as he would have on the English plan (i.e., fixed rent) for equally fertile and well-situated land equipped with the same capital, and in a place in which the normal ability and enterprise of candidates for farms is the same."

The Moral: One must be wary, then, of claims of gross inefficiency in the allocation of resources. Someone is hurt if income is lower than it could be, giving the someone an incentive to remedy the inefficiency by evading the law or by creating the nonexistent property right or by amending the faulty contract that gave rise to the inefficiency in the first place. The argument is no warrant for supposing that we live in the best of all possible worlds. But it is ample warrant for supposing that tales of inefficiency are more complex than even economists believe them to be.

A good example is intellectual overfishing of the fisheries model. In one variant it is called "the tragedy of the commons," the tragedy being that grazing on the nonprivate land in English villages long ago is supposed to have resulted in overgrazing, like overfishing, and a tragic loss of output. The alleged history has been used as the standard case of ravaging of resources and the difficulty of cooperation. The history is wholly mythical, however, on two counts. For one thing, the land was private. Medieval villages knew property. For another, when it was threatened by overgrazing the village adopted rules called "stints" that prevented the inefficiency. To suppose that a village of 30 or 40 families would sit by and watch its grazing land be ruined by lack of private property or of public rules is ersatz economics in the guise of economic analysis.

Summary If factor markets are functioning as they should, then workers of the same quality will receive the same wage, regardless of their location. The self-interest of the workers in seeking the highest pay will allocate the workers to their best employment until at the margin all products are equal. The factor market simulates the rule of rational life.

The argument can be used to answer the question of who benefits from an increase in agricultural productivity, or the question of how much is lost by putting too many fishermen in one ocean, or the question of whether sharecropping is a bad system in agriculture. The Green Revolution case illustrates the theorem in taxation that the inelastic party gets the benefit or bears the cost. The fisheries case illustrates why scarce resources must be treated as scarce, that is, must be priced. The sharecropping case illustrates that one must look below the machinery of marginal productivity to the human incentives and market opportunities that drive it. An analysis that leaves some people worse off for no good reason is not much of an analysis. The economist should be as wary of inexplicable starvation as of the free lunch.

EXERCISES FOR SECTION 24.1

1. Stanley Lebergott works on a 150-acre Sac City corn farm owned by Helen Hill Updike. In 1920 the value of his marginal hour at various number of hours per year was the column marked Old:

Hours of Lebergott Labor per Year	Value of an Additional Hour	
	Old (1920)	New (1940)
0	0.56	0.65
200	0.53	0.62
400	0.50	0.59
600	0.47	0.56
800	0.44	0.53
1000	0.41	0.50
1200	0.38	0.47
1400	0.35	0.44
1600	0.32	0.41
1800	0.29	0.38
2000	0.26	0.35
2200	0.23	0.32
2400	0.20	0.29
2600	0.17	0.26
2800	0.14	0.23
3000	0.11	0.20
3200	0.08	0.17

- a. If Lebergott's earnings in alternative employment were 20 cents an hour—that is, if Lebergott supplied his hours at 20 cents an hour—how much of his labor would Updike buy? How much per year would he cost her?
- b. At this amount how much *total* revenue per year would Updike earn? (*Hint:* Add up the marginal revenues for each successive hour. Make sure you include all 200 hours between each step.) If his labor is her only expense, how much therefore does Updike earn in profits per year?
- c. What would Updike earn in profits if she employed Lebergott (and no one else) only 1000 hours a year? How much does she lose by making this mistake?
- d. If Lebergott has 3200 hours available, what does he do with his excess time? What is his *full* income?
2. Suppose that the coming of hybrid corn and tractors by 1940 has pushed the value of the marginal product of Lebergott's labor out to the column marked New. Suppose that Lebergott still earns \$0.20 in alternative employment.
- a. How many hours of Lebergott's time does Updike employ? How much does Lebergott earn from Updike? How much is his *full* income? Has it increased from 1920 to 1940?
- b. How much is the gross revenue in 1940 of the Updike farm? How much does Updike get in net revenue? (Again assume that Lebergott is the only expense.) Who has gained from hybrid corn and tractors in the first instance? How much?
3. Updike's gain was gain "in the first instance." Suppose (as was the case) that millions of other American farmers took advantage of hybrid corn and tractors between 1920 and 1940. What would happen to the amount of corn produced? What would happen to the price of corn? What would this in turn do to the curve of the value of marginal product?

What therefore would happen to Updike's net earnings? Who benefits from the fall in the price of corn caused by the inventions? Who therefore gains in the last instance? Updike? Lebergott? Both as consumers of corn?

PROBLEMS FOR SECTION 24.1

1. The value of the marginal product of labor in Indian wheat growing is the marginal physical product times the price of wheat. The coming of the Green Revolution to Vilyatpur alone obviously does not much affect the whole supply of wheat and therefore does not affect the price. What happens to the analysis in the text if 50,000 Vilyatpurs adopt the Green Revolution? Who benefits?
- o 2. In the eighteenth and earlier centuries England's ancient system of "open fields" was eliminated (enclosed). The impact was similar to the Green Revolution in type if not in magnitude: The marginal product of labor (and of land, but focus on labor) moved out. The enclosure movement was slow, affecting only a tiny portion of England's villages in any one year. The facts on output before and after enclosure are very poor. Rents of land, however, are easily available. *True or false:* The rise in the rent on a village enclosed in 1788 is equal to the rise in output attributable to enclosure.
3. Is it to the advantage of employers as a class to obstruct or assist the mobility of labor? One's instinct might be to answer "yes," but it is in fact more ambiguous. In answering the question, consider the following simple case. There are two regions with a given labor force to be distributed between the two. The only other factor of production is land, and it is immobile between the two regions. Employer-landlords own land everywhere (that is, each one has a portfolio of land in both regions). Initially, before mobility, there is a certain allocation of the labor force, *not necessarily optimal*
 - a. Set up the model in a diagram, assuming no labor mobility to start with. Identify the reward to labor as a whole and to land as a whole.
 - b. Construct an example with the sort of diagram you used in (a) of a case in which the introduction of mobility *increases* the reward to land as a whole and one in which it *decreases* it.
 - c. Can you make any general *ceteris paribus* statements about the conditions under which the total returns to land are more likely to rise than to fall when mobility is increased?
 - d. What mobility policy is in the interest of each individual employer-landlord if they own a portfolio of land concentrated in one region? Would concerted class action be likely in this case? What do you conclude about the fruitfulness of the model of class interests in explaining various obstructions to labor mobility (such as the Settlement Laws in England in the seventeenth through the early nineteenth centuries)?
4. The payment of a wage to cowboys equal to the wage the cowboys could earn elsewhere can be justified on the grounds that if it were lower cowboys would be overused (socially speaking) in herding cows. *True or false:* The payment of a pure rent to owners of cattle land in excess of what the owners could earn from the land in alternative employments (as garden plots, say) can be justified on the same grounds; that is, without the rent the cowboys would be overused.
- o 5. Economists are fascinated by allocation and delight to find in misallocation corrected the cause of economic growth. Tariffs, taxes, racial discrimination, entrepreneurial sloth, and so forth put resources in the wrong places. But the loss of income—and therefore the gain of income from eliminating the misallocation—can be tiny even in what would seem to be favorable cases.
 - a. Suppose that the economy is divided into halves by tariffs, racial discrimination, linguistic prejudice, caste, distance, mountains, habit, or whatever. Each half has the production

This

function $Q = cK^{0.5}$, where Q is the homogeneous regional product and K is the region's capital stock (or labor force or whatever: the only essential point is that the other factors of production are held equal in the background, collapsed into the constant c assumed to be equal in both places). This particular production function is chosen partly because it represents reasonably well the usual facts and partly because it can be manipulated with the otherwise useless square root key on your calculator. The units of Q and K can be chosen such that c is 1.0—that is, when $K = 1.0$, Q also = 1.0. Let the total capital stock to be allocated between the two regions be 2.0 units. In view of the symmetry of this division of the economy, what do you suppose is the optimal allocation of the capital stock between the two regions? What is the corresponding *national* income? This is the maximum attainable income, isn't it?

b. If you have had calculus, you know that the marginal product of capital in a region will be $dQ/dK = (0.5)K^{0.5-1.0} = (0.5)K^{-0.5}$. If you have not had calculus, take this on faith. Now, a very large difference in factor rewards for mobile factors of production is a difference of 100%, that is, wages in one region twice what they are in another or interest rates in one social class twice what they are in another. It is a large difference because a lot of money can be made overcoming it. Migrants from Europe to the United States in the nineteenth century, for instance, doubled their incomes for their (considerable) trouble. In view of this, show that an allocation of 0.4 of the total capital stock of 2.0 units to one region and 1.6 to the other will produce this degree of difference (twice) in the factor rewards. What is national income with such an allocation? How does it compare with the income attained under the correct allocation? Show that the correct allocation is in fact correct. If you were attempting to explain a doubling of national income over 50 years or so, would you put your intellectual faith in the power of reallocations to explain it?

6. Tithes were taxes on agricultural output for the support of the church, by tradition 10% of the value of output. By the eighteenth century they had existed unchanged for many centuries.

a. *True or false:* "No [eighteenth-century] landlord could honestly believe that the [tithe] payments robbed them of any parts of the rents to which they were justly entitled."

b. *True or false:* "Nor could any tenant honestly complain that tithes increased the burden of his rent."

c. Does it follow from your answers to (a) and (b) that no one would have benefited from the elimination of tithes without compensation to the church?

d. "But the real practical grievance was . . . [that a tithe was] a charge which increased by good farming, or diminished by bad,—a tax on every additional outlay of money and labour,—a check upon enterprise and improvement." Discuss.

- o 7. Alfred Marshall, the great English economist of the turn of the century, was chief among those who believed sharecropping to be analogous to a tax. Marshall's solution to the sharecropping problem is known to be incomplete, even wrong (although on the page after the incomplete solution Marshall himself completed it). Yet an economist studying sharecropping in northeastern Brazil (where the crop share is about 50%) applies Marshall's model without getting Marshall's results. That is, he develops a model in which landlords do *not* specify and enforce the amount of labor to be applied to an acre under sharecropping (that is, the sharecroppers themselves decide), fits the model to the facts, and finds by simulation that sharecropping produces a socially optimal solution. In fitting the model to the facts, he alleges that the wage in alternative employment facing sharecroppers (after allowing for costs of search and travel) is only about half the wage paid by landlords to steady workers. Why does he get the result that sharecropping does not lead to inefficiency? What assumption in his model would you advise him to submit to creative self-doubt before he publishes a paper on his finding?

8. A worker in an agricultural community can either work for a fixed wage on the landlord's land or become a tenant on some of the land, paying a fixed rent. In the wage case the worker gets a low but stable income, because the worker bears none of the risk of variability in agricultural income; in the rental case the worker gets a high but variable income, bearing all the risk.

a. Draw the Edgeworth box for this situation, using total agricultural income and total variance as its dimensions. Locate the wage contract and the rental contract.

b. Under sharecropping the worker and landlord share whatever income there may be in some prearranged proportion (instead of one or the other taking a fixed income—a wage or rent—and leaving all the variability to the other). Place a typical sharecropping arrangement in the Edgeworth box. Why is it superior to either a wage or a rental contract alone?

c. Notice the word “alone” in the last sentence. If landlords and workers can engage in some wage work and some renting, together, what do you think of the common proposition that sharecropping has unique advantages in sharing the risks of agriculture?

True or False

9. The speed limit of 55 miles per hour will reduce the productivity of trucking labor but will not reduce its price.

10. A government policy for stopping inflation by prohibiting wage increases that are not matched by productivity increases in the industry will lead to inefficiency.

○ **11.** Cab drivers (keeping them distinct from owners of cabs and the artificially limited licenses to operate them) are rational to support rises in fares.

12. In taking an examination it is desirable to work on a question until your marginal point product of a minute on the question is equal to the marginal point product on every other question.

13. A proportional tax on the rent of agricultural land (that is, on its marginal product) will not reduce the incentive of the last acre to remain in use and will therefore not affect agricultural output.

24.2 Monopoly in Factor Markets

What to Read For

How does monopoly or monopsony affect the markets for factors? What is the difference between the value of marginal product and the marginal value product? Between average and marginal factor costs? How does one measure the monopsony power of the owners of baseball teams? What is the octopus effect?

Smallness Yields Optimality

The Rule of Rational Life for a firm is to hire labor until the marginal cost of labor equals its marginal benefit. If the firm is small in all the markets it faces, then the rule reduces to hiring labor until the wage equals the value of the marginal product of labor (that is, the price of the product multiplied by the marginal physical product). Another hour of labor costs a wheat farm facing a large number of other competitors for the supply of labor the going wage. Another hour earns a farm facing a large number of other suppliers of the demand for wheat the amount of wheat the hour can produce multiplied by the going price of wheat. The smallness of firms in an industry has, as usual, pleasing conse-

quences: The valuation that workers put on the last hour they supply (namely, the wage) is set equal to the valuation that consumers put on the product of the hour. If the last hour of bakers should become more valuable in some other use—if they should develop a passion for Latin, say, and would rather spend the hour studying Latin—then consumers would have to pay more for the bread produced in the hour and would reduce their consumption of it. The society would be induced to spend its hours where they were valued most, in studying more Latin and eating less bread.

Bigness Spoils the Optimality

But of course all is spoiled if the firm is large in one or more of the markets it faces. If General Motors has monopoly power in selling buses it will recognize that selling another bus spoils the price and will therefore respond to a marginal revenue lower than price. In the hiring of labor to make the additional bus, therefore, it will view its benefits as the *marginal revenue product* (the additional revenue, allowing for spoilage, that another hour produces) not the *value of marginal product* (the additional revenue, *not* allowing for spoilage).

Likewise, if General Motors has monopsony power in buying skilled bus designers, it will recognize that buying another hour spoils the wage. The marginal cost of hours is no longer equal to the average cost (namely, the wage) but is higher by the higher wages General Motors will have to pay to *existing* workers in the course of inducing another to join the company.

The upshot is that a firm with both monopoly and monopsony power sets the marginal revenue product (lower than the value of marginal product) equal to the marginal cost of labor (higher than the average cost). The diagram is as shown in Figure 24.4. The four possible equilibria are marked as Perfect Competition, Monopsony Only, Monopoly Only, and Both Monopoly and Monopsony (this being a loosely reasoned diagram in which competitive and monopolistic situations can be compared directly).

Note that the firm (or, when it is a monopoly, the industry) must always pay the Average Labor Cost, or wage: It must remain on the supply curve of labor. The point of equilibrium marked Monopsony Only exists in the mind of the monopsonist, not in observed behavior in the labor market. It is analogous to the point of equality of marginal cost and marginal revenue for a monopolist. Just as the monopolist picks an observed point on the demand curve, so too here the monopsonist picks an observed point on the supply curve (the Average Labor Cost), indicated by the dashed arrow.

The equation mimicking the diagram is

$$\begin{aligned} \text{Marginal revenue product} &= P(MPP_L) \left(1 - \frac{1}{E_Q^D}\right) \\ &= w \left(1 + \frac{1}{E_L^S}\right) \\ &= \text{marginal labor cost} \end{aligned}$$

P is the price of the product, MPP_L the marginal physical product of labor, E_Q^D the elasticity of demand for the product, w the wage (or the average labor cost, itself dependent on how much labor is demanded), and E_L^S the elasticity of the average labor cost (that is, of the supply curve of labor). The appendix to this section gives a proof by calculus, but the equation is plausible without

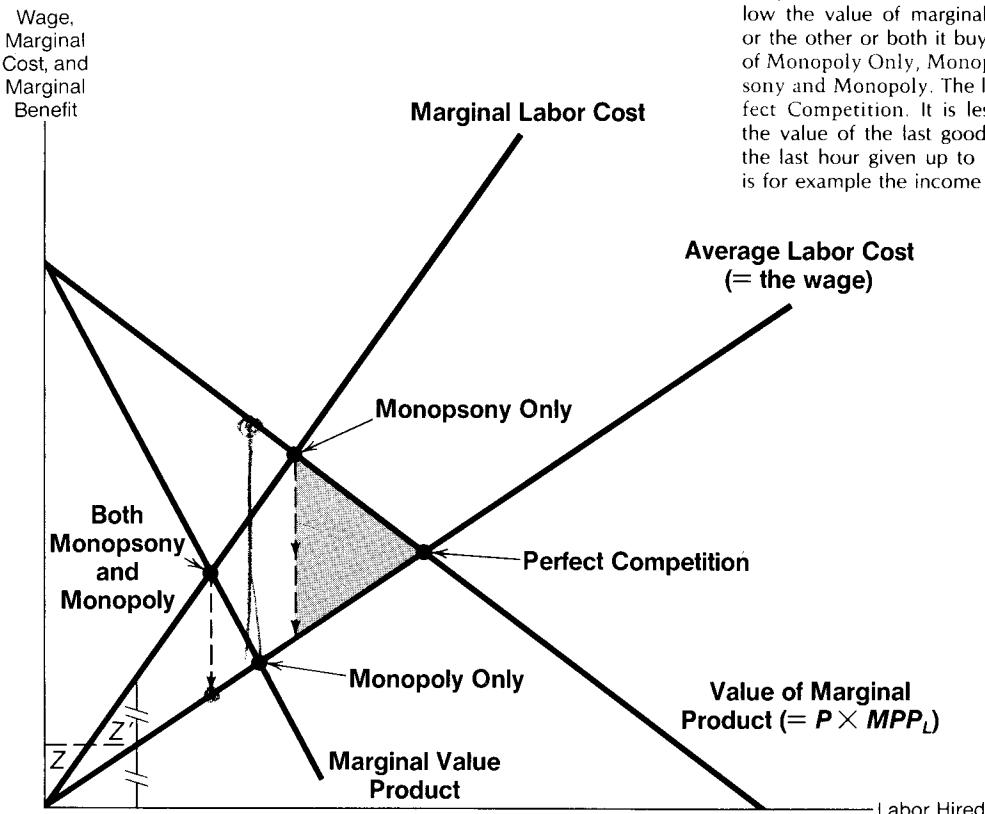


Figure 24.4
Monopoly or Monopsony Power Leads Away from the Competitive Point in the Labor Market

For a monopsonist in the labor market, the marginal cost of labor is above the average cost. For a monopolist in the product market, the marginal revenue product is below the value of marginal product. If the firm is one or the other or both it buys labor only out to the point of Monopoly Only, Monopsony Only, or Both Monopsony and Monopoly. The labor hired is less than at Perfect Competition. It is less than desirable in view of the value of the last good produced and the value of the last hour given up to produce it. The shaded area is for example the income lost from Monopsony Only.

it. When the elasticity of demand for the product is very high, approaching perfect competition in the product market, $1/E_Q^D$ is very low and the left-hand side reduces to $P \times MPP_L$. When the elasticity of supply for labor is very high, approaching perfect competition in the factor market, the right-hand side reduces to w . In other words, under *perfect* competition the value of the marginal product of labor is set *equal* to the wage. Under monopoly or monopsony it is not.

Applications of Monopsony

The most prominent labor monopsony in the American economy is major league sports, which has long been given special exemption from the laws against monopoly. The owners of the teams complain loudly when from time to time a new league springs up to share the monopsony and monopoly profits of the old. In 1973, for example, a new hockey league was competing for players against the

old National Hockey League, resulting in a sharp rise in salaries for players. Sportswriters, swallowing (as they invariably do) the arguments of the owners of the few franchises in the old league, complained that the fans, not the owners, would pay for fancy salaries for Bobby Hull and other stars.

Q: Who pays the higher salaries resulting from more competition in the player market?

A: The new league breaks the monopsony power of the old owners, forcing them to treat the wage as their marginal cost of labor. The player market moves from Both Monopsony and Monopoly to at worst Monopoly Only and at best Perfect Competition. More players will be hired by the leagues, new and old. More hockey will be played. The owners will no longer be able to exploit so well their monopoly in the market for professional

hockey games. On this account, prices at the gate will go down, not up. The profits from a franchise will go down, transferring money from owners to players. Any monopoly profits of the franchise remain to be divided up between the owner and the players. The fans are unaffected. They pay a high price at the gate if supply is restricted, as it is *without* the new league. But they do not care whether the high price goes to the owner or the player. It is not Bobby Hull's salary that makes the price high but whatever monopoly in the sale of hockey survives the coming of new franchises.

Using the algebra of monopsony one can measure the monopsony power of owners of sports franchises. Before baseball players were liberated they were paid only 20% of their marginal revenue product (and major stars were even more underpaid).³

T or F: The elasticity of supply that an owner faced was apparently only 0.25.

A: The equilibrium condition for the owner is to set his hiring such that the marginal revenue product equals the wage times the value of $1 + 1/E_L^S$. If the marginal revenue product is taken to be 1.0, then the wage is 20% of it, or 0.20. Solve the equation

$$1 = 0.20 \left(1 + \frac{1}{E_L^S}\right)$$

for E_L^S , the elasticity of supply of labor. The solution is

$$E_L^S = \frac{1}{1 - 0.20} - 1 = 0.25$$

Therefore, true.

Discriminatory Monopsony Monopsony in a labor market, like monopoly in a product market, can be discriminatory, even perfectly so. If it is, the social loss from the monopsony is lessened:

Q: Whaling was the chief occupation of New Bedford, Massachusetts, in the early nineteenth century. Much whaling labor was highly skilled and the labor all lived in New Bedford. Suppose that the market for whaling labor was competitive initially.

1. *True or false:* If the owners of whaling ships in New Bedford band together to set up a central hiring hall, the employment of whaling labor will fall and the income of the owners will rise.

2. *True or false:* If the owners now charge to each laborer a personal fee for entrance to the hiring hall,

their income will rise still further and employment will rise back to the competitive level.

A: 1. If the owners band together they will face the entire supply curve of labor and will therefore cut back their hiring to the point where the value of marginal product equals the marginal cost of labor (see Figure 24.5). Employment will indeed fall and the income of owners rise. Therefore, true.

2. If perfectly adjusted, the personal fee will extract *all* the benefit each person earns from exchange. No

³ Gerald W. Scully, "Pay and Performance in Major League Baseball," *American Economic Review* 64 (December 1974): 915–930.

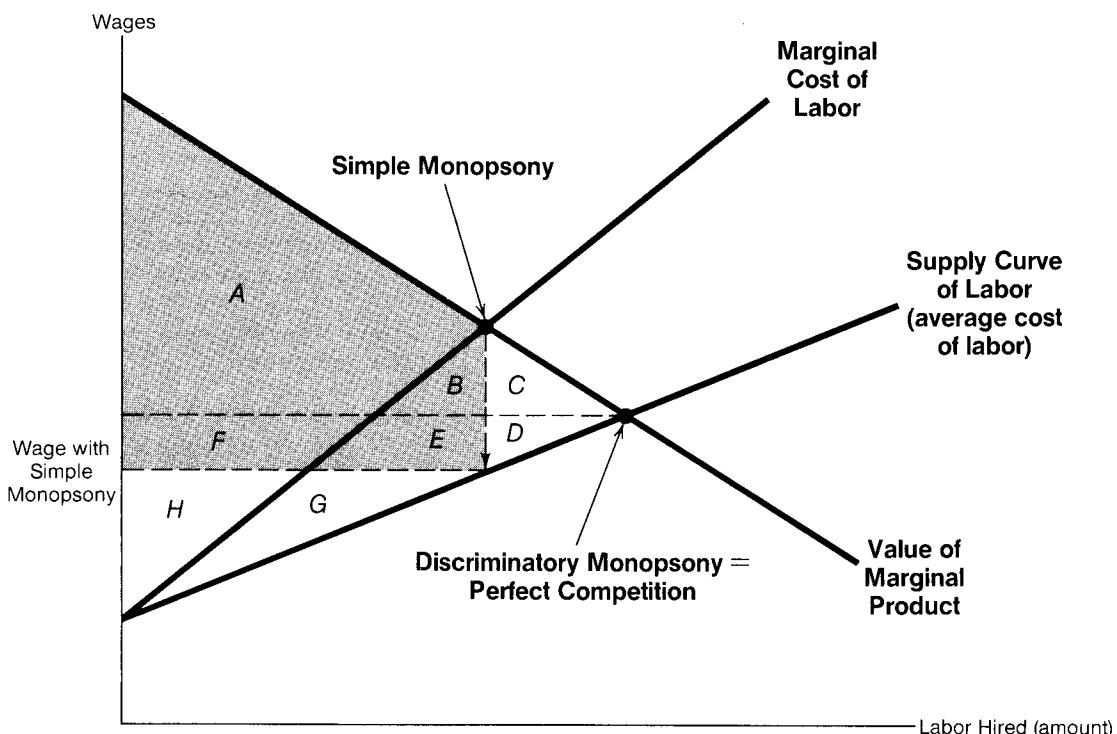
longer does the hiring of another worker increase the wage that must be paid to previous workers. Workers earn no economic rent. Their income does fall still farther and their employment rises. Employment rises until marginal benefit (the value of marginal product) equals marginal cost (which is now the wage necessary

to induce one more hour of labor to supply itself, that is, the wage along the Supply Curve, not the Marginal Cost of Labor). As usual in such cases, the equilibrium for perfectly discriminatory monopsony is the same as the equilibrium for perfect competition. Discriminatory monopoly can achieve efficiency.

Labor Monopoly Thus far all the monopoly power has rested in the firm that hires the workers. A baseball firm faces both ways, selling tickets to fans and buying services from players, and its power in either market affects its behavior in the market for players. The players themselves, however, can also exercise monopoly power, at least if they can form a union and strike, as they did in the summer of

Figure 24.5
Discriminatory Monopsony Can Raise Employment but Reduce the Welfare of Workers

A competitive industry buying labor earns the triangular area $A + B + C$. If it forms itself into a monopsony, it will cut back its hiring to the level of Simple Monopsony, earning the larger shaded area. If it is able to discriminate perfectly, paying each person or hour exactly what is acceptable and nothing more, then clearly it will earn the entire triangular area between the demand curve and the supply curve (namely, all the lettered areas). Employment will increase.



1981. The economics of monopoly applies in a straightforward way to unions and other restrictions on the supply of labor.

A monopsonist such as the organization of baseball owners facing a monopolist such as the players' union is engaging in pure bargaining, about which little can be said. A strike or lockout is a failure to reach any bargain to exchange labor for money. The bargain struck will presumably make both sides better off than no bargain. Aside from such obvious points of terminology, the situation resists analysis. It is the oligopoly problem revisited, the problem of predicting behavior when two intelligent people bargain without competitive constraints. Only one-sided monopoly or monopsony is easy to analyze.

Labor monopoly has two peculiarities. First, like governmental monopoly such as the postal service in the 1980s and unlike enterprise monopoly such as the international steel rail cartel in the 1890s, labor monopoly is legal. Second, and more important for present purposes, it is a monopoly of a good used to make other goods. That is, labor has a derived demand. The power of a monopoly can be summarized in its elasticity of demand: The lower the elasticity, the more power has the monopolist to raise its price. The power of a labor monopoly to raise the wage, therefore, depends on the elasticity of the labor's derived demand.

Applications of Monopoly in Factor Markets: The Octopus Effect

A monopolist tries to make as much money as possible. Cling to this truth in answering the following question.

Q: Suppose that the Teamsters hold a monopoly on truck-driving labor and suppose that they exercise the monopoly on behalf of their members, fixed in number. Consider a competitive trucking company, with downward-sloping demand for labor.

1. Where will the Teamsters set the wage to the company? What are the areas of profit to the Teamsters and nonlabor income to the company? By comparison with the wage and employment in the absence of the Teamsters, what is the social loss from the labor monopoly?

2. If the Teamsters owned the company, what wage would they set (now to themselves)? If they paid off the former owners of the nonlabor inputs, what would the Teamsters earn in additional profit?

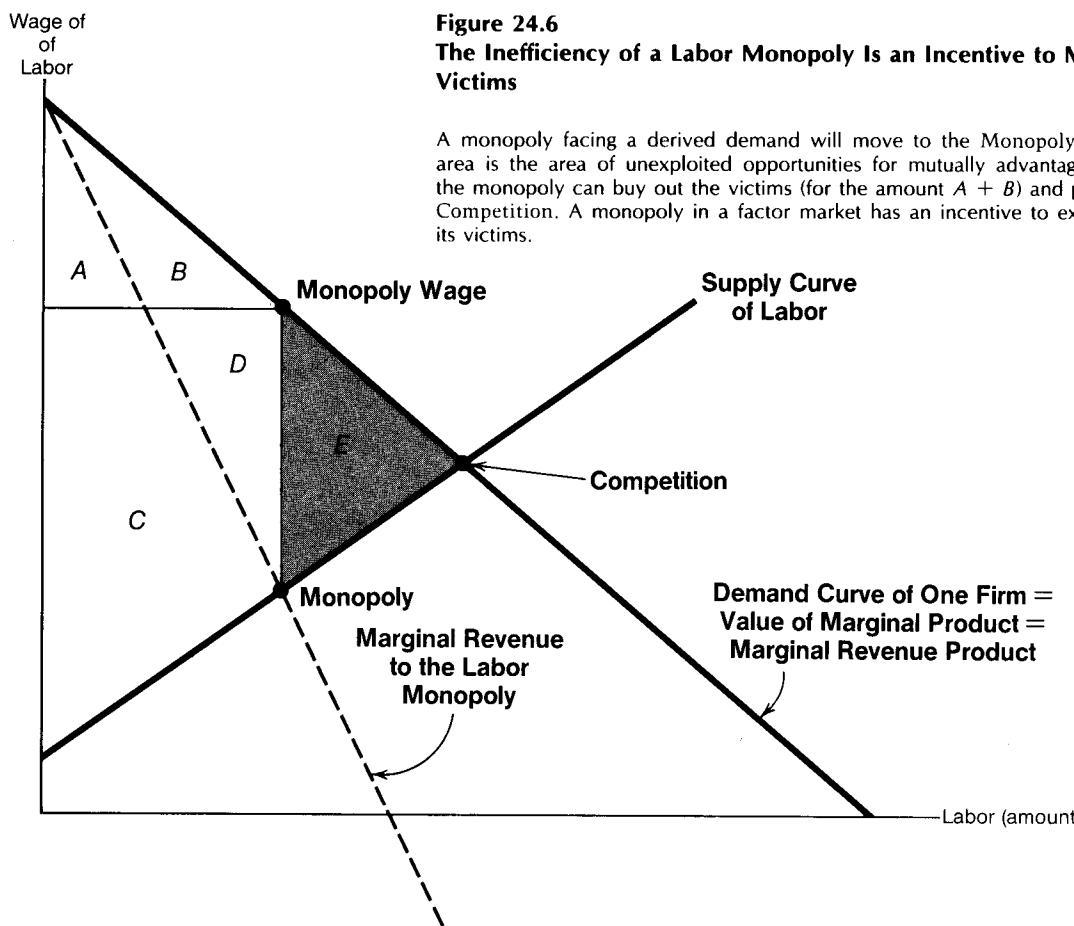
3. Is there what might be called an *octopus effect*, that is, a tendency for a monopoly in a factor market to extend its tentacles forward into the product market?

A: 1. The diagram appears in Figure 24.6. The dashed line is the curve marginal to the company's demand-for-labor curve (it is emphatically *not* the marginal revenue product, which is in this competitive case identical to the value of marginal product). On behalf of their members the Teamster's charge the high Monopoly

Wage, the wage that maximizes labor income. The excess of labor income over its value in alternative uses (that is, over the value given by the Supply Curve) is $C + D$, the conventional area of monopoly profit. The company earns in nonlabor income what remains from labor income, namely, the triangle $A + B$. The social loss from the monopoly is clearly the shaded area E .

2. The Teamsters need pay only $A + B$ to buy out the company. Once they own it they will set employment at the point Competition, getting for their members the additional area of profit E . The merger of Teamsters and the company eliminates the social loss.

3. There is an octopus effect. If the labor monopoly merges with its victims, it becomes both the supplier and the demander. As a seller alone its interest was to maximize a trapezoidal area of monopoly profit, $C + D$. As both a seller and a buyer, its interest now is to maximize the entire triangular area of consumer's and producer's surplus, $A + B + C + D + E$. The point Competition, of course, is where such a maximization takes place, although the income of labor is higher than it would be in ordinary competition. Labor unions—or any monopoly of an input, such as banks monopolizing inputs of funds—have an incentive to buy out the firms they sell to.



The octopus effect is a thinly disguised restatement of the proposition that monopoly is inefficient, or is inefficient unless the monopoly is perfectly discriminatory. One can think of the labor monopoly as "discriminating against itself." Under monopoly there exist opportunities for mutually advantageous exchange between buyer and monopolist. The merging monopoly takes advantage of these. Everyone is made at least no worse off: The customers and the former owners of the firm are no worse off and labor is better off. More monopoly can be good for you, at any rate if you ignore the additional resources expended to pursue or protect the richer monopoly.

The octopus effect depends on some degree of "substitutability in production." If the industry demands labor in fixed proportion to output, and has no leeway for substituting machines or material for people, then the labor monopoly can take its reward in the labor market or the output market indifferently. The monopoly will wish to send out a tentacle only if the exercise of monopoly in the labor market will cause an inefficiently small amount of labor to be hired there, a small amount because other factors have substituted for labor.

Summary The failure of property rights is one source of misallocation in the factor market, as in any market. The exploitation of bigness is another. A big firm buying labor is a monopsony, a big firm selling it a monopoly. In the usual way of market power, either will lead to fewer exchanges of labor than is optimal.

A special case is the reduced exchange and higher wage of a labor monopoly. The labor monopoly will have more success in raising the wage the lower the elasticity of derived demand for labor. A monopoly of labor, or of any factor, is favored by stickiness in the world and by cloots of habit. In the long run, when all things are adjustable, the elasticities are higher. The octopus effect is a use of these ideas: If the elasticity of substitution is anything but zero, the inefficiency from monopoly tempts the monopolist to take over his or her victims' affairs.

APPENDIX TO SECTION 24.2 Proof of the Equation of Imperfect Competition in Factor Markets

The profit to be maximized, π , is revenue, PQ , minus cost, taken to be labor cost alone, wL . In the case of a large firm, the price, P , is a function of how much Q is produced, Q is a function of L hired, and w is a function of L hired. The expression to be maximized is therefore

$$\text{Profit, } \pi = P(Q)[Q(L)] - w(L)L$$

Taking the derivative with respect to L (since L is what is to be chosen) gives

$$\frac{d\pi}{dL} = P\left(\frac{dQ}{dL}\right) + Q\left(\frac{dP}{dQ}\right)\left(\frac{dQ}{dL}\right) - w - L\left(\frac{dw}{dL}\right) = 0$$

The expression can be rewritten in elasticity form. Factor out from the first two terms dQ/dL (the marginal physical product of labor, styled MPP_L) and notice that $Q(dP/dQ)$ lacks only division by P to be the inverse of the elasticity of demand. Likewise, multiplying and dividing the last two terms by w leads to a second term that is the inverse of the elasticity of supply of labor. The result is, defining all elasticities as positive numbers,

$$P\left(1 - \frac{1}{\epsilon_Q^D}\right)MPP_L = w\left(1 - \frac{1}{\epsilon_L^S}\right)$$

which is the expression in the text.

EXERCISES FOR SECTION 24.2

1. The Northfield Woolen Mill (of Section 17.2) was a monopsony in buying labor in Northfield, Vermont, but perfectly competitive in selling wool blankets on the American market. What was its equilibrium condition for hiring labor? If the elasticity of supply of labor it faced was 1.0, what percentage of the full value of the marginal product would the wage be? What would happen to these if a new highway eliminated its monopsony power?

2. Suppose that the Northfield Woolen Mill after becoming competitive in its labor market became part of American Wool Blanket Monopoly, Inc. Suppose that the monopoly faced

an elasticity of demand for blankets of 2. What now is the equilibrium condition and the ratio of the wage paid to the value of marginal product?

3. Suppose that the road is closed, and Northfield Woolen (a subsidiary of the same monopoly) reacquires its local monopsony power. What is the ratio of wage to the full value of the marginal product?

PROBLEMS FOR SECTION 24.2

1. What are all possible areas of profit in Figures 24.5 and 24.6?
2. League rules restrict the ability of professional athletes in baseball, basketball, football, and hockey to move to the team that will pay them the most. The owners defend the restrictions as necessary for equal competition among teams, equal competition being necessary to make the games close and worth paying to see. Without the restrictions, the owners say, the teams with the most money would buy up all the best players. Suppose that the New York Knicks already have five superstars (Sam Williamson, Gary Fethke, Tom Pogue, Bill Albrecht, and Sam Wu) and that the Chicago Bulls have none. The Knicks customarily beat the Bulls 130 to 80. Suppose that the acquisition of Albrecht by the Bulls would make the customary score 110 to 95, with occasional very close games. Who would pay more for Albrecht, the Bulls or the Knicks? Is the owners' argument sound?
3. OPEC is a monopolist of oil. Suppose that there are no economies of scale at any stage of producing or using oil.
 - a. If oil were used only for gasoline and were made into gasoline by a process using fixed proportions of other factors of production in refining, would OPEC care whether or not it integrated backward into owning all the world's oil refiners?
 - b. If the same facts held as in (a) but gasoline were not being made with fixed proportions, would it care?
 - c. If oil were used also for petrochemicals (other than gasoline), would it care?
4. Suppose that Tay Bridge Company has a monopoly on truck crossings from Perthshire to Fife; it has no costs (except for a negligible insurance payment, so invulnerable to disaster is it). Will it raise the monopoly profits of the bridge owners if the company integrates forward into trucking, that is, acquires control of the (competitive) trucking industry that uses the bridge? (*Hint:* Consider the demand for bridge passages derived from the demand and cost for truckloads crossing the Tay. Show the truth of your answer for the special case of a straight-line, but downward-sloping, demand curve for trucking services, a flat marginal cost of trucking, and as mentioned zero costs. Why is the result different from the case of the Teamsters in the text?)

True or False

5. The point Monopsony Only must entail a larger employment of labor than the point Monopoly Only.
6. The marginal revenue product is marginal to the value of marginal product as marginal revenue is to the demand curve.
7. The imposition of an effective minimum wage on a monopsonist will sometimes reduce the number of workers employed by a monopsonist.
8. Since the steel industry faces an inelastic supply of iron ore, Bethlehem Steel is a monopsonist.

- **9.** A rise in demand for cotton would trap cotton producers in a cost cage as the price of labor in cotton rose.
- **10.** An iron ore monopoly merged with a steel monopoly will now sell ore to itself at a competitive price.
- **11.** The victims have as much incentive to buy up the monopoly as the monopoly has to buy up the victims.

The Supply of Labor

25.1 The Leading Idea in Labor Supply: Compensating Differentials

What to Read For

What is the principle of compensating differentials? Do safety regulations make workers better off? How are compensating differentials determined in the market? If most people dislike jobs as pipefitters relative to jobs as professors, which profession will earn a compensating differential?

Equal Pay for Equal Unpleasantness

Assumes equilibrium tends to become true

The labor and other factors demanded according to marginal productivity must also be supplied. They are supplied to any one job according to the *principle of compensating differentials*. Among the most useful of economic doctrines, the principle says simply that a person's pay in a job must equal her pay in another. More exactly, the *whole* pay, in happiness as well as money, of the *marginal* worker, just indifferent between the two jobs, must be equal in the two. If money pay is not equal, the worker must be compensated by other differentials between the jobs—differentials in working conditions, say, or in the reliability of employment.

The principle is merely a condition of equilibrium: If it is violated the workers will move around until it is reestablished. A professor of economics, for example, earns less than she could in business. If the difference is \$15,000 a year she would have to "earn" \$15,000 of differential pleasure as a professor, perhaps the pleasure of being largely her own boss or of associating with other people devoted to the life of the mind. If this were not so, she would change jobs.

The applications of the argument are limitless.

T or F: The introduction of workmen's compensation, under which employers pay all the costs of injuries on the job, raised wages of college professors and file clerks relative to those of coal miners and trapeze artists.

A: Before the law the risky occupations earned a premium to compensate for the risk, a premium that vanished with the coming of workmen's compensation. That is, true.

T or F: Legislation on safety in coal mines does not affect the welfare of miners.

A: The conclusion seems inescapable if workers are free to move into and out of coal mining. High wages compensated for risky mines. That is, true.

The natural tendency will be to resist such reasoning and to point out that in the bad old days not only were mines risky but wages were low. The point is irrelevant. To be sure, people were poor long ago. Their poverty is their low income in total, money income plus the money value of working conditions. What matters for the question is the *composition* of the total. Legislation on safety makes the miners take their income, whether low or high, in safety rather than in money. Any other result would cause them to exit or enter mining. One can escape the conclusion only by arguing (as one might) that miners do not see their own best interest in safe mines.

Another surprising application of the idea of compensating differentials is the following.

T or F: A legal minimum wage may not cause unemployment (excess supply of labor) if working conditions can adjust (by increasing the intensity of an hour of work, say, or by reducing the employer's expenditure on safety devices on the job).

A: Again, entry and exit fix the whole pay. A 16-year-old working at McDonald's can get \$2 an hour (below

the minimum) and a leisurely pace or \$4 an hour (above the minimum) and a hectic pace. The hectic pace may well choke off what would otherwise be an excess supply of labor at the high wage. The worker may be indifferent between the two deals, as might the employer, leaving the society unharmed by the minimum wage, or harmed only a little. That is, quite possibly true.

The Whole Pay Is Determined by Competition, Not by Its Composition

Bargaining over wages, then, cannot affect the welfare of workers unless the bargaining changes the amount of competition the workers face. The whole pay of miners will rise only if coal miners can extract a promise from owners to hire only at union wages and if the miners can restrict entry to the union. The whole pay of workers at McDonald's is not affected by air conditioning of the kitchens. The whole pay of college professors is not affected by a bargain to raise their money wages faster than supply and demand would warrant. Working conditions will adjust.

"Working conditions" include more than legal privileges. For example, theft from the company is part of the whole pay. The company will have to pay more in money wages if it does not allow a professor to use college stationery to write her mother, the waitress to eat the last piece of pie, or the longshoreman to steal television sets off the docks. The company can pay in money or it can pay in stationery, pie, and television sets.

T or F: The company has no incentive to prevent stealing if the value of the goods stolen is as high to the worker as to the company.

A: The money wage plus the value to the worker of the goods stolen is the whole wage, given by competition from the outside. Take the case of television sets stolen from the docks. The company will not care about the stealing if the longshoreman can sell the stolen televi-

sion for the same price as the company would get, because he will then accept wages lower by exactly the value of the television set. The company does not care whether it pays \$100 worth of wages or \$100 worth of television sets. On the other hand it will care a lot if the longshoreman only gets \$50 for the hot television set from his fence. The longshoreman will accept only \$50 less wages, even though the company values the stolen television at \$100. That is, true.

Ships and Chips The history of the English Navy yards provides a spectacular case of the principle of compensating differentials in action:

Q: In the seventeenth and eighteenth centuries in the royal shipyards, where fighting ships were built, workers claimed the right to remove pieces of waste wood (called "chips") and sell them as lumber. Guards, floggings, and hangings failed to stop the practice, which ended only in the early nineteenth century, when mechanization took away the opportunities to create "waste" wood in cutting. Before mechanization only one sixth of the wood coming into the yard for ships went out as ships. The rest went out on the backs of workers, as chips.

1. *True or false:* The workers were made better off by being able to take and sell the chips (the chips sold for less than their cost to the Navy).

2. *True or false:* The coming of mechanization was therefore ruinous to the workers, another case of capitalism robbing the workers of their dignity and self-determination.

3. *True or false:* Both the workers and the Navy would have been better off if workers had been paid exclusively in money and had been prevented from taking chips.

4. *True or false:* Nevertheless, the Navy could sustain no such mutually beneficial agreement to pay only in money if bad cutting (which produced chips on purpose) was difficult to detect.

5. *True or false:* If all scrap lumber could be burned by the Navy, no problem of chips would exist.

6. *True or false:* The official burners were bribable; the problem would return.

A: 1. No, the workers were not made better off by being able to take the chips. Competition fixed the *whole* pay of the workers. Taking the pay in chips meant

The principle of compensating differentials is a sharp tool. An extreme application of it is to the question of whether urbanization and industrialization made people better off.

T or F: If it is literally true that all differentials in pay are compensated by differences in the value of working conditions, then, despite the observed rise in money income during the Industrial Revolution, no increases in happiness came about from the movement of people from farms to factories.

merely that the money wage the Navy would need to pay was lower.

2. Therefore, no, the workers were not ruined by the coming of mechanization. Each worker quite properly viewed his own chips as a benefit to him, just as you would view part of your pay as a benefit. As people do, the workers did not understand that eliminating chips would raise their money wages; or at any rate the historian of the workers does not.

3. Yes, both the workers and the Navy would have been better off with money pay alone. The chips sold for less than they were worth to the Navy as ships. The wood was badly used, leaving a social gain from using it better to be shared out between the workers and the Navy. The Navy could get cheaper ships yet still pay the workers more in money alone than they earned in money and chips combined.

4. True, such a mutually beneficial agreement would be subject to free riding, because each worker has an incentive to cheat and to take chips. The Navy in fact offered the workers an agreement, which they accepted. The agreement broke down when workers took not only the higher pay but also the chips.

5. Yes, if all chips were burned as soon as they were made, there would be no incentive to create them artificially for purposes of stealing and selling them. The utterly spiteful measure of burning up what is a valuable resource (the chips) would prevent another waste (creating the chips to order).

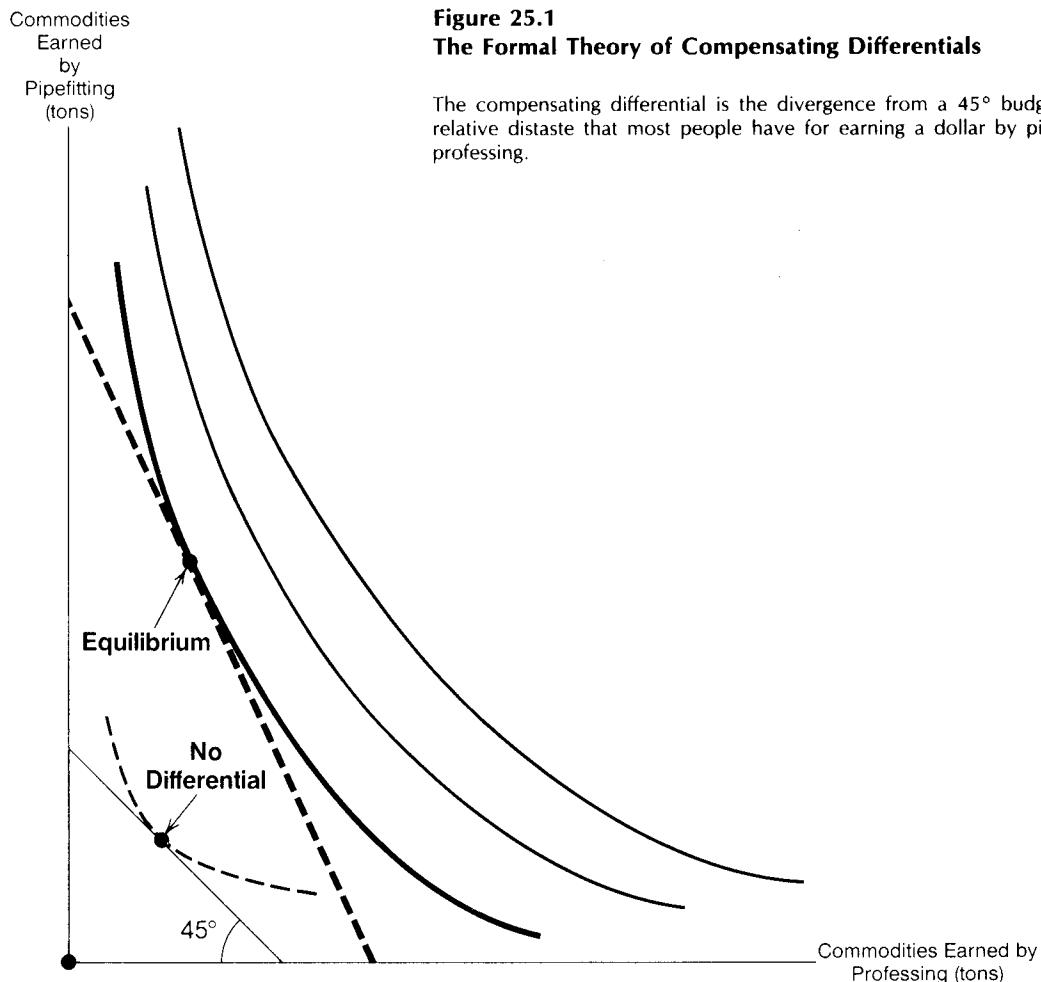
6. Yes, the official burners could be bribed to give up their chips. In fact the inspectors hired to prevent the theft of the chips in the first place were bribed to overlook the theft.

A: True. The higher wages in the factory were not clear gain if they were merely compensation for bad working conditions. We wish to measure people's happiness, not their holdings of greenish portraits of George Washington (or King George). Conventional national income, which rose when people moved to factories, measures paper not happiness.

The Indifference Curves for Equalizing Differentials

The theory of equalizing differentials can be put formally as a choice between two commodities, such as income earned as a professor and income earned as a pipefitter. The analogy with a person consuming is exact. Instead of allocating money income between hamburgers and housing, the person allocates hours of time between pipefitting and professing. Just as each dollar buys a certain number of commodities, each hour earns a certain number of dollars to buy commodities, determining the budget line between the two fruits of the hours (see Figure 25.1).

People have feelings about how they would like to earn their daily bread, represented here by the solid indifference curves. The dashed budget line gives the terms of exchange between the two ways of using hours to earn. Just as money income is in consumer theory the scarcity behind the budget line, here the amount of hours is the scarcity. As drawn, professors earn less than pipefitters.



T or F: That Mendels, Loschky, and Komlos do not usually do both pipefitting and professing implies that their indifference curves either intersect the axes or are straight lines.

A: They will do both if the indifference curves have the usual, convex shape and do not intersect the axes, be-

cause the equilibrium point will be interior, as in the diagram. Therefore, since we do not in fact observe people doing all sorts of jobs, indifference curves must have the shape that yields corner solutions. In other words, although people do not usually specialize in consuming food or housing, they do specialize in employment. Therefore, true.

The diagram, therefore, best represents the whole market, not one person. You can read it as determining relative wages. The dashed line would be the relative wage if professors and pipefitters were demanded by the economy in fixed amounts given by the point Equilibrium. Since most people are portrayed here as disliking pipefitting relative to professing, the pipefitters earn a compensating differential; that is, they earn more per hour.

Summary

For the marginal worker just indifferent between two jobs, the whole pay must be the same in both jobs. If working conditions are not the same, as they are not in coal mining and store clerking, the wage adjusts; if the wages are not the same, the working conditions adjust. This is the principle of compensating differentials. Entry and exit fix the whole pay, leaving the composition of the whole to vary. Therefore, safe miners will be miners with lower money pay; McDonald's workers with high pay will be workers with hot kitchens; dockers who steal television sets will be dockers with lower money pay, or with unemployment if they insist on high money pay as well.

The theory can be put in the same way as the theory of consumption, showing how tastes for one as against another way of earning commodities determine the equilibrium differential between pay in different occupations. For a single person, the diagram shows how the given differential leads her to choose one occupation or the other. For a whole economy, the diagram shows how a given composition of demand leads to the differential pay.

EXERCISE FOR SECTION 25.1

1. Nowadays any professor of economics can earn \$10,000 more working as an economist for a private company.
 - a. Who is the "marginal" economist that is choosing between professing and businessing?
 - b. How much to the penny does she "earn" from the differential advantages (net of disadvantages) of professing?
 - c. Professor X "earns" in differential advantages the following, with money values: ability to set own schedule (\$8000); loose relation to boss (\$5000); contact with genial, intelligent, eager undergraduates (\$5000); contact with obnoxious, stupid, sullen undergraduates (minus \$4000); feeling of being to one side of the affairs of the world (minus \$3000); contact with new ideas (plus \$5000). What is her total differential? What happens if it is over \$10,000?
 - d. Professor Y once felt exactly the same way as Professor X. As he gets older, though, his valuations of the various items change in this way:

Own schedule	Up \$2000
Loose relation to boss	Up \$2000
Contact with genial undergrads	Down \$1000

Contact with obnoxious undergrads	Down \$4000
Feeling of being out of it	Down \$4000
Joy of new ideas	Down \$2000

What does Professor Y do?

- e. If large numbers of Professor Ys acquire a strong dislike for obnoxious students, what happens to the total supply of professors? What then happens to the unfavorable monetary differential?

PROBLEMS FOR SECTION 25.1

1. Liability for accidents in a plant is placed on the employer. *True or false:* The employer will spend no more in preventing accidents in his plant than he would if the liability (that is, the cost of hospital care, etc.) were placed on the employee.

- 2. Federal judges have more power and prestige than do most practicing lawyers. The salary of federal judges, however, is well below what a good practicing attorney can make. *True or false:* If the offer of judgeships is random with respect to ability, this implies that federal judges on average will be worse lawyers than practicing attorneys.

True or False

3. That a college professor earns less than NASCAR racers (whose average speed through life is 200 mph) is evidence of the perverted values of Americans.

4. You would expect bartenders to be worse off in countries where it is not customary to leave a tip on the counter.

- 5. The wage of garbage collectors is lower than that of engineers because their work is dirty and disgusting.

- 6. The choice among occupations is not affected by a proportional tax on money income.

7. An unemployment compensation law (in which all unemployed are eligible for benefits) would reduce the wage rate in depression-prone industries relative to that in safe industries.

8. A worker in a remote district who has to buy all his food at high prices in the company store is worse off than one in a big city who can spend his wages wherever he wants.

25.2 The Choice Between Work and Leisure

What to Read For

What is the economist's definition of leisure? Should its value be included in the national income? What is the analogy between the theory of consumption and the theory of the supply of labor? How can supply curves slope backward? Does an income subsidy reduce work? How? What is a negative income tax?

The Shadow Wage of Leisure

In short, like many other theories in economics, the theory of labor supply is a theory of choice under scarcity. The scarcity is hours. The scarcity of hours requires not only a choice between two occupations but a choice between being occupied and being unoccupied. Being unoccupied is called conventionally *leisure*, but it is best to think of it as all uses of time other than working for wages. The "leisure" can be sitting in the sun or sleeping late. But it can also be raising four children or writing the great American novel, more demanding

than most paid jobs. Leisure in the technical sense is not sloth but merely time not paid a wage.

The point is that the wage is the opportunity cost of whatever is produced at home—whether a happy family or merely a suntan. One can speak of a *shadow wage* as being the dollar value of what is produced at home, in which case a person's participation in the labor force depends on whether or not his shadow wage is higher than the available market wage.

T or F: As long as women are paid less than men, women will specialize more in housework.

A: You can assume that women are as skilled as or more skilled than men at housework (whether the skill is natural or acquired is not at issue). That is, their shadow

wages in housework are the same or higher. But the opportunity cost of women's housework is less than men's if women earn less in the market. If anyone does, it will be rational for a wife, not a husband, to stay home. Therefore, true.

The idea brightly illuminates the domestic scene.

T or F: A narrowing of pay differentials between men and women will lead to more equal sharing of housework between husbands and wives.

A: By running the argument just given in reverse, true.

It is customary to exclude the value of housework from the national income. The custom arises from the difficulty of measuring shadow wages (except for the housewife who by jumping into and out of the market labor force exhibits indifference between wage work and housework: For her the market wage is the shadow wage). For the same reason the national accounts ignore most other nonmarket uses of time, such as most home vegetable gardens, much do-it-yourself home repair, all driving time in a car, and all leisure in the narrow sense. Measured income rises when people buy a tomato instead of growing it, hire a plumber instead of doing the work, buy a cab fare instead of driving it, and sell time instead of enjoying it.

That shadow wages are hard to measure is not much of an objection to including all the income of the nation in a measure of national income. All things are valued at their marginal values. The superhousewife with a high shadow wage earns so to speak a consumer's surplus from consuming housemaking, but so does a pizza lover with a total valuation of pizza higher than its price. The exclusion of consumer's surplus from national income is not a flaw if you believe that national income pretends only to measure the nation's budget constraint, not its happiness. But it *is* a measure of happiness, and is notably incomplete.

The Theory of the Supply of Labor

The single supplier of labor faces a budget constraint in allocating her hours to leisure or to wage work, a budget constraint whose slope is the real wage (see Figure 25.2). The endowment point is 24 hours, where the dashed lines begin. The units of the slope of one of the dashed lines are dollars of wages per hour divided by dollars of price per ton of the commodity. This reduces as it should to tons per hour. Since leisure and commodities are goods, not bads, the indifference curves have the usual shape. The hours or work supplied by the worker in the figure are those left over from the 24 hours in a day after the worker has settled on, say, an Equilibrium amount of leisure.

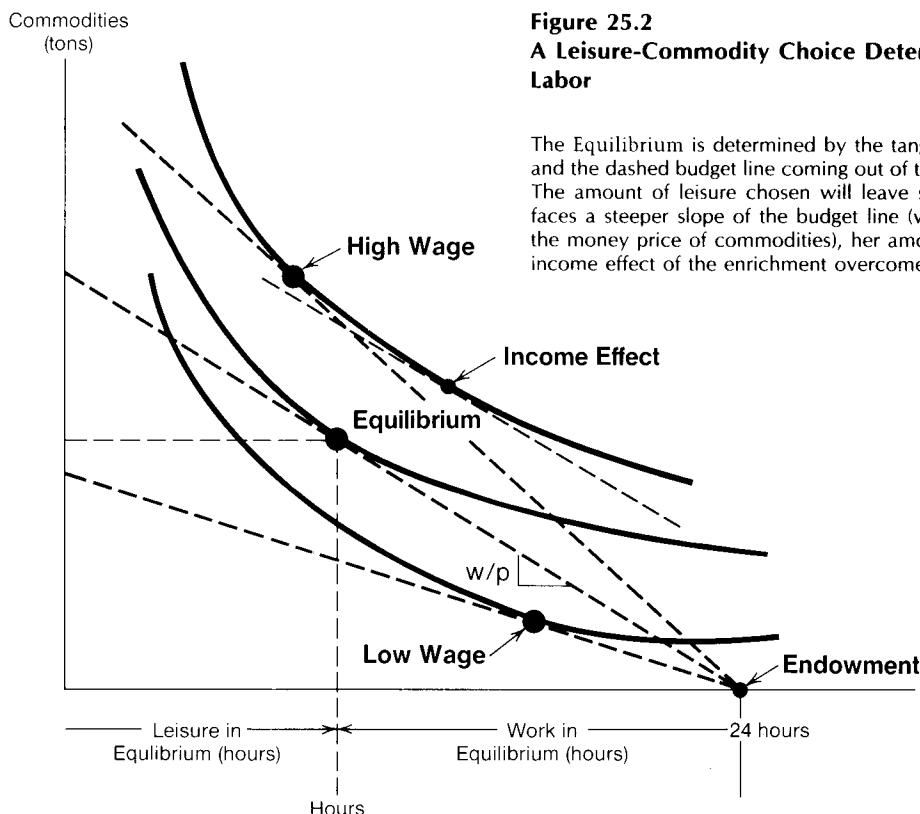


Figure 25.2
A Leisure-Commodity Choice Determines the Supply Curve of Labor

The Equilibrium is determined by the tangency between an indifference curve and the dashed budget line coming out of the endowment of 24 hours of leisure. The amount of leisure chosen will leave some hours for work. As the person faces a steeper slope of the budget line (which is the money wage divided by the money price of commodities), her amount of work will increase, unless the income effect of the enrichment overcomes the substitution effect.

The apparatus is nothing but consumption theory with one commodity taken to be leisure. The astonishing usefulness of the diagram will illustrate once again the astonishing usefulness of consumption theory. The first use is to undergird the supply curve of labor. As the wage varies, the equilibrium varies, from Low Wage to Equilibrium, and then to High Wage. By the usual geometry, the rising hours of work could be plotted in a diagram below this one against the rising wage to give an upward-sloping supply curve of hours.

Backward-Bending Supply But consider the following:

T or F: Without violating any usual features of indifference curves, the point High Wage can be to the *right* of Equilibrium (implying *fewer* hours of work supplied at a higher wage).

A: Look at the point Income Effect. It represents what would have happened to the person's consumption of commodities and leisure had income gone up without a rise of wages. (Notice that the dashed wage line that gives the point Income Effect is parallel to that giving

Equilibrium.) It represents, therefore, the pure income effect, without any substitution effect. If leisure is a normal good (as of course it is natural to assume it to be), then the hours of leisure increase as income does, leaving less time for work. The pure substitution effect of the cheapening of commodities relative to leisure is the move from the point Income Effect to the point High Wage. You can see that the substitution effect depends on the curvature of the indifference curve—that is, on the substitutability of commodities for leis-

sure. If stereos and McDonald's hamburgers were poor substitutes in consumption for flute practice and home-made bread, then the final High Wage point could easily

A single supplier, then, might not exhibit a conventional supply curve of labor. The singer Perry Como earns so much in one performance that he gives fewer performances than he gave before his reputation was established. But the proposition is much less likely to apply to all singers as a group, for the following reason.

T or F: A rise in real wages will increase the percentage of people who do *some* market work.

A: Before the wage goes up, some of the people who do not work are at a corner (the rest are around the corner, but will reach it at a higher wage). That is, they earn no wages (or no wages in singing), spending all their time at leisure (or in nonsinging jobs). The

entail *less*, not more, labor supplied than the point Equilibrium. The corresponding supply curve would show a portion that bent *backward*.

A rise of wages has for them *no* income effect: They start by sliding up the same indifference curve as soon as they are induced to work at all. It is quite possible, and even likely, that the entry of new workers (or singers) will offset the growing reluctance of the old to give up more scarce leisure. At the least, the idea of a backward-bending supply curve looks much less attractive if one allows for new entrants to the labor force.

Consider the following, for instance.

Q: Suppose that Robert Ankli, a Scottish highlander, is initially self-sufficient, producing oats by sacrificing some of the one other commodity that interests him, leisure. Describe briefly his position, assuming that Ankli has an initial endowment of 24 hours of leisure and the usual shape of the production possibility curve. Suppose that he is now presented with the opportunity to become a hired worker, working for constant hourly wages on a neighboring farm. What is the minimum wage that will induce Ankli to abandon entirely self-sufficient farming? Suppose that the wage actually offered to him is exactly this minimum. Will his hours of leisure be larger than, smaller than, or equal to the hours under self-sufficiency?

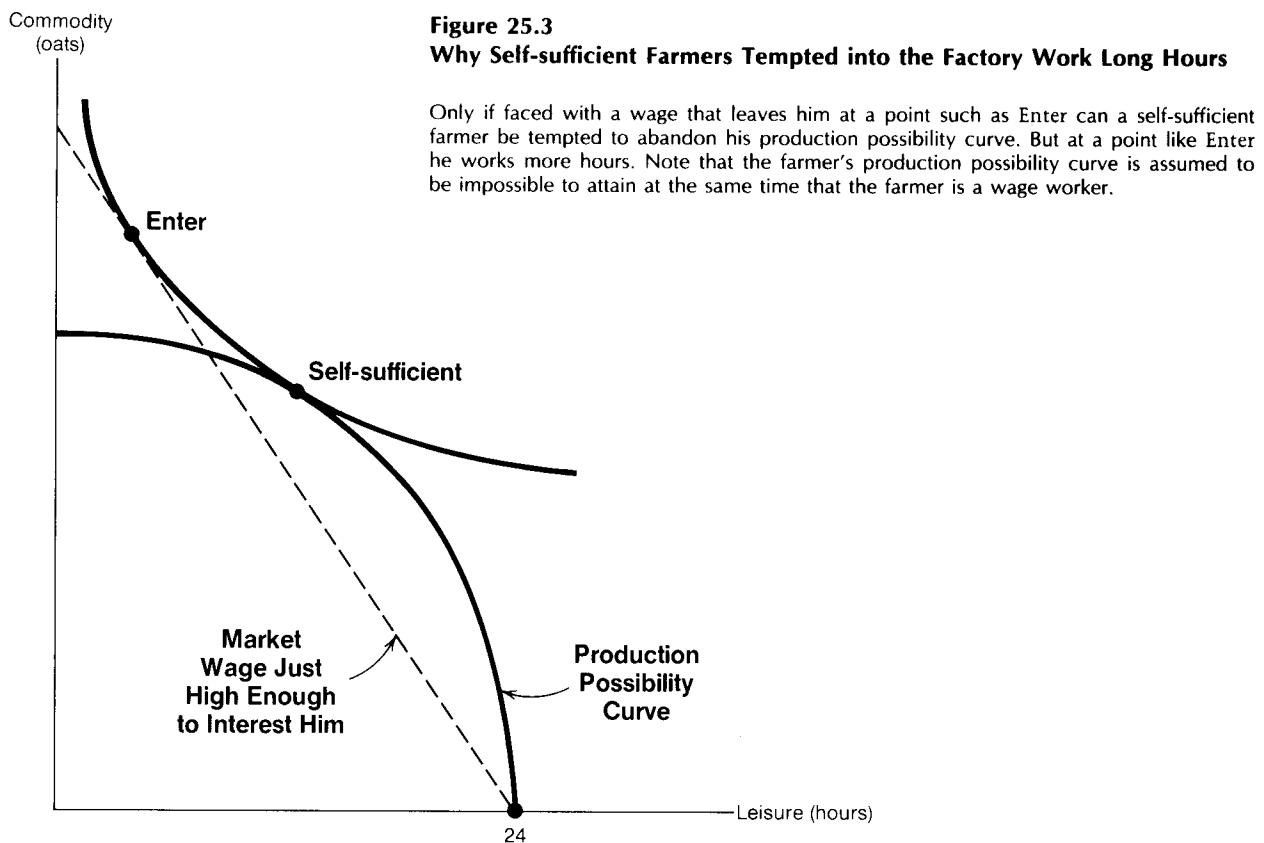
A: Ankli's initial position is Self-sufficient along his production possibility curve in Figure 25.3. The dashed line is the Market Wage Just High Enough to Interest Him. That is, it is the line just steep enough to touch the indifference curve through Self-sufficient. It touches it at a point such as Enter, with *less* leisure than at Self-sufficient. The tale is one of early industrialization: In the presence of diminishing returns to labor in subsistence agriculture, the first recruits to the factories chose to work long hours.

Applications of Income Effects: Inheritances and Subsidies

Quite naturally, a larger endowment of commodities will reduce the amount of work that a person supplies. If Jeffrey Williamson receives a large inheritance, his endowment will move up off the leisure axis in Figure 25.4 to the point Subsidy. The terms of the inheritance permit him to still earn a wage, but he earns it along the dashed line parallel to the original one. Since leisure is a normal good, he will always work less: Other things equal, people rich from *unearned* income work less.

The argument and the diagram apply also to the poor. The Subsidy point might then be a socially acceptable minimum of income. A straightforward subsidy in that amount will put some of the poor above the utility of the Subsidy point, moving to point Better at the expense of taxpayers. Such a poor person will work less (Better is to the right of Poor) but have more, because the poor person supplements the subsidy with earned income.

A clumsy mechanism for keeping the utility of the poor low but above the



minimum is to specify that anyone who accepts the Subsidy cannot earn any more. In early nineteenth-century England a subsidy without strings attached was called outdoor relief; with strings it was called indoor relief—that is, indoors in the so-called workhouse, which contrary to its name was a prison designed (by economists, regrettably) to prevent people from moving from the point Subsidy to the point Better by working for pay. The workhouse made the point Subsidy less pleasant by setting the inmates to useless but difficult tasks and by separating the sexes, just as modern versions of such subsidies make it less pleasant by giving it through surly clerks, long lines, and inquisitorial social workers.

The effect of the workhouse or its modern equivalent is perverse, inducing many people who were willing to work to choose to do nothing but accept the point Subsidy. The terms of the subsidy are that one cannot take it and *also* be in the labor market. The budget line is now the old dashed line augmented by the point Subsidy (recall the analysis of free education early in the book). Whenever Subsidy is on a higher indifference curve than some point on the old budget line the recipient abandons the work line entirely. Only the work lover (or commodity lover) among the poor stays on the job, because for that person Subsidy is on a lower indifference curve. The rest go on the dole full time.

Commodities

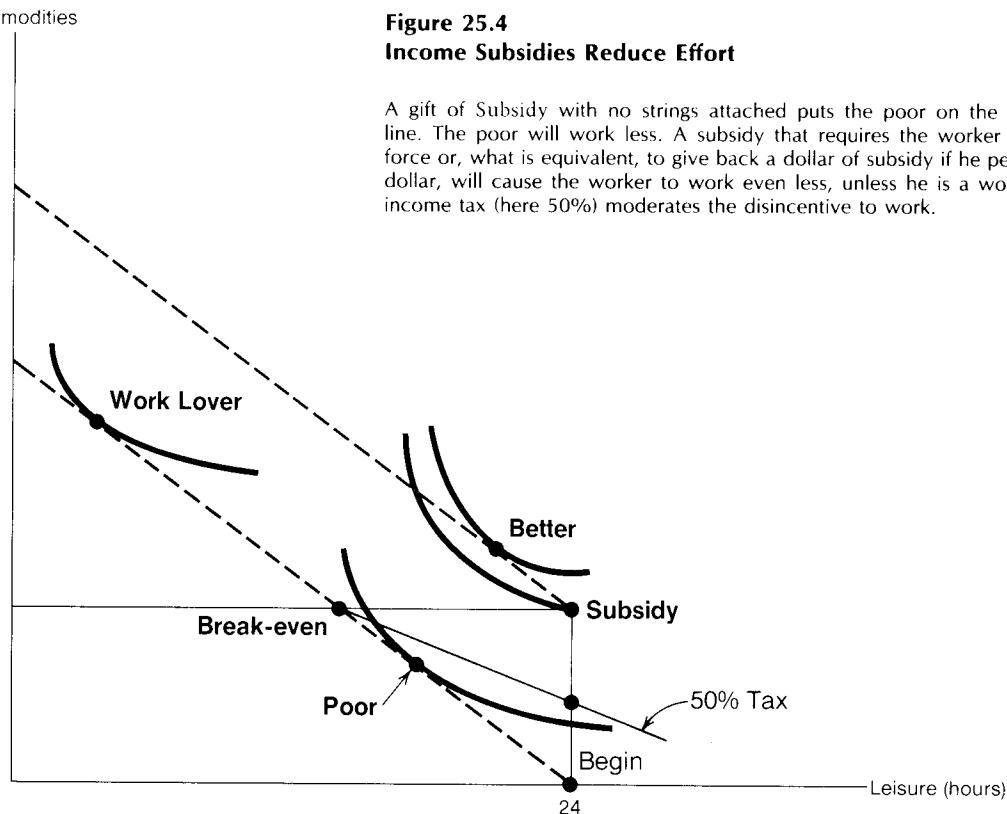


Figure 25.4
Income Subsidies Reduce Effort

A gift of Subsidy with no strings attached puts the poor on the higher dashed budget line. The poor will work less. A subsidy that requires the worker to abandon the work force or, what is equivalent, to give back a dollar of subsidy if he personally earns another dollar, will cause the worker to work even less, unless he is a work lover. The negative income tax (here 50%) moderates the disincentive to work.

T or F: A scheme in which earned incomes below the level of Subsidy were subsidized until they came up to Subsidy would have the same effects on the incentive to work as the scheme in which Subsidy is an all-or-nothing offer.

A: Subsidy would bring all incomes below Subsidy up to the level of the horizontal solid line through the

point Subsidy. The worker would be faced with virtually the same budget line as the all-or-nothing offer. Only Subsidy would be chosen by someone at Poor because no one with a positive marginal valuation of leisure would take up a position on the horizontal line. In short, true. Unfortunately, at present many welfare programs have this feature.

The Negative Income Tax

It is fruitful to look at the problem another way. If earned income is below Subsidy, for each dollar a poor person earns the government takes away a dollar of subsidy. The scheme amounts to a tax on income, a tax with 100% marginal rates. The government taxes effort heavily and therefore the poor workers supply little of it. A partial solution to the problem is to let the poor *keep* some of the extra money they earn, *reducing the tax rate*. This scheme, very popular among economists of all political stripes, is called a *negative income tax*. The diagram shows a tax rate of 50% on the poor, for example, the highest imposed on the earned incomes of the rich. As the tax gets lower, the line swings clockwise around the Break-Even point, giving less and less incentive to go on the dole, though also giving less and less of a dole. The desires of the society to encourage

work and to maintain incomes above some minimum are inconsistent: To some degree one must be sacrificed to achieve the other.

Summary The theory of labor supply is the theory of the consumption of leisure, where “leisure” is “any activity that does not earn a wage and takes time from activities that do.” If the shadow price of leisure is high relative to the market wage, labor is not supplied. Time is used at home.

All the usual features of consumption theory apply. For example, it is important to distinguish between the income effect and the substitution effect of a rise in wages. The income effect can make the corresponding supply curve of labor bend backward, perversely reducing the amount of labor supplied as the wage rises. It can, that is, but does not always do so: The entry of people into labor and out of leisure is a substitution effect alone, always working in the conventional direction. Like consumption theory in general, the theory of the consumption of leisure—the supply of labor—can be used to analyze changes in the budget line of the consumer imposed from outside. The chief case in point is a subsidy to the poor. If the subsidy is an income subsidy the diagram shows that it reduces the amount of labor supplied; if it is a wage subsidy it increases the amount supplied. In either case the theory casts a bright light on the perennial attempts to help the poor without altering their incentives to help themselves.

EXERCISES FOR SECTION 25.2

1. Suppose that the coordinates of the points in Figure 25.2 are

	Leisure Taken (hours)	Commodities Gotten (tons)
High Wage	15	4
Income Effect	18	3.4
Equilibrium	16	3
Low Wage	20	1

a. What are the hours of labor supplied at the three actual points (Income Effect is hypothetical)? What is the *hourly* wage at each? What, then, are three points on this person’s supply curve of labor? (Recall: A supply curve relates quantities supplied—measured on the horizontal axis—to the price per unit—measured on the vertical axis.)

b. What is, in terms of hours of labor supplied, the income effect of the rise in wages from those through Equilibrium to those through High Wage? How much is it in terms of commodities?

2. Suppose that a poor person is originally at Poor in Figure 25.4. She works eight hours a day and earns \$6000 per year. Measure “commodities” in dollar’s worth per year.

a. The government gives the poor person a \$10,000 subsidy, but will reduce the subsidy by a dollar *for each dollar she works*. If she nonetheless goes on working eight hours a day, what will her payment be from the government? How much money will she earn in total? How much leisure will she consume?

b. If she decided not to work at all, what would her payments, income, and leisure be? Would a rational person go on working?

c. The government, seeing its error, goes to the other extreme, removing entirely any reduction of subsidy if she works. Under this deal, when would she work at all? Would she ever work as much as she did without any subsidy?

d. The government, seeing its next error (and the large payments to the poor person), decides on a negative income tax, with the following schedule:

If the Poor Person's Income from Work Is:	The Government Gives the Poor Person:
0	\$5000
\$ 1,000	4500
2,000	4000
3,000	3500
4,000	3000
5,000	2500
6,000	2000
7,000	1500
8,000	1000
9,000	500
10,000	0

How much is the poor person “taxed” on each additional \$1000 she earns from work (that is, how much does the government cut her subsidy)? What is the rate of tax? Draw the budget line that the negative income tax gives to the poor person.

PROBLEMS FOR SECTION 25.2

- 1. In 1974 the average wage was \$760 per month in the United States and \$640 per month in Japan. Suppose (for simplicity) that women could earn this wage. Suppose further that in Japan most married women choose to stay at home and that in the United States most work in the market. *True or false:* The difference between the shadow wage at home in the two countries is less than the difference between the market wage; indeed, it may be that a housewife (whether or not that job is chosen) is more valuable in Japan than in the United States.
 - 2. Examine the effect on the labor force participation of women of
 - a. Increased subsidies for the construction of day-care centers.
 - b. A subsidy per child given to married couples.
 - c. An increase in the income tax rate.
 - d. A stiffening of immigration rules for domestic help and other unskilled labor (warning: some women are unskilled, too).
 - e. The invention of the vacuum cleaner.
 - f. A reduction in the amount of discrimination against women.
 - g. A law forbidding women to work in dangerous industries.
- 3. How would you interpret the following facts:
 - a. The rate of labor force participation of married women is higher for the more educated.
 - b. In families with more children the rate of labor force participation of the mother is smaller.
- 4. Assuming that all suppliers of labor have the same indifference curves, rank the following from least to most work encouraging:

- a. A gift of \$10,000 per person.
 - b. A floor on income of \$10,000 per person.
 - c. A proportional subsidy to the wage that equals \$10,000 per person in equilibrium (that is, as actually paid).
5. A cheap pill is invented that makes four hours of sleep a night equivalent to the present eight hours. *True or false:* If leisure and commodities are normal goods, people will work longer hours and will be paid a lower wage per hour.

True or False

- 6. If household output is taken to be a measure of economic welfare, this measure would rise less rapidly during cyclical upswings and fall less rapidly during cyclical downswings than real national income as we currently define it.
- 7. You would expect that a “gold rush,” in the form of panning for gold, occurred during the Great Depression of the 1930s.
- 8. By self-selection the wage that emigrants from Ireland to America get in America are likely to be higher than the wages their distant cousins could get in America if they, too, emigrated to America.
- 9. If the marginal valuation of leisure depends only on the amount of it consumed (and not on the amount of goods consumed as well) and if much unemployment in “normal” times consists of temporary layoffs—known to be temporary by both employees and employers—then unemployment insurance will increase the amount of measured unemployment.
- 10. At the point of backward bending in the supply of labor, the elasticity of demand for goods is 1.0. (*Hint:* Use the offer curve, not the supply curve or algebra.)
- 11. The amount of leisure might rise if the Scottish highlander of Figure 25.3 could work part time at home and part time in the factory, as he might if the factory were not down in Glasgow but up in Inverness.
- 12. Taking H to be all hours available, H_w working hours, H_L leisure hours, P the price of the commodity, Q its quantity, and w the wage, the whole income is $wH = wH_w + wH_L = PQ + wH_L$.
- 13. In view of Problem 12, if the income elasticity of demand for leisure hours is ϵ , then the income effect of a rise in w is $\epsilon(wH_L/wH)w^* = \epsilon(H_L/H)w^*$ (Recall that asterisks indicate rates of change.)
- 14. A negative income tax with a 100% marginal rate will induce some people above the Subsidy level of commodity income to move to Subsidy.
- 15. A proportional income tax reduces work supplied, even if the reduction in work supplied raises to some degree the wage of work in the market.

25.3 The Choice of Schemes of Payment

What to Read For Where do the budget lines facing a supplier of labor come from? How does overtime induce workers to work more for less? Is a piece rate always better than payment by the hour in getting workers to work? How can the leisure/commodity diagram be modified to show punishment as well as pay?

Explaining Budget Lines The various budget lines for helping the poor or taxing the nonpoor are social experiments, not social events. That is, they are the result of explicit planning,

not of the interaction of economic actors. Unless the economist can extend economics to the political reasons for the enactment of Aid to Dependent Children or the minimum wage or the progressive income tax, she cannot say much about why such budget lines are as they are. The ambition of a political economy is a noble one, but far from achieved.

On the other hand, the economist has readily at hand an explanation for the budget lines that businesses offer to their workers: They offer the budget line that makes them the most money. The one that makes them the most money depends on the character of the workers and the character of the competition for their services. The peculiarities of wage payments, in short, should be explicable in economic terms. The key to explaining them is to search for the mutual advantage in the peculiarity: Because both workers and employers must agree on it, both must be made better off.

Explaining Delayed Wages

Pensions, bonuses, stock options, and the like, for example, are budget lines with wages delayed, delayed in order to assure good behavior.

Q: In the seventeenth and eighteenth centuries, many American immigrants were indentured servants (in 1776 about a third of the white population were indentured servants or descendants of indentured servants). The terms were typically that the servant gave five or seven years of service and in exchange was given passage to the New World and subsistence while in service. Indentured servants could escape and would be more likely to do so the more experienced they were. At the end of their term, however, they received a grant of land or a cash bonus. Why?

A: To keep them to the terms of their agreement. A servant who escaped received no land or cash at the end of the indenture term. The greater the total pay piled up at the end of the term the less likely would the servant be to escape. But the passage to the New World (chiefly to the Chesapeake Bay region, incidentally) was a very expensive and early payment to the servant: Some decided therefore to escape anyway, though less than would have without the bonus. Pensions work the same way.

Overtime

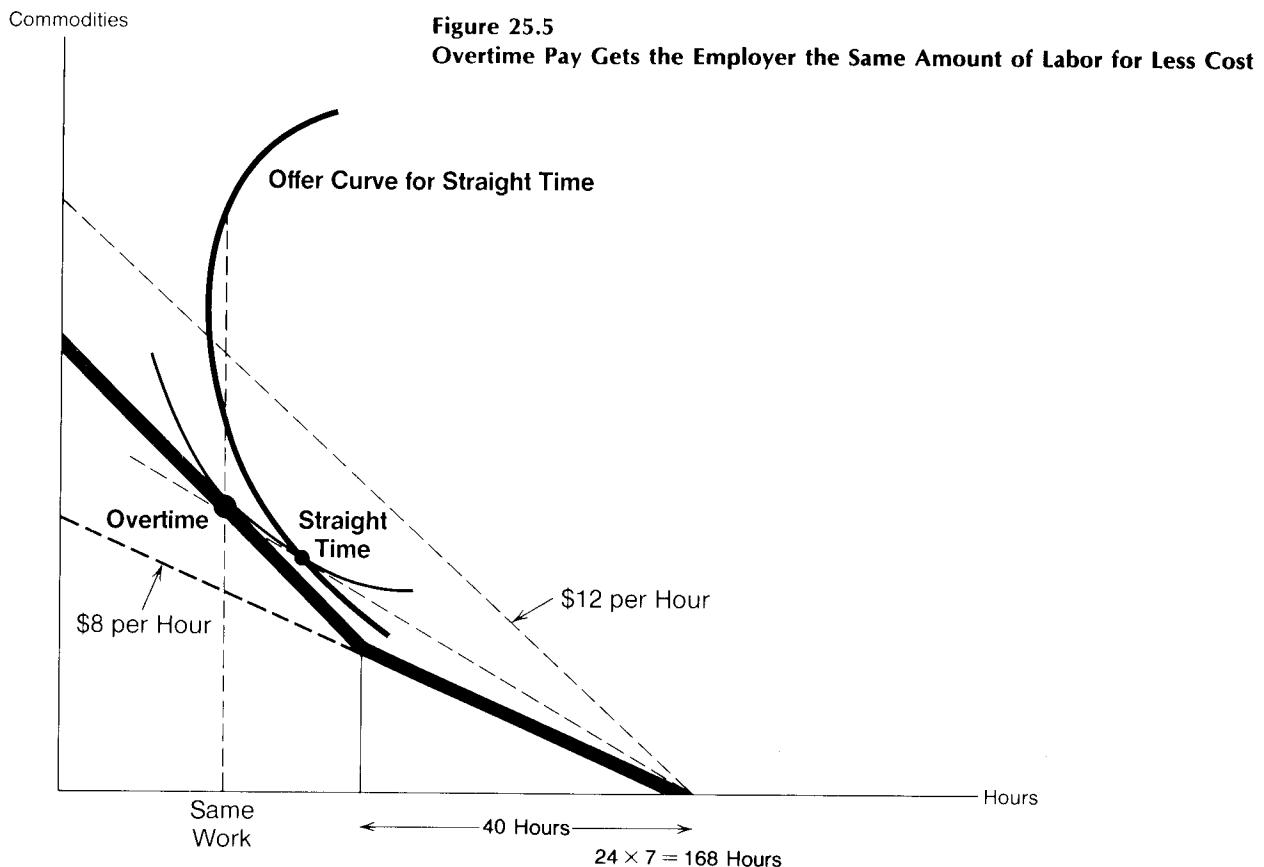
Overtime pay is another important example of how the wage offer itself is an object of choice between worker and employer. Steven Webb, who works as a cook, gets \$8 an hour up to 40 hours a week and \$12 for every hour above 40.

T or F: By having such a two-part system, the restaurant gives Webb less wages in total for the same work than it would have to give if it paid the same amount for all hours.

A: Webb chooses the point Overtime along the solid two-part budget line in Figure 25.5. Look at the point Straight Time, which is on the same indifference curve as Overtime. Since the dashed budget line tangent to Straight Time must have less slope than at Overtime (which

has a \$12 slope), the point Straight Time must be to the right of Overtime. The entire offer curve for Webb's straight time must therefore be above or to the right of overtime. If Webb's employer wants to get the same work out of him as at Overtime, he will have to move up the offer curve. But such points involve more pay to Webb, that is, a higher value along the commodity axis. If the restaurant pays straight time only, it will have to pay Webb more. In short, true.

Schemes of two-part pay are schemes of price discrimination, and a firm can do better if it can discriminate. The question remains why some other restaurant does not offer Webb the one-part scheme (straight time) that he prefers. The answer must be that the other restaurant would in fact offer less pay to Webb



in total, for some reason connected with the pay scheme. The discrimination does not actually make Webb worse off than he could be working for another restaurant, only worse off than he would be if the same restaurant would pay straight time.

Pay by Time Versus Pay by the Piece

The alternative to paying people by time is to pay them by the piece, that is, by how much they make. The choice between the two provides the last, best example of the uses of the commodity-leisure diagram. It is generally believed that people will work harder on a piece rate than on time pay, the argument being that a piece rate gives an incentive to put in the last ounce of effort. If you recognize the parallel to the fallacious argument that a renter is more energetic than a sharecropper, you will suspect that the argument is incomplete. It is.

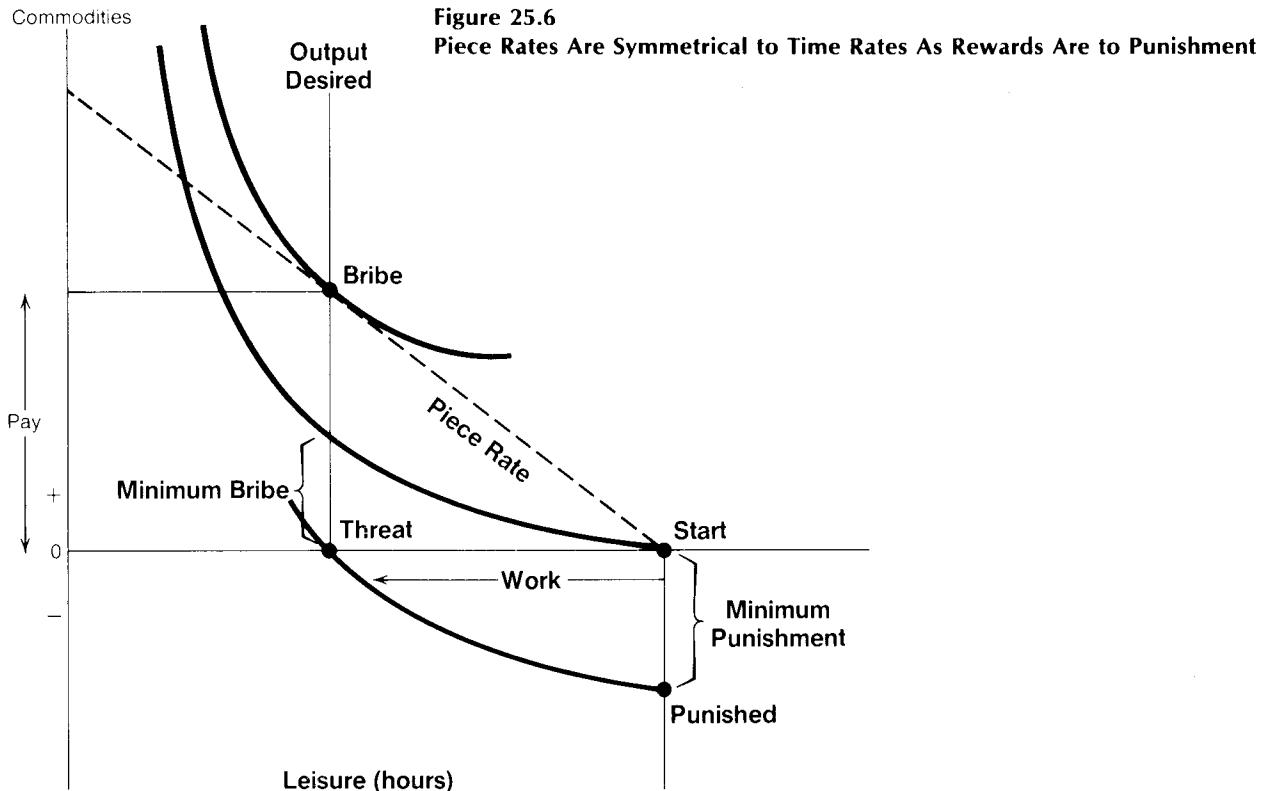
Suppose that Arcadius Kahan works in an eighteenth-century Russian iron-works. To keep the same axes as before, suppose that it is impossible to withdraw effort from an hour but that it *is* possible to withdraw hours (the analysis could be framed in terms of effort instead of hours, with similar conclusions). His shirking if he is paid by the day takes the form of cheating on the number of hours he puts in: He agreed to put in 12 hours a day, but can secretly withdraw

some of the hours. If he is paid by the piece, on the other hand, his input of hours is automatically confirmed, because each hour results in, say, a ton of iron and he is paid by the ton. It would seem that the piece rate is always superior from the employer's point of view.

Not so. As a first approximation both employer and employee are indifferent between the two systems (unsurprisingly, since both exist, sometimes side by side). The reason becomes clear when the two deals are made explicit. Under a piece rate in making iron, the employer presents Arcadius with a schedule of payments and lets him choose the output. In Figure 25.6, he moves along the dashed Piece Rate (remember: the "piece" here is merely an hour) to Bribe. (Ignore Threat.)

Under a salary (or time) rate the employer presents Arcadius with the output desired and with his total pay (determined by market competition) and *tells Arcadius that he will get fired if he does not perform*. This is the key. In general the negative prospect of being fired works just as well as the positive prospect of earning an extra ruble for the extra hour. Both systems end at the point Bribe.

Some minor advantage, not a gross difference in efficiency, explains why one system is in fact adopted over the other. For example, the option of threatening to fire a worker may be more advantageous to the employer if the worker has large costs of searching for a new job. As usual one must handle such an asym-



metrical argument gingerly, because one party to a mutual agreement cannot impose unfavorable conditions on another. The ability of each worker to choose his output may be inconvenient to a coordinated enterprise, such as an assembly line or an office. Such places will give salaries, while places that sew collars on shirts will give piece rates. The cost of monitoring output may be high—as in an office, again, while it is low if one has merely to count seven collars—in which case the more or less vague threat of firing can inspire self-monitoring by the worker personally. And so forth. The possibilities are many. The main point is that the reward of a schedule of piece rates is symmetrical with the threat in an agreement for a salary and can achieve the same result. Similar remarks apply to the schemes for worker participation, profit sharing, cooperatives, and so forth: More conventional arrangements can and do mimic their results.

Bribery Versus Punishment The symmetry runs deep.

Q: If an employer has the power to keep Arcadius from other job offers, what in Figure 25.6 is the minimum bribe he can offer to get Arcadius to produce the output desired? If the employer's power over Arcadius is still greater, so that he can *take* from Arcadius as well as give to him, taking off his hide with a knout if he has nothing else to be taken, what is the minimum punishment the employer can inflict sufficient to get Arcadius to produce the output desired? If punishing Arcadius, like rewarding him, is expensive, how would the lord (for with such powers that is what he is) treat the serf?

A: Look at the diagram. The analysis is identical to that of alternative taxes and subsidies. The indifference

curve through the Threat point defines how much of a punishment from Start is necessary to induce Arcadius to prefer the point Threat to the point Punished. The lord gets the work he wants. But likewise he could get it by bribery, the minimum bribe being the one of perfect wage discrimination. The lord can choose the carrot or the stick. If punishment is expensive (in overseers to give it, or in resentment and sabotage once it is given) then bribery may be better, even for a lord of serfs. The power to inflict punishment does not imply that punishment will always be chosen. The matter is one of economic, and moral, choice.

The government's monopoly of violence puts it in the position of lord of its citizens. Therefore, it too faces a choice between punishment and bribery. For example, in getting soldiers to defend our vital interests in Addis Ababa or Alicante or Antigua the government can choose to hire soldiers with a wage high enough to get volunteers or it can choose to draft them with jail sentences high enough to get compliance. Gangsters face the same choice: They can get what they want by offering you a good price or by making an offer you can't refuse, namely, to dump you in the river if you do not give in. In competitive markets, by contrast, the range of methods of competing is narrowed to mutually advantageous exchange.

Summary Budget lines do not come down from God, or from the Office of Management and Budget. They come from the behavior of people, people employing and people employed. The deal that employer and worker make must be mutually advantageous. Thus, a bonus at the end induces indentured servants to stay until the end of their term, reducing the hiring costs of the employer. Higher pay per hour induces workers to work more for less pay than a one-price system

would, but most offer some compensating advantage. Piece rates appear at first to extract more work than pay by the hour, but the threat of dismissal contradicts the appearance. The extreme use of the principle of mutual advantage is slavery: Here the principle still holds, except that the slave has and keeps zero advantage. Punishments and rewards are symmetric ways of dealing with slaves, or with citizens. The choice between them, too, is economic.

EXERCISES FOR SECTION 25.3

1. Write an *annual* wage contract to induce an agricultural worker not to leave the job during the harvest. (*Hint:* Look at the indenture contracts.)
2. Write a contract for a competitive equilibrium piece rate on sewing shirt collars. A sewn collar sells for 3 cents; your costs are $2\frac{1}{2}$ cents.
3. Write a contract for a weekly salary for a worker who can produce 20 sewn collars an hour for 12 hours a day six days a week. Are the results the same as for Exercise 2?

PROBLEMS FOR SECTION 25.3

- 1. Leisure is a good, work a bad. Redraw the analysis of piece rates versus time rates in a commodity-worktime diagram.
- 2. Give a diagrammatic interpretation of the following argument: "In the professions of law, medicine and the like, it is the reputation of enjoying a large practice which attracts new clients. Thus a successful barrister or physician generally labours more severely as his success increases." (*Hint:* Is the budget line as described straight?)
- 3. Molly McClelland is paid \$2.00 per hour for the first eight hours that she gives up from her daily leisure. She is paid \$3.00 per hour for any additional hours that she chooses to give up to get more income. When she can choose the number of hours to be worked, she works ten hours per day. *True or false:* Therefore, if a single price of \$2.20 per hour is paid for all the hours, and she can choose, she will surely choose to work less than ten hours.
- 4. The harvest work on each of many quite different fields in a medieval village could be done at the best time for each field "if each landholder enlisted the help of other peasants. . . . [But] each individual would spend much of his time working for his neighbors rather than for himself, and all the problems associated with working for others would significantly reduce the effective input of labor." Comment. The writer offers no evidence on how "significantly" the effective input would be reduced. Is an estimate necessary for his conclusion?
- 5. In preindustrial societies, such as northwestern Europe in the seventeenth century, the pace of work was slow and many days of the year were holidays. During industrialization, the pace of work quickened and the number of holidays declined; that is, the supply of labor hours increased. This increase has been explained in terms of rising supplies of energy (calories, say) making it possible for laborers to supply more effort.
 - a. If the total amount of energy, E , is unaffected by the wage, w , or consumption of goods, C —that is, if it is exogenously given—and if an hour of leisure, H_1 , requires e_1 units of energy and an hour of work, H_w , acquiring the consumption good, C , by working at a wage of w requires e_w units of energy, write down the energy constraint below which a worker must operate in C, H_1 space. In other words, write down the inequality connecting C and H_1 (consumption and hours of leisure) in terms of e_1 , e_w , and w . What is the slope (that is, what is $\Delta C / \Delta H_1$ for a given endowment of energy)?
 - b. Write down the market-wage constraint, that is, the trade-off between C and H_1 .

when the worker can purchase C by giving up H_1 at the rate w . What is its slope? Plot both constraints on a graph of C against H_1 . What do you conclude about the relative slopes of the two constraints if work requires more units of energy than does leisure? Show the attainable combination of C and H_1 . If both constraints are binding in equilibrium, where does a worker end up?

c. If wages remain constant and the energy constraint is always binding, what will happen to the supply of labor, H_w , if the amount of energy, E , increases? It is this effect that was mentioned earlier: More energy changes the amount of labor supplied. Does it matter if w increases at the same time E does?

d. Now suppose that energy is determined by the amount of C by the relation $E = aC$. Write down the new energy constraint. What is the new $\Delta C / \Delta H_1$ (the slope of the constraint in C, H_1 space)? Why does (a) $\times (w)$ have to be greater than e_w ? What, therefore, is the sign of the slope of the new energy constraint?

e. Graph the new energy constraint in the C, H_1 plane. Include the wage constraint and graph the attainable area. Show how it changes when w increases. If the energy constraint is always binding, what will happen to the supply of labor, H_w ? Does the introduction of an energy constraint, even if binding only at low wages (low w), increase or reduce the likelihood that effort will rise when w rises? (*Hint:* Look at the situation when the energy constraint is binding at low wages and ask whether eliminating the energy constraint would raise or lower the H_w supplied.)

f. What do you conclude about the plausibility of the theory outlined in parts (a)–(c)?

True or False

6. A draft is better than a volunteer army because it is cheaper. (*Hint:* To whom?)

- 7. A tax on all wages will decrease the labor supplied by less than a tax with the same yield to the Treasury on overtime pay alone.

CHAPTER

26

Capital's Supply and Demand

26.1 The Interest Rate

What to Read For Why is a dollar later worth less than a dollar now? What is the full rental rate of capital? What is the difference between nominal and real interest rates? What are discounting, present value, perpetuities, and the general discounting formula? How does the expected future return affect asset prices now?

The Opportunity Cost of Waiting

A dollar tomorrow is not worth as much as a dollar today, because there is a positive rate of *interest* (the next section discusses the reasons for the positive rate of interest). Because you can earn interest, you will always choose the dollar now if offered a choice between a dollar now and a dollar—*without* interest—a year from now. You could put the dollar now in a bank to earn interest at, say, 10%. The \$1.00 would become \$1.10 a year from now, bettering the dollar-for-a-dollar deal by \$0.10. To enter a deal that earns less than \$1.10 later for each \$1.00 given up now is to forego interest.

The point is one of opportunity cost.

Q: Improvements on late-eighteenth-century English farms were financed largely out of profits from farming, not from borrowing. *True or false:* Therefore the rate of interest on borrowing (roughly equal to interest on lending) was not a relevant cost to the farmers when deciding to invest.

A: Whether or not the money came from borrowing, the interest rate measures what the funds could have earned

in lending, a use other than investment in farms. When the interest rate is i , the opportunity cost of a fund of $P_K K$ (being, say, the cost of K feet of fences times the price per foot) is $i(P_K K)$. That is, false. Sources of funds are fungible, and the interest rate was relevant. When it was low, farmers would face a bad alternative to investing in the farm and would invest. When it was high, they would face a good alternative and would not invest.

The Full Equation for the Cost of Capital

The annual opportunity cost of investing in a machine, building, or other long-lived thing is in fact a little more than merely the interest foregone, though all the additions to a full accounting are similar to interest in being rates per

year. The question is, what is lost each year by spending $P_K K$ on K machines? The interest $iP_K K$ is lost. But so too is lost the wearing out or obsolescence of the machines: if these take place like evaporation of water, at a steady percentage rate of δ per year on the remaining value, then the whole "depreciation" is $\delta P_K K$. Finally, even if it did not wear out, the machine might increase or decrease in price over a year: if the rate of change of P_K is P_K^* , then the "capital gain" is $(P_K^* P_K K)$ per year, which offsets other costs if it is positive. The whole cost of capital is therefore $(i + \delta - P^*)P_K K$, the term $(i + \delta - P^*)P_K$ often being called the *rental rate of capital*, abbreviated R . That is, one could rent the machine from a competitive firm at R if R covered in a year the year's opportunity cost of interest plus evaporation of the amount of the machine minus the rise in value of a nonevaporated machine.

T or F: Since the interest rate was higher in the United States than in Britain during the nineteenth century, an investment would have to be more productive in the United States to justify doing it.

A: The cost of capital includes more than the interest

rate alone. In the United States depreciation might have been lower (it was not), capital gains larger (they were), or the price of capital goods, P_K , smaller (it may or may not have been). Although the assertion is true other things equal, it may not have been true unconditionally.

Uses of the Idea of Opportunity Cost

A project over the next year or the next month must earn, then, at least the rental rate on the capital invested. Holding some wheat in storage after the harvest in the Middle Ages is such a project. The earnings from it is its rise in price, the price becoming higher every month until the next harvest by the monthly opportunity cost of the money tied up in wheat. As usual, the argument is simply an equilibrium condition. If the price were not expected to rise from November 15 to December 15 enough to compensate wheat storers for the rental rate on the money invested for a month, the storers would remove some wheat from storage on November 15 and sell it. The sale would drive down the price on November 15. By reducing the store available to sell, it would drive up the price on December 15. The price curve could continue to tilt until the slope just equaled the rental rate, at which point further removals from storage would stop. A similar story from the other direction assures that a slope that was excessive would fall to the rental rate as well. Consider, then, the following.

Q: In England during the thirteenth and fourteenth centuries, 1075 pairs of prices of wheat in the same year and in the same village show an average increase of 2.7% per month. What was the rental rate per year? If depreciation (rotting) was 10% per year, what was the interest rate per year on safe loans in the Middle Ages?

A: The rate must be compounded at least each month to get the relevant figure. It is $1.027^{12} - 1 = 37.7\%$. Since the full rental rate (capital gains are not relevant here) is $i + \delta$, the interest rate, i , was $37.7\% - 10.0\% = 27.7\%$. By the sixteenth or seventeenth centuries, in contrast, it had fallen to 10%, a great change in the environment for investment.

The Real versus the Nominal Rate of Interest

The interest rates over 25% in the Middle Ages do not look high when compared with recent experience, rates 20% or higher for the prime customers of banks as recently as 1981. The high interest rates in the late 1970s and early 1980s, however, reflected inflation of prices in general, *not* the real rate. If prices in general are going up at 15% per year, the price of a sausage machine, say, will rise 15% in the year, reflecting the rising price of sausages and machines. A borrower can borrow \$100 to buy a sausage machine and then pay back \$115

in a year simply from the inflation. On the other side, a lender getting only 15% on his money is only just keeping up with inflation. Lenders will require *more* than 15%. And if the project of buying and using a sausage machine has real productivity, it can earn something real over and above mere inflation. If the prevailing real something is 4 or 5% (as it is in developed economies nowadays, in contrast to over 25% five centuries ago), then the “nominal” interest rate will be 20%:5% “real” interest plus a 15% allowance for inflation.

T or F: When countries such as Israel, Chile, and Brazil have annual inflation rates over 100%, you can expect to see interest rates there over 100%.

A: A 105% interest rate in an economy with a 100% inflation is really only a 5% interest rate. That is, the

goods you get back by investing goods in some project are only 5% greater, not 105%. Therefore, true. Your inducement to save and invest is the real rate, not the nominal rate. You give up present satisfaction for a future yield in the expectation of earning the 5% real rate.

The Ubiquity of Interest Rates

The interest rate on invested capital pervades a society, no less than does the wage of labor. The price of a future dollar in terms of a present dollar affects any choice to hold assets from the present into the future. For example, it pervaded the market for animals in the Middle Ages, confirming the astonishingly high interest rates revealed in the wheat market. Like holding wheat, the project of holding a flock of sheep for a year must repay the opportunity cost of the money invested in it.

Q: On one of the Bishop of Winchester’s estates in the thirteenth century the flock was worth £56 (about 22 person-years’ worth of sheep, to give an idea of the size of the enterprise), the sheep required little in the way of care and produced £25.4 of wool, pelts, cheese, and lambs per year. What is the implied interest rate?

A: If costs of caring for the sheep could actually be neglected, then it would be $25.4/56 = 45\%$ per year. In truth, there were some few costs, so 45% is somewhat too high an estimate. But it accords with the finding of high interest rates in the Middle Ages.

The Far Future Is Worth Less

Waiting for an investment to bear fruit, then, has an opportunity cost. For an asset not subject to depreciation and not expected to gain in capital value, the cost is the interest you would get if you lent the money. To put it the other way, in an economy with 10% interest rates you would be willing to pay only \$0.91 for the privilege of getting \$1.00 a year from now: $(1.10)(\$0.91) = \1.00 . Note the inverted order.

Such downward valuation of future income in the presence of an interest rate is called *discounting*. Many bonds, for example, are sold as promises to pay a \$1000 on a certain date, say, a year from now. The bonds sell now at a discount below their *face* or *par* value (\$1000), namely, at a discount just sufficient to earn anyone who holds them for a year the *going* interest rate. This is a condition of equilibrium in the market, not a conscious plan; arbitrageurs would rush in or out if the price were lower or higher. At 10% simple (that is, uncompounded) interest, for instance, the one-year bond would sell for P in the equation $(1.10)P = \$1000$, or \$909.09. As the date of *maturity* approaches, the price will rise. For instance, if maturity is only half a year away the price will be halfway from \$909.09 to \$1000. On the other hand, the farther away the date of maturity, the lower the price of the bond, because the price has to be low enough to earn that many years’ worth of interest.

T or F: If the interest rate is 10%, a project that will yield \$1000 five years from now will sell now for only \$620.92.

A: There are five years' worth of 10% interest to be covered by the eventual rise of the price from P to \$1000. Therefore, using compounding between years but not within them, P is the solution of $(1.10)(1.10)$

$(1.10)(1.10)(1.10)P = \$1000$. Your calculator reduces this to $(1.6105)P = \$1000$, or $P = \$620.92$. Therefore, true. A more sophisticated calculator can measure the daily compounding, which is in fact closer to how such prices are calculated. The correct price is a little lower, P in $(1 + 0.10/365)^{5(365)}P = \1000 or \$606.60 or a little different if you allow for leap years.

*Do
land
wastage
conservation*

The wider point arising out of discounting is that very distant payoffs in economies with very high interest rates are not worth having. We are urged to save oil "for our grandchildren." The policy is dubious. For one thing, our grandchildren may invent ways of doing without oil entirely—uncertainty about the state of technology or tastes in the far future is a reason for discounting it. For another, our grandchildren are very likely to be richer than we are, which means that the policy transfers oil from the poor to the rich. For still another—and the central point here—at 10% interest a dollar of benefit for a grandchild 50 years from now is worth now only P in $(1.10)^{50}P = \$1$, or *less than one cent*. To put it the other way, we would have to value each dollar's worth of pleasure by our grandchildren at over 100 times our own (117.4 times our own to be exact) to want to save oil for them. Even at 5% interest we would have to value them over ten times our own, a figure that does not accord with the actual behavior of even the most loving grandparents toward their grandchildren.

The Algebra of Valuing Distant Returns

The central idea is *present value*. The right to a stream of income in the future from any asset, such as a State of Iowa bond or a piece of land at Melrose Avenue and Grand Avenue Court, has a price today. The price today can be no more or less than the sum of money now that could earn the stream of income if invested in a bank at the going rate of interest. This sum is the *present value* of the stream. For example, suppose that a voyage to the Orient begun today will earn \$1000 a year from now. If the interest rate is 10% the present value of the \$1000 is, as you have come to expect, $\$1000 \div (1 + 0.10) = \909.09 . The right to collect the \$1000 earnings from the voyage could not sell for more than \$909.09. If it was quite certain that the \$1000 would be forthcoming, nobody but a fool would sell it for less.

The formula being used here (and earlier) is obviously $P = R_1 / (1 + i)$, where P is the present value, R_1 is the principal plus yield coming all at once one year from now, and i is the interest rate that could be earned on the sum P used to buy the stream. That is, $(1 + i)P = R_1$: the sum P invested for a year gives back itself plus iP at the end of the year. The formula will prove important in the next section.

The Perpetuity and Its Uses

The formula important in this section is the present value of a stream of annual returns of \$1000 lasting forever, a "perpetuity." In such a situation there is no repayment of the principal—or, if you wish, it is postponed until the end of time. The return is the net return, above principal. The present value of the stream is $P = R/i$, where R with no subscript is the annual return repeated into the indefinite future. You can see that the formula is true by rearranging it as $iP = R$: If a sum P is invested forever at an annual interest rate i , the

owner gets iP —call it R —on each New Year's Day forever. Using the formula, then, the present value of \$1000 forever (if the interest rate is 10%) is \$10,000.

T or F: A fall in the real rate of interest will increase the value of china plates relative to paper plates.

A: One can pretend that china plates survive forever, paper plates one year. The pretense is not too misleading: 20 years, as will be shown in a moment, is nearly forever at high interest rates. An equivalent stack of paper plates might well sit in a pantry for two years before being used up. In any case, the present value of the china is the use value (one meal), R , divided

by the interest rate, i ; the present value of the paper is $R/(1 + i)$, only one year's worth of discounting. The price of china relative to paper is therefore $(R/i) \div [R/(1 + i)] = (1 + i)/i$. If i is 5%, for instance, a china plate sells for 21 times a paper plate of comparable beauty and usefulness—which does not in fact seem to be far from the truth. If i fell to 3% the ratio would rise to 34.33. That is, true. A fall in the interest rate favors the use of relatively durable things relative to flimsy things.

It will now be clearer what Chapter 11 on the firm meant by saying that the firm seeks to maximize its “value.” Its value is its present value, that is, what the price would be of the right to collect the profits of the firm forever.

T or F: Since capitalists are mortal, but forests (if properly managed) are immortal, capitalism will lead to premature cutting of forests.

A: The reasoning suggests that Gary Walton the capitalist would not care about returns after his death. But Walton can at any time sell his ownership in the forest for the amount R/i , where R is the annual return from proper management, even if Death is at his side. He cuts off the future R 's and reduces now the value of

his forestry firm if he adopts “short-run maximizing.” In truth, what is best for the short run is what is best for the long, namely, maximizing the present value of the forest. Therefore, false. Capitalism is not shortsighted. Quite the contrary: a socialist country that for doctrinal reasons did not permit the taking of interest would be the shortsighted one, for there would be no way for future events to affect present incentives through present values.

The Annuity and Its Uses The formula R/i gives the price of a perpetuity, a stream of income R per year forever. The formula for any stream of income R_1 in the first year, R_2 in the second, and so on for N years, is

$$P = \frac{R_1}{1+i} + \frac{R_2}{(1+i)^2} + \frac{R_3}{(1+i)^3} + \cdots + \frac{R_N}{(1+i)^N}$$

This is the general discounting formula. The squared, cubed, and higher-order terms reflect the increasing force of compound interest at more and more distant dates. A special and especially useful case is an annuity, in which all the R 's are equal, say, R unsubscripted. The annuity formula can be written more neatly, it turns out, as

$$P = \frac{R}{i} \left[1 - \frac{1}{(1+i)^N} \right]$$

which is worth memorizing, so useful is it. Notice what it says: the present value of an annuity is some fraction of the value of the same amount in perpetuity (namely, R/i), the fraction being closer to 1.00 the longer is the number of years and the higher is the interest rate. At $N = 20$ years and $i = 10\%$, for instance, the fraction is 0.85. At $i = 20\%$ it is 0.97, which says that at high real interest rates 20 years is virtually forever.

The formula has many uses. For example, it illuminates the tragically important subject of slavery.

Q: 1. Slavery was abolished in Brazil in 1888. In the 30 or so years before abolition one could rent a 25-year-old slave for a year in Rio for 330 mil reis, which must therefore have been the annual earnings from a slave. If the rate of return on investments having the same risk were 25% and if a slave lived forever (effectively they did: the life expectancy of a 25-year-old slave was about 20 years more, and with such high interest rates this is "forever"), what would be the price of a slave?

2. As far as their owner is concerned, salves "die" when they are freed. What do you make of the following pattern of slave prices for 25-year-olds in Rio in the 1880s?

1880:	1223	1883:	589	1886:	54
1881:	1700	1884:	700	1887:	24
1882:	1128	1885:	543	1888:	0

A: 1. Apply the perpetuity formula: $330/0.25 = 1320$ would be the price. Notice that this is roughly the price in 1880–1882 (the 1700 of 1881 reflects perhaps temporary optimism about the future value of slaves, from a sudden rise in the price of their output—in this case, coffee).

2. Evidently *people anticipated abolition*, expecting by 1883 that it would come in the near future. Slaves continue to earn their annual rental regardless of whether or not they are going to be freed next year.

The Price of an Asset Contains an Estimate of Its Future Income

The problem shows well that the price of any asset embodies expectations about the future. Expected returns, as it is said, are *capitalized* into the value of the asset earning them, that is, made into a capital sum that can be bought and sold today. Someone with special foresight could make money by anticipating the future better, forming a more accurate estimate of the value of the asset. The money-making is limited, however, by the principle of entry and exit and its brother, the American question ("If you're so smart, why aren't you rich?"). The corollary here is that public information is already capitalized into the value of the asset.

Q: The price of stock in O'Grada Irish Creamery Ltd. falls as a result of a report announcing poor prospects for sales of Irish cream. *True or false:* The price fall makes the stock a bargain: you can make much money if you buy it.

A: False. The price is low now because the future for

So the price is the capital value of an annuity of 330 mil reis per year, R , for N years at an interest rate, i , of 25%. Using the formula for an annuity,

$$P = \frac{R}{i} \left[1 - \frac{1}{(1+i)^N} \right]$$

(in continuous compounding it is

$$P = \frac{R}{i} (1 - e^{-iN})$$

which gives virtually the same results), in 1883 the expected life of slavery was the solution of N in

$$589 = \frac{330}{0.25} \left(1 - \frac{1}{1.25^N} \right)$$

You can solve this by experimenting with N 's until you get about the right one. Or you can draw on college algebra and use logarithms. That is,

$$N = \frac{\ln [(R/(R-iP))]}{\ln (1+i)}$$

or

$$N = \frac{\ln [330/(330 - (0.25)(589))]}{\ln 1.25} = \frac{0.59}{0.22} = 2.65 \text{ years}$$

In 1884 their pessimism decreased— $N = 0.76/0.22 = 3.43$ —but they predicted abolition with accuracy. What is striking about these results is that there was no legal announcement that 1888 would bring abolition.¹

the company looks bleak. That is, the market has already capitalized the knowledge of reduced prospects into the price, making the knowledge useless as a profit maker for someone who hears about it late. Holding the stock will now earn only a normal return—as should be expected from only normal knowledge of what the future will bring.

¹ Pedro de Mello, "The Effect of the Abolition Movement on the Market for Slaves in Brazil, 1871–1888," in Robert William Fogel and Stanley L. Engerman, *Without Consent or Contract: Technical Papers on Slavery* (New York: W. W. Norton, forthcoming).

The point is that it is usually foolish to believe that one has access to information superior to that already capitalized into the price. In particular, it is exceptionally foolish to believe that one can profit much from a study of the pattern of movement of the price itself. A successful study would be a money-making machine of unlimited scope. There is usually little reason to believe that you are the first to seek and to find such a treasure.

T or F: Hog prices move up and down in a regular three-year cycle.

A: Suppose they did. Then Cathy McHugh, the hog farmer, could make money selling at the peak. But she

would have to believe, contrary to fact, that she had more information than did other hog farmers, such as Paul Munyon. The three-year cycle is there for all to see. All would act on it, eliminating it. The only rational expectation is that such cheap machines for making unlimited amounts of money do not in fact exist.

Summary To wait for a return without interest is to forego the interest. The cost of investing in a machine or other asset is the interest foregone, with depreciation subtracted and capital gains added. For example, stored wheat is an asset, which must have earned for its holders the interest rate plus the depreciation of the grain. The interest rate involved is a real one, corrected for the rate of inflation.

If the real rate is high, then, a distant prospect is worth little, flimsy things are worth nearly as much as durable things, and economic behavior is shortsighted, though properly so. Whether high or low the price of an asset—a bond or a business or a slave—embodies an opinion about the future, capturable in precise equations for discounted value. Only abnormal knowledge of the future can earn an abnormal return.

EXERCISES FOR SECTION 26.1

1. Calculate at 10% interest the present value of the following streams of R 's in three successive years:
 - \$1, \$1, \$1.
 - 0, 0, \$1.
 - \$1, 0, 0.
 - \$1.10, \$1.21, \$1.331.

2. The price of slaves in the American South during the Civil War remained at prewar levels for many months into the War. What do you conclude about the attitude of southerners?

3. Annual membership in a club is \$50. What should lifetime membership sell for, roughly, supposing a 10% interest rate and a long life?

4. "Stock splits" are advantageous to owners of stock in a company. That is, if people knew a split was coming they would be willing to pay more for a stock. The Securities and Exchange Commission prohibits "insider trading"—that is, buying and selling of stock by company officers likely to have especially good knowledge about whether things like stock splits are going to happen. What do you make of the tendency of prices of a stock to rise a couple of weeks before a stock split?

PROBLEMS FOR SECTION 26.1

- 1. The period of the French Wars (1793–1815 with interruptions) witnessed a large number of "enclosures" (expensive projects of agricultural improvement) in England. Yet interest

Interest rates rose up from 4% before the wars to $5\frac{1}{2}\%$ during them, leading some historians to doubt that interest rates had much to do with the decision to enclose. The rate of inflation rose from zero before to $2\frac{1}{2}\%$ during the wars. Are the historians' doubts well founded?

- o 2. Why do you suppose American railways were flimsy and British railways solid in the nineteenth century?

3. Slave prices in the United States rose during the 1850s. Did slaveowners anticipate emancipation?

4. Most emancipation of slaves in the North was gradual, freeing not the slaves but the children of slaves and often freeing the children only after some long period of "apprenticeship" (28 years of age in the Pennsylvania law of 1780, for instance). The owner of a female slave owned her children. Explain, in view of interest rates of 5 to 10%, why you are not surprised to learn that such emancipation cost slaveowners only a few percent of the value of their slaves.

5. News item, *The New York Times*, December 2, 1973, under the heading "Australia Again Bars Export of Her Prized Merino Rams": "The Australian Merino is considered the top wool-producer in the world as a result of more than a century of intensive selective breeding. In the current situation, with exports allowed, a fine stud Merino may sell for more than \$10,000 and some have brought many times that sum in auctions. Merino-raisers oppose export of the studs in the belief that increasing the number abroad would undercut the Australian wool market. Some experts dispute this, however." As an expert in economic analysis, if not in sheep raising, dispute this. In particular, would it be reasonable for specialized Merino-raisers to oppose exports?

True or False

6. Since modern corporations finance much of their investment from their own earnings, not from issues of bonds, the interest rate is irrelevant to their decisions.

7. The cost of capital is the same in Lahore and in Cleveland when the interest rates are equal.

26.2 Supply and Demand Curves

What to Read For

**How do markets fix the price of the factor of production, capital?
How is the choice between present and future consumption
represented diagrammatically? What is the separation theorem?
Internal rate of return? Human capital?**

Some Capital Controversies

The theory of interest is a theory of demand for capital, where "capital" is the mass of long-lived and reproducible implements of production such as ships, roads, machines, education, and houses. Materials, such as coal and cloth, are implements but are short lived. Laborers are not implements because they own themselves. Land is not reproducible.

An investor in capital wants to know what its rate of return will be, in order to choose whether to invest in it or to save the trouble and invest in a bond. The rate of return expresses the marginal productivity of capital. Just as a laborer should be hired to dig a ditch if his marginal product in that activity is greater than his wage (which is his marginal product elsewhere), so too a machine

should be built if its marginal product is greater than its opportunity cost. You can draw a demand curve for capital as you can draw one for labor.

Among the most controversial assertions in economic thinking is the assertion that capital has a demand curve like other things and that the interest rate is related to its price. The violence of the controversy suggests that something more than mere logic or fact is involved. What is involved is a metaphor describing the economy.

The simplest possible metaphor is elaborated here. Figure 26.1 is presented as a metaphor for the economy. You are warned to rely on it as one would on the assertion that the age from 1300 to 1600 was a "rebirth" or that the orbit of the moon around the earth is an "ellipse." Fortunately, less hinges on the literal truth of the capital metaphor than on the truth of these.

A Seminal Diagram

The metaphor is contained in the definition of the axes of a diagram of present grain plotted against future grain, shown in Figure 26.1. You will be asked in a moment to believe that the axes can represent all consumption possibilities for the economy in the present and future. In the meantime, the story of grain is entirely convincing. The curve that looks like a production possibility curve is just that: present grain can be put back into the ground as seed to yield more grain next year. Notice in particular that the investment of seed yields diminishing returns. The seed can be viewed as capital, K_t , in a production function relating output next year, Q_{t+1} , to inputs of capital and labor, L_t , this year: $Q_{t+1} = F(L_t, K_t)$. If one were portraying the whole economy, L_t would be fixed by population, and returns might diminish both with respect to capital and with respect to scale; if portraying one small sector, L_t would be varied optimally to match the rise in K_t , and only diminishing returns to scale could produce the bulge in the curve. In any event the slope of the curve is the marginal product of capital: it is the change in output caused by diverting some present consumption to investment in seed capital. It is in the same units as the interest rate plus 1: if the marginal bushel in seed now yields 1.15 bushels later, the return is 15%.

The other curves are of course indifference curves between present and future consumption. Their slope is impatience, known as the rate of time discount or time preference. If Ginalie Swaim's bushel now were just as good for her as her bushel later, she would exhibit zero time preference; her rate of time discount, which is the slope of the curve, would be 1.00. If on the other hand she required 1.10 bushels later to offset a loss of 1.00 bushel now, her rate of time preference (also analogous to an interest rate) would be 10% and the time discount would be 1.10.

If Swaim faced a marginal product of capital of 1.15 and a time discount of 1.10, she would want to invest more in seed. As she did so the marginal product would fall and the time discount rise, until at Equilibrium they were equal. At Equilibrium she has maximized her utility. In the usual fashion, it would be easy to draw below the present diagram a pair of supply and demand curves that were the slopes of the two curves and that intersected at their own equilibrium. The supply of and demand for capital would be equal at some slope, measured in units of 1 plus the interest rate. The standard theory of interest says essentially this: that the tastes for present as against future goods interact

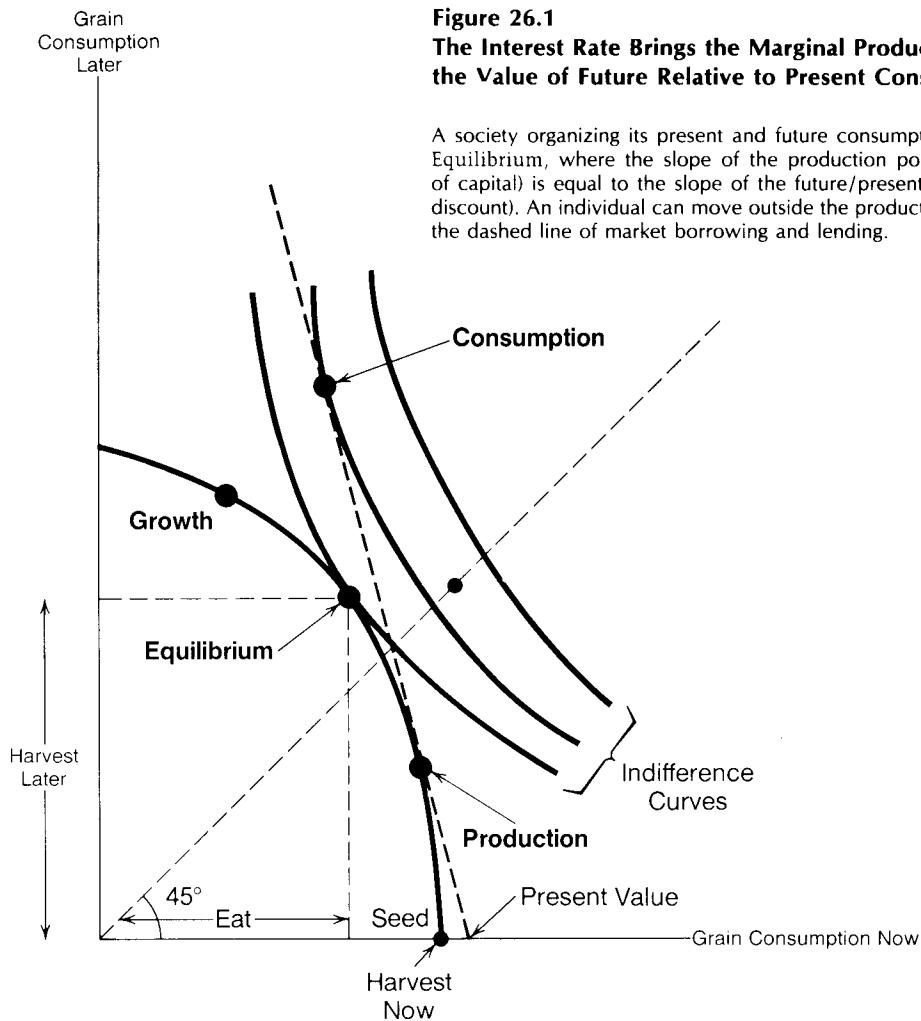


Figure 26.1
The Interest Rate Brings the Marginal Product of Capital into Equality with the Value of Future Relative to Present Consumption

A society organizing its present and future consumption as well as it can will pick point Equilibrium, where the slope of the production possibility curve (the marginal product of capital) is equal to the slope of the future/present indifference curve (the rate of time discount). An individual can move outside the production possibility curve by trading along the dashed line of market borrowing and lending.

with the technology of producing future goods from present goods to determine the price of future goods in terms of present goods, namely, the rate of interest.

The Seeds of Growth Theory

The point Equilibrium has been arranged to also illustrate the core of the standard theory of economic growth. Notice that the vertical distance Harvest Later happens to be equal to Harvest Now. If the harvest now did *not* reproduce itself in the future, the economy would contract. The production possibility curve between later and much later would start with a lower harvest, that is, with a lower income available for consumption (eating) or investment (seeding). On the other hand, if the Harvest Later exceeded the Harvest Now, as it does at the point Growth, the economy would enlarge in each period. The enlargement contains the seeds (as it were) of its own demise. As the production possibility curves move outward, the growth point moves closer to Equilibrium, reaching

it sometime in the future. The capital stock stops growing, and income stops rising, a condition known as the steady state.

Borrowing and Lending at Market Interest Rates

The economy as a whole must of course stay on its production possibility curve. But a single person can trade outside it, just as a single country can exchange wine for cloth. If the interest rate is high enough, for instance, the person will be able to produce at point **Production**, putting a little seed into the ground and loaning a lot of it at the high interest rate (which will be someone else's rate of time discount or marginal product of seed). Swaim will be able to consume little grain now in exchange for getting a lot to consume later. If the interest rate were lower than the slope at **Equilibrium**, she would be on the borrowing instead of the lending side: she would invest much seed herself in the ground, borrowing along the line of market interest to get back to a point of higher consumption now and lower later.

Uses of the Idea of Capital: Theoretical Points

If you accept the diagram as a metaphor of society, then it finishes the tale of the distribution of income by marginal product. Taste and technology determine the interest rate and how fast capital grows; the interest rate (plus depreciation) multiplied by the price of capital goods (grain here) determines the rental on capital. In short, markets determine the incomes of capitalists as well as of laborers and landlords.

The uncontroversial use of the idea is simply as a convenient summary of scarcity over time: you must, alas, abstain a little from consumption now if you wish to consume anything later; and the amount of abstention, if you are wise, maximizes the present value of your plan of production. Look back again at the diagram. The point **Production** is the best possible in a world with an interest rate equal to the slope of the dashed line minus 1. Notice the point at which the dashed line cuts the horizontal axis. It is the **Present Value** of the **Production** plan. It is also the highest value of the whole production possibility curve between present and future. Choosing any other production plan would get lower present value, because a line through the plan and parallel to the dashed line would intersect the horizontal axis at a smaller value.

Notice that the borrowing and lending line separates the decision of how to invest from the decision of how to consume. The proposition is known as the *separation theorem*. It says that a given slope to the market line determines the production point and also determines the consumption point. The two decisions are independent of each other. Michael Edelstein, the farmer, makes his decision of how much to invest in seed, machinery, and storage bins by choosing the investment (that is, production) plan that maximizes the present value of his farm at the going interest rate. Having done this he has maximized his wealth—"wealth" being the present value of all one's prospects and assets. He can then move along the market line of borrowing and lending out to wherever he is happiest. If the capital market (that is, the borrowing and lending market) were so poorly developed that he did not have access to a dashed line, he would be forced back into producing in each year exactly what he consumes, just as a country burdened by tariffs or transport costs is forced back into autarky. The utility gain from the capital market is the move from the point **Equilibrium** to the point **Consumption**.

**The Rule of Investment:
Maximize Net Present Value**

T or F: Because it pays back in one year rather than in two, a project costing \$1000 now and yielding \$2000 and zero in the next two years is better at a 10% interest rate than another project costing \$1000 now and yielding \$500 and \$1700.

A: Calculate the present value of the returns. The first project has a present value of $R_1/(1 + i) = \$2000/$

$1.10 = \$1818$. The second has a present value of $R_1(1 + i) + R_2/(1 + i)^2 = \$500/1.10 + \$1700/(1.10)(1.10) = \$455 + \$1405 = \1860 . That is, false. The businesswoman who chooses the second project gets higher wealth than from the first, by $\$1860 - \$1818 = \$42$, enough for a decent business dinner for one in New York City. The general point is that the criterion of *payback*, though commonly used, is irrelevant.

The wise businesswoman will in fact do both projects if they are not mutually exclusive, because both increase her wealth. Both cost only \$1000 now but give present values now of \$1818 and \$1860. Good deals. The argument can be taken further. She should do all projects that have a net present value greater than their present cost. Or, to put it another way, she should do all projects that add to her wealth.

The Internal Rate of Return Is Not an Infallible Guide

A somewhat backhanded way of satisfying the criterion is to find the interest rate, i , that brings the present value of the stream of benefits into equality with the present value of the stream of costs. The rate is called the *internal rate of return*. For example, for the second project it is the i that solves

$$\text{Present value of costs} = \$1000 = \frac{\$500}{(1 + i)} + \frac{\$1700}{(1 + i)^2}$$

= present value of benefits

The procedure for solving a quadratic equation learned in high school and since forgotten says that you first put the equation in standard form, $1000i^2 + 1500i - 1200 = 0$, then apply the forgotten formula to find that $i = -62\%$. When you have recovered from your surprise that an apparently desirable project has a large negative rate of return, you will remember that quadratic equations have *two roots*. The other is 112%. You will leave the exercise satisfied that a project earning 112% is worth doing when the opportunity cost of the funds is 10%, but a little wary of a technique so heavy in algebra that you cannot do it and so wide open to absurd results. Your wariness is justified: stick to present values.

The most engaging use of the economics of investment is to understand investment in human skills, for example, your own.

T or F: Since their present incomes are low, college students are poor.

A: Their present value—their wealth—is high, higher than that of many people of similar age who cannot

go to college and higher even than most older people who did not go. The students are at present investing, but that their consumption is low now does not mean that it will be low later; quite the contrary. Therefore, false.

T or F: In view of the question just answered, low tuitions at state universities are subsidies of the rich (the students) by the poor (the taxpayers).

Someone getting an education or learning skills as an apprentice or running 15 minutes three times a week is investing, investing in human capital. The very idea has been fruitful.

T or F: If education is painful, one will buy less of it than will maximize one's wealth.

A: The whole demand curve for years of education is the vertical sum of the marginal valuation of it as investment and the marginal valuation of it as consumption.

A given rising curve of opportunity cost for each successive year intersects with the summed curves at a number of years lower than the years that would maximize wealth (namely, the years implied by the marginal valuation of investment alone).

Another example of this sort of reasoning is:

T or F: It will never pay an employer to pay for the training of workers if after completing the training the workers can move to another employer.

A: If the training is specific to this one employer, it is trivially false. Telling your employees the location of the bathrooms in your plant does not raise their value to other employers. If the training is general and *does* make the employee more valuable to other employers,

the employee will recognize this. The employee will be willing to work for a wage lower than his or her current marginal product in the plant in which he or she is acquiring the education. In other words, the employer is acting as a schoolmaster and is paid for this service the difference between the employee's current marginal product and the (lower) wage paid. That is, false.

The metaphor of investment in human capital is itself derived from the metaphor of capital. Like capital, it is a powerful one.

Summary The idea of capital contains many puzzles, none of which have been discussed here. The image of capital accumulation as an investment of seed avoids the puzzles and keeps what is worthwhile in the story. What is most worthwhile is the notion that future consumption requires a sacrifice of some present consumption, that is, savings and investment. The best way of making the sacrifice is the way that maximizes the present value of production, for this way enlarges most the opportunities for consumption. A market line of borrowing and lending separates the decision of how to consume from the decision of how to produce. Any capitalist will do well to maximize present value, which is to say to maximize wealth. And workers can be seen as capitalists of labor, unable to sell their capitalized value but able nonetheless to augment it by investment and to borrow on its promise. The theory of labor and of capital are at bottom united.

² See Armen Alchian, "The Economic and Social Impact of Free Tuition," *New Individualist Review* 5 (Winter 1968): 42–52, reprinted in his *Economic Forces at Work* (Indianapolis, Ind.: Liberty Press, 1977); highly recommended.

EXERCISES FOR SECTION 26.2

1. What is the internal rate of return on a project lasting one year in which the present cost is \$1000 and the return in a year \$1100?
2. Polonius said to Hamlet "neither a borrower, nor a lender be." Where in Figure 26.1 is he recommending that Hamlet go and remain?

PROBLEMS FOR SECTION 26.2

1. Prove diagrammatically that the growth process described in the text converges to a point of tangency such as the point Equilibrium. (*Hint:* Use a compass to plot each successive harvest later back onto the Harvest Now axis, starting a new production possibility curve; use roughly parallel production possibility and indifference curves.)
2. The European Common Market announces a temporary one-month experiment in free trade for meat; this announcement increases the price of meat in Argentina and increases the quantity sold. Next month it makes the free trade permanent, to last by law for at least 15 years; *this* announcement further increases the price of meat in Argentina but *decreases* this year's quantity sold. What is happening?

True or False

3. If the rate of interest at which Paul Uselding can borrow is sufficiently above the rate at which he can lend, he will stay at the autarkic equilibrium.
4. If Eric Jones expects his future income to be larger than his present income, Jones will always save.
5. That human capital cannot be sold, or even partially precommitted at a far future date, will lead to underinvestment in it.
6. It can be argued that the exemption for students in college from the (now defunct) military draft in the United States would in the long run help those who did not go to college by changing the distribution of income.
7. If today or at any time in the next ten years you can make an investment of \$1000 that will yield a return of \$2718.52 ten years from now (in 1885), then you will make the investment now rather than later only if the interest rate is below 10%.

Appendix: Answers to Odd-Numbered Exercises and Selected Problems

CHAPTER 1

Exercises for Section 1.1

1. True. An hour spent finding out George Marr's background, likes and dislikes, habits, and so forth is an hour less for studying or watching a movie. You sacrifice something: friendship is desirable but scarce.
3. Mosk isn't, nor is anyone else. Or, rather, all are, because all are fungible.
5. False. The week of time could have been used another way. A choice was made to use it for a vacation, and now it is gone forever. The other ways of using it are part of the cost, just as much as are the other ways of using the money for the hotel and so forth.

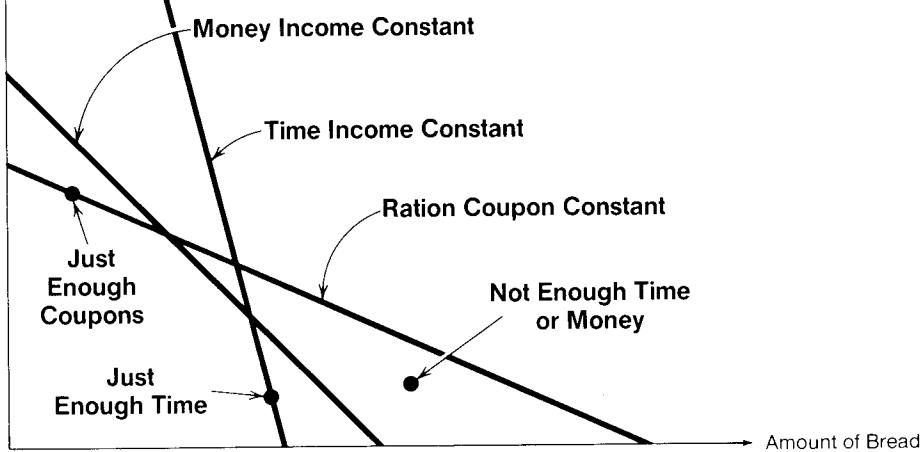
Problems for Section 1.1

4. a. There are three budget lines, but Jones can operate only along the crossed segments of each (otherwise he would exceed one of the budgets, as at a point like Not Enough Time or Money) (see Figure A1.1). This arrangement of the three lines is one possibility; the true arrangement depends on the relative prices implicit in the three currencies, money, time, and coupons.
- b. True. At Just Enough Time, for example, the rate at which Jones must surrender units of Housing to get one unit of Bread is higher than it is at a point such as Just Enough Coupons. The slope is steeper. The real price of Bread in terms of Housing has risen.
6. False, as common experience should tell you. To be sure, children do involve these money expenditures, represented by Money Budget in Figure A1.2. But children involve the use of their parents' *time* as well, which, like money income to spend on diapers and diplomas, is also limited. If time is involved in Other Things as well fit certainly is: eating, watching television, and the like require time as well as money), then the amounts "bought" of Children and of Other Things is limited by a budget line of time (shown in the diagram) as well as by a budget of money. The combinations of Children and Other Things that are available are those in the shaded area, for which *both* budget constraints are satisfied. Point Not Enough Time, for example, is not available: the parents could, to be sure, cut back their money expenditures on Other Things enough to "afford" 10 children, but, the diagram asserts, they would already have exhausted their budget of time; Not Enough Time is inside their money budget line, but outside their time budget line.

Amount
of
Housing

Figure A1.1
Many Budgets

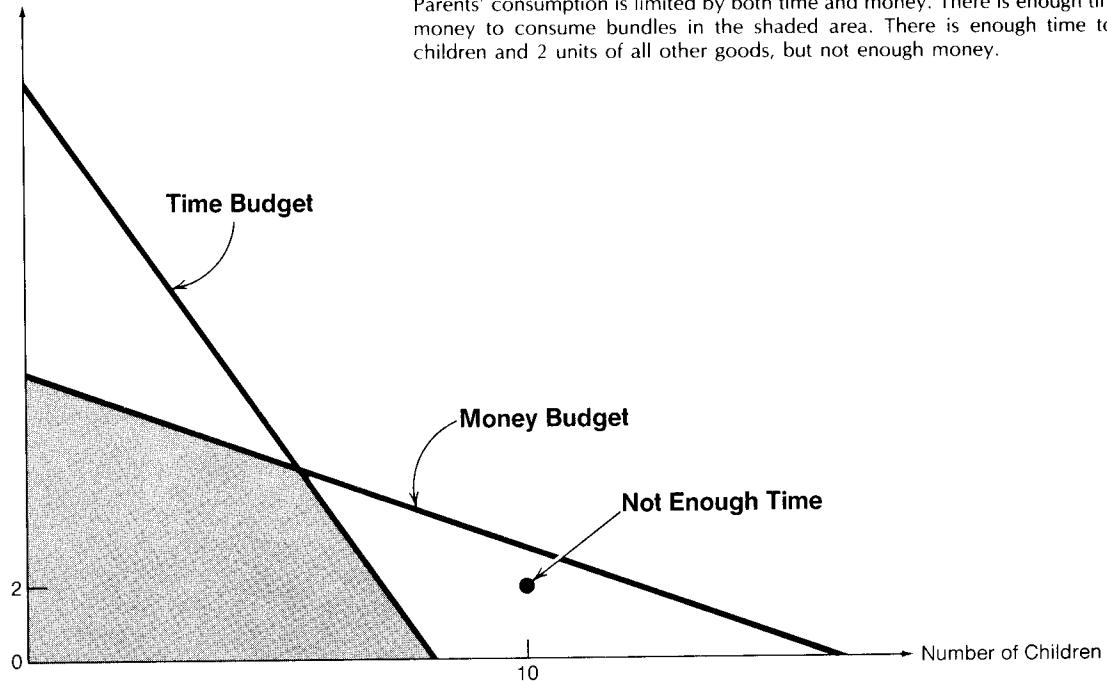
The consumer facing many constraints must satisfy all of them. The Just Enough Time and Just Enough Coupons points lie inside or on all three constraints and are in the consumer's opportunity set. The Not Enough Time or Money point is not. At the Just Enough Coupons point, the price of housing in terms of bread is determined by the slope of the ration coupon income constraint, since that constraint only is binding.



Amounts
of
Other
Things

Figure A1.2
Two Scarcities

Parents' consumption is limited by both time and money. There is enough time and enough money to consume bundles in the shaded area. There is enough time to consume 10 children and 2 units of all other goods, but not enough money.



COMMENT The diagram is directly applicable to point rationing, such as that used in the United States during World War II. Butter, sugar, and tires cost money then as now, but they also cost ration points. That is, one needed a certain number of ration points in addition to the 50 cents in money to purchase a pound of butter. Such diagrams are applicable, in fact, to any situation in which more than one type of currency is used to buy commodities and in which one "currency" (dollar bills, ration points, hours of time, or whatever) cannot be converted into another. "Currency," "buy," and "commodities" are all to be defined broadly. In the normal sense you buy a skiing vacation in Colorado with money for plane tickets, hotel, equipment rental, and lifts; but you also buy it with the use of your time or energy or other opportunities foregone in the act of consumption. A good grade in an economics course is bought with purchases of books and allotments of sheer hours of reading, but also with allotments of attention and mental energy within the budget constraint of the student's attention span and intelligence.

If currencies were convertible it would not matter how much of each one had. That one has ten \$5 bills rather than fifty \$1 bills is of no consequence if the two types can be exchanged, as they can. There is one budget constraint, not two. In the Child/Other Things problem it is implicitly assumed that hours devoted to childraising or consuming the Other Things cannot be transformed into money income: but of course some of them can, by selling the hours in the market at a money wage. The method for handling such (partially) interdependent constraints is developed later.

Exercises for Section 1.2

1. You would rather have had the \$80, since the best use you can put the lamp to is to sell it for a mere \$20. The gift makes you \$20 richer, ignoring any riches of mutual love that cheerfully accepting and keeping the thing may give. You can buy \$20 more of movies or food. Her gift does not move your budget line out by \$80 because you weren't already in the market for hideous lamps.

3. True. People have different incomes. But they all buy Cadillacs and corn for the same price. At least that is what the simple theory assumes. It can be modified, though, to allow prices to vary from person to person. For instance, people in California pay less for vegetables relative to beef than do midwesterners.

5. You want to solve the first equation for Q_C . To isolate it on one side, bring the other term, $P_A Q_A$, over to the Y side, changing it (of course) to negative. In other words, subtract $P_A Q_A$ from both sides. The result is

$$P_C Q_C = Y - P_A Q_A$$

Get the P_C out of the left side by dividing both sides by it. The result is what you want.

Problems for Section 1.2

7. The money constraint is $P_A A + P_F F \leq Y$; the coupon constraint is $c_A A + c_F F \leq C$. (Recall that " \leq " means "is less than or equal to.") When coupons can be bought for money (or money bought for coupons) the income in money (\$ Y) actually spent is equal to the original endowment of income (call it Y^*) plus whatever is "earned" by selling coupons. If coupons are bought instead of sold the "income earned" is negative. In other words, if the quantity $C^* - C$ is the amount of coupons sold and if P_C is the money price of coupons, then it is true by the definition of money income that $Y = Y^* + (C^* - C)P_C$.

Now because of the buying and selling of coupons the consumer will always be at a corner point at which *both* of the budget constraints hold true. Were this not the case, the price of coupons would change until it was the case. Look at Figure A1.3.

A point such as E_0 cannot be an equilibrium. Only coupons are an effective constraint at E_0 , because the consumer is inside the money budget line. Therefore the consumer would take some of his excess, worthless money and buy coupons with it. The result would be that the coupon budget line would move out and the money budget line would move in until the point of rest for the consumer was in fact at a corner, as in the right-hand panel of Figure A1.3. At a corner point both coupons and money

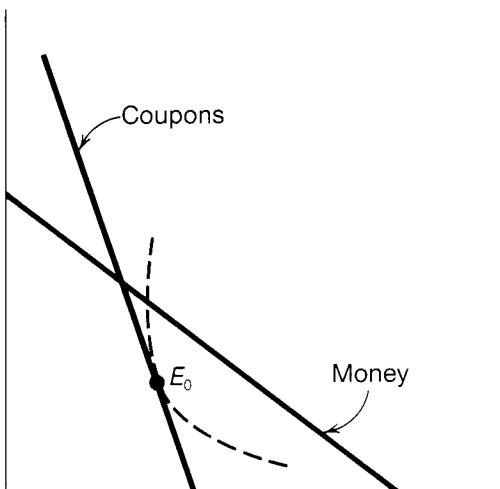
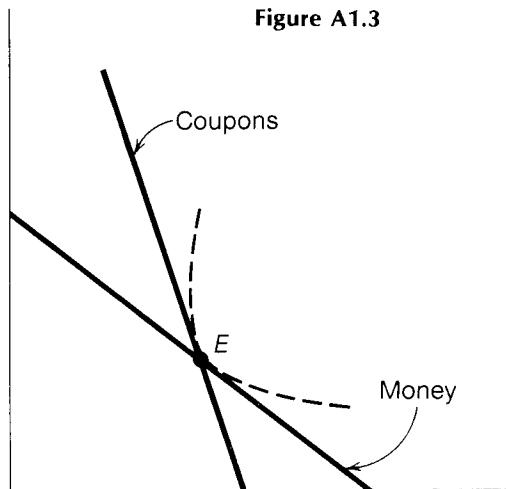


Figure A1.3



are scarce, and there's no further incentive to trade one for the other. The budget lines stop moving. Both hold true as equalities, not inequalities.

The three equalities in equilibrium are, to repeat:

1. $P_A A + P_F F = Y$
2. $c_A A + c_F F = C$
3. $Y = Y^* + (C^* - C)P_c$

Eliminate the variable Y by setting the left side of (1) equal to the right side of (3), getting:

$$P_A A + P_F F = Y^* + (C^* - C)P_c$$

Then use (2) to substitute out the variable C in this equation, getting

$$P_A A + P_F F = Y^* + (C^* - c_A A - c_F F)P_c$$

This equation can be rearranged to make it look more like a budget constraint:

$$(P_A + P_c c_A)A + (P_F + P_c c_F)F = Y^* + P_c C^*$$

The first set of variables is the full price of All Other Goods in money: it is the money price plus the money equivalent (the number of coupons required times the price of coupons) of the coupon price. You can see that there is a similar price for Food. The right-hand term is full income: it is money income plus the money equivalent of coupon income.

The argument can be applied to a much more prevalent situation than ration coupons, namely, the situation in which there is a time as well as a money price of consuming things. It takes, for instance, \$7 plus a half hour of time to get a haircut, \$2 plus 10 hours of time to buy (and read) a paperback novel, 60 cents plus five minutes to eat an ice cream cone. In such an application the income side of the equation will be one's nonlabor income (that is, returns from bonds or rents from land) plus labor income, the notion being that labor income arises from trading time for money, just as coupon income comes from trading coupons for money. The labor income will be the total amount of hours available for work or consumption multiplied by the wage in dollars per hour. Rearranging the equations will result in prices involving money plus the money equivalent of the time used. These are the true, full prices of goods.

Exercises for Section 1.3

1. The assertion is meaningless. You should refuse to answer meaningless questions. An ounce of gold is, to be sure, more expensive than one haircut. But who said we were dealing with ounces?

One millionth of an ounce of gold is in fact cheaper than a haircut. But aside even from the absence of units in the assertion, gold and haircuts don't serve the same function. So it doesn't make sense to talk about the cheapness of one compared with the other.

3. False. It will be less. A constant price of transport per orange, as in the Sterno–Napoleon brandy problem, raises the high-priced item less relatively. Therefore, by the Law of Demand, relatively more high-quality oranges will be consumed in the Northeast.

Problems for Section 1.3

3. True. The meaning of "the same tastes" is that the two groups can be treated as one consistent consumer, facing in Germany and Italy different conditions of income and prices. The question, in other words, is whether or not the German and Italian data satisfy the Weak Axiom of Revealed Preference. Well, the German coal miners could have bought the Italian bundle of goods (it is cheaper than the one they in fact bought), but did not. If the Italian bundle in Italy were revealed preferred to the German bundle the WARP would be violated. But in fact the German bundle could not have been purchased by the Italian coal miners (the German bundle is more expensive at Italian prices than is the Italian bundle). The Italian bundle is not revealed preferred to the German bundle. The WARP is not violated and one cannot reject the hypothesis that the differences in the things Italians and Germans buy (Italians more spaghetti and the Germans more bratwurst) arise from the lower relative price of, say, spaghetti in Italy or the higher incomes of Germans, rather than from national differences in tastes.

CHAPTER 2

Exercises for Section 2.1

1. a. A fall in the price of corn will have the same effect as a rise in the price of books. The relative price of corn falls, which implies that the demand for it will increase (and for books fall). The budget line swivels counterclockwise, as in the problem in the text.
- b. A fall in Zecher's income causes his budget line to move in parallel. If corn and books are normal, the quantity demanded of both will fall. The new combination of corn and books will be southwest of the old.
- c. A decline in his desire to read has the effect of shifting his indifference curves. His tastes have changed. Look at Figure 2.1 in the text and imagine a set of indifference curves with a tangency at the point Corny.

Problems for Section 2.1

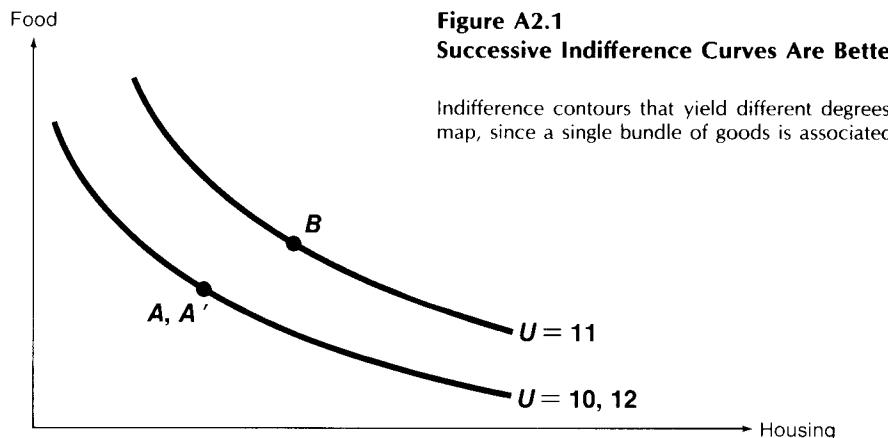
2. True. The contours around an overhang might be as in Figure A2.1, in which one contour represents two different altitudes, one directly above the other. On a real mountain a climber might well start at A at altitude 10, climb to B at altitude 11, and continue (with some difficulty) to A' (directly above A) at altitude 12. But were he climbing the hill of utility such an overhang would imply that B (with more of both food and housing) is inferior to A' ; and, further, that the identical bundles A and A' gave differing utilities.
5. Surprisingly, true. So do $U = (BC)^3$ or $U = \alpha + \beta(BC)^{1/2}$ or, in fact, any function of (BC) whose graph rises without ever falling (that is, is "monotonically increasing"). To take the simple case $U = \alpha + \beta(BC)$, the expression to be maximized (having substituted for B by using the budget constraint) is

$$U = \alpha + \beta \left[\frac{Y}{P_B} - \frac{P_C}{P_B} C \right] C = \alpha + \beta \frac{Y}{P_B} C - \beta \frac{P_C}{P_B} C^2$$

The derivative of this is to equal zero, that is,

$$\frac{dU}{dC} = \beta \frac{Y}{P_B} - 2\beta \frac{P_C}{P_B} C = 0$$

But notice that β can be divided into both sides, leaving exactly the same expression as for $U = BC$.



Exercises for Section 2.2

1. Notice that to answer the question you have to decide what each pair is used for. In different uses they may be substitutes or complements, permitting a variety of answers.
 - a. In covering tall buildings, substitutes. In small buildings, with standard windows, complements.
 - b. Substitutes, which is an important point about the "monopoly" power of steelmakers.
 - c. Substitutes for individual drinking, unless your intent is to get sick. Complements if you are supplying a large party, a certain predictable percentage of which will be wine drinkers and the rest beer drinkers.
 - d. Complements, since exactly two molecules of hydrogen match with each of oxygen.
 - e. Substitutes. The fact is again important for the "monopoly" of American production by three makers. Not much of a monopoly, which Toyota, Volkswagen, and Fiat can muscle in on.
 - f. Substitutes. Oil or coal.
 - g. Complements. Paper and pencils.
3. At a price of oats equal to about 62% of the price of barley. The prices move together in this ratio.

Problems for Section 2.2

2. True. At a given real income the proportion in which the consumer consumes butter with his bread may be insensitive to their relative prices. But as his income rises he may well spread the butter thicker, as in Figure A2.2.
3. True, if he can buy Heroin at a constant price in terms of Bread (that is, if his budget line is straight, as in Figure A2.3. At E he is on the highest attainable indifference curve. Were the other end of the budget line at B_1 he would be indifferent between Heroin and Bread addiction; were it at B_2 he would specialize in consuming Bread. The case of Heroin shows that concavity and the specialization implied by it are not always unreasonable. The marginal valuation of heroin rises rather than falls as more is consumed relative to bread. Points such as E are called, for obvious reasons, "corner solutions."

Exercises for Section 2.3

1. The Taxed Cost is *inside* the Full Cost, just as the Subsidized Cost is *outside*. All lines start from the same point on the axis of all other goods (if no calls are made—which is the case on the All Other Goods axis—no subsidies or taxes are paid).
3. The budget line is parallel to the old, but *starts* at Public Education. No longer must the consumer choose.

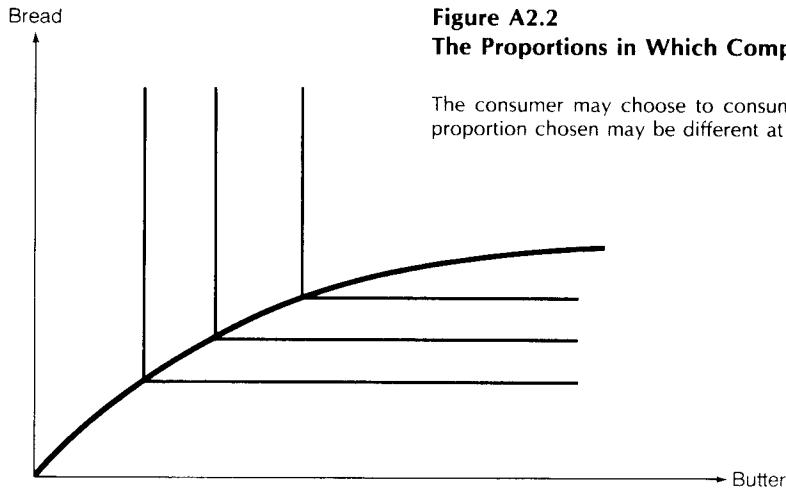


Figure A2.2
The Proportions in Which Complements Are Consumed May Change

The consumer may choose to consume bread and butter in a fixed proportion, yet the proportion chosen may be different at different levels of income.

Exercises for Section 2.4

1. a. The marginal benefit of information is the increment to the value of knowing it, the marginal cost is the sacrifice of time, money, and so forth in acquiring the increment. Think of searching for ideas for a term paper. You search until the two are equal.
- b. The marginal benefit and cost of *price* information is a clear example: one more hour of cost spent looking for the best deal on a standard wool skirt has the cost of whatever else the hour could be used for and the benefit of having a chance of getting a better price. Again, you search until the two are equal.
- c. The marginal benefit of another 10 seconds of weeding is another weed sent to its Maker. The marginal cost is the increment to that growing pain in your back. The optimal point to stop is when the extra gain is just worth the extra pain.
- d. You join the line when the marginal benefit is greater than the cost. The marginal benefit is the movie; the marginal cost is the time in line. The cost grows as the line gets longer. The line stops getting longer when the cost equals the benefit, that is, when the line is so long that no one else cares to join it.
- e. When the marginal benefit to the Eager Student's Education equals the value of the increment to weariness of the Distracted Author's Brain, he stops, as right now.

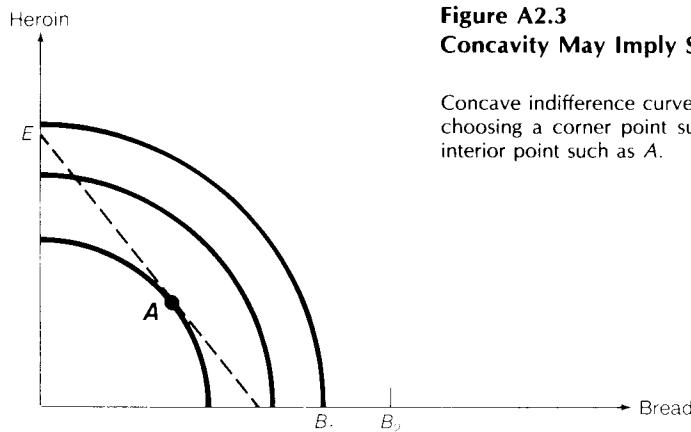


Figure A2.3
Concavity May Imply Specialization in Consumption

Concave indifference curves plus a straight-line budget constraint result in the consumer's choosing a corner point such as *E*, at which only Heroin is consumed, rather than an interior point such as *A*.

Problems for Section 2.4

4. False. Consider the alternative expression for the two equations from marginal utility theory: $MU_F/MU_H = P_F/P_H$. P_F/P_H is fixed. But if the marginal utility of food to Zecher depends only on the amount of food consumed and the marginal utility of food declines as more food is consumed (with the marginal utility of housing remaining unchanged), then equality cannot be maintained. Food and housing cannot therefore be normal goods to Zecher.

5. True. Solving for Margaret's demand functions for food and housing using the method presented in the Appendix to Section 2.1 yields

$$H = \frac{Y}{P_H} \quad F = \frac{Y}{P_F}$$

These demand functions say that Margaret will increase her amounts of housing and food (prices held constant) in proportion to increases in her income, Y . Further note that the demand function for one good is not a function of the price of the other good. Margaret will therefore spend the same shares of her income on housing and food regardless of their prices or her income.

CHAPTER 3

Exercises for Section 3.1

1. All but (d), which has *increasing* marginal utility. All the others have decreasing marginal utility, which is all that is necessary to get Coatsworth to stop eating hot dogs eventually. The rate at which the utility falls or the "units" in which it is measured don't matter as long as utility isn't measurable anyway.

Exercises for Section 3.2

1. Because Field gets the same bundle as at Lucky (his house is not flooded) *but must pay insurance money* to get it for sure. The insurance payment expressed in pounds of hamburger is shown in Figure 3.2.

3. Set up the gamble versus the sure thing:

$$\frac{1}{1000}(U_{\text{winning the lottery}}) + \frac{999}{1000}(U_{\text{safe}}) = U_{\text{lucky}}$$

That is, he could stay at Lucky for sure if he chose to. We know the utilities of Safe and Lucky already. Substituting them in leads to

$$\frac{1}{1000}(U_{\text{winning the lottery}}) + \frac{999}{1000}(910) = 1000$$

Solving for $U_{\text{winning the lottery}}$ yields

$$\left[1000 - \frac{999}{1000}(910)\right]1000 = 90,910 \text{ joys}$$

Problems for Section 3.2

2. a. He would choose the gamble. At these odds the utility of the safe, present income is below the utility of the gamble itself plus the utility of gambling:

$$36.25 < \frac{19}{20}(35) + \frac{1}{20}(40) + 2 = 37.25$$

b. You would not know that Van Hoorn got 2 joys from the act of gambling. If you assumed that he got no joys from it, you would calculate his utility from his present income as

$$\frac{19}{20}(35) + \frac{1}{20}(40) = 35.25, \text{ which is less than } 36.25$$

That is, you would underestimate the true utility of the safe choice by one joy, because Van Hoorn gets utility from gambling. When there is no gambling involved Van Hoorn gets 35, 36.25, and 40 joys from the three incomes; when there is gambling involved he acts as though he values safe options less than he in fact does.

c. Evidently, a love or hate of the act of gambling itself will cause you to mismeasure Van Hoorn's utility. As was noted in Footnote 3, this is the chief criticism of the gambling method. Many a gambling addict can tell you that the criticism is important.

Exercises for Section 3.3

1. He is just indifferent between Safe and the gamble of a $\frac{9}{10}$ chance of Lucky and a $\frac{1}{10}$ chance of Unlucky. So

$$U_{\text{safe}} = \frac{9}{10}(2000) + \frac{1}{10}(1000) = 1900 \text{ joys}$$

3. The dollar taken from Butlin is worth less to him than is the same dollar given to Field: Butlin's shaded column in Figure 3.4 in the text is shorter than Field's. So if one's idea of a "good idea" is to create the highest *sum* of utilities of the two, a Robin Hood policy is a good idea.

Problems for Section 3.3

3. True. As long as Rich's income is higher than Poor's, a dollar taken from Rich has less marginal utility than one taken from Poor. Any taxing of Poor will result in more lost utility than taxing of Rich. If the government's lust for revenue exceeded the difference in income between Rich and Poor, then the best tax would drive Rich's income down to Poor's and thereafter take equal amounts from each.

COMMENT

This problem is an exercise in hedonistic utilitarianism, the principle that society should be organized to give the greatest happiness to the greatest number. The greatest exponent of this self-contradictory principle of greatness (one cannot in general have the greatest of two things simultaneously, such as the greatest family life and the greatest career in economics) was the English philosopher and economist Jeremy Bentham (1748–1832). Interpreting it to mean simply the greatest sum of happiness, the sum is maximized by thoroughgoing equality of income: as long as Rich earns a dollar more than Poor, the sum could be made greater by equalizing their incomes. But the sum is meaningless. Utilitarianism is intellectually and morally questionable.

Exercises for Section 3.4

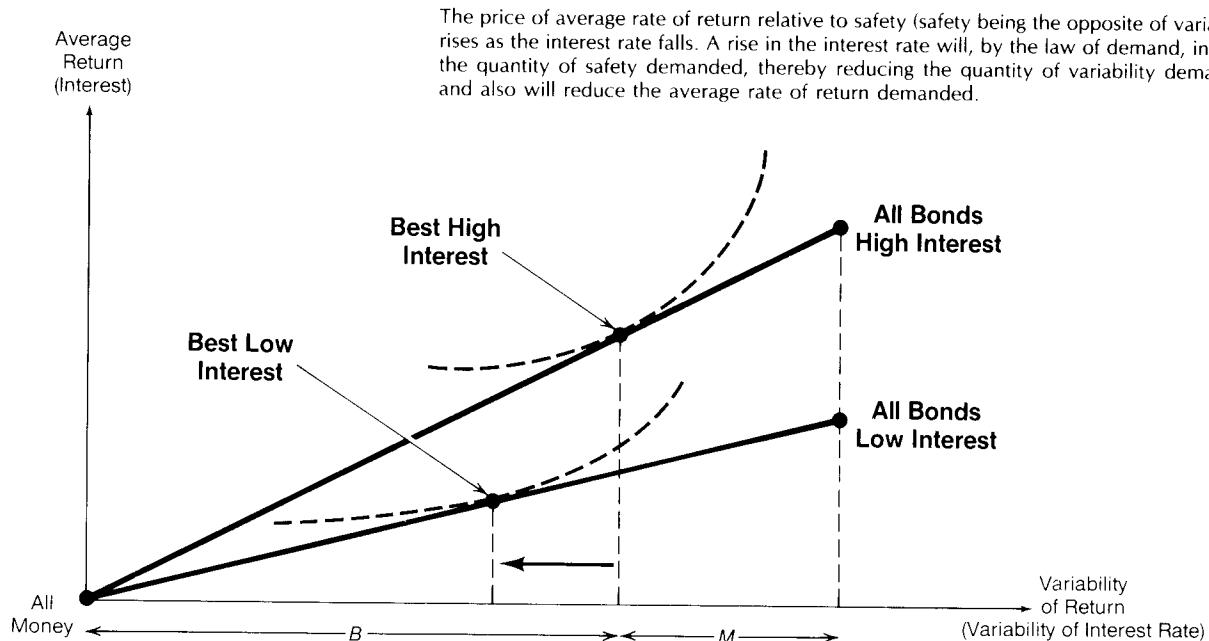
1. True. Each move to a higher indifference curve would increase utility by larger and larger amounts. The increase for one unit of all other goods (that is, the slope of the total utility curve) would be getting larger and larger as more income was achieved. The total utility curve would be concave *upward*, not downward.

Problems for Section 3.4

1. a. The average and variability of return on money balances are both zero. That is, if he puts all of his \$100 in them he is at the point All Money in Figure A3.1; if he puts all of it in bonds, on the other hand, he is at a point such as All Bonds High Interest. He can pick any point on the line between the two and will pick Best High Interest. The proportion he puts in money can be measured along the Variability axis as $M/(M + B)$ (or, for that matter, by a similar construct, along the Average axis: the measure is the same because both Average and Variability for the combination he chooses move in proportion to the amount put in bonds).

- b. If the average return (interest rate) falls he has a new budget line, that is, All Money–All Bonds Low Interest, and chooses a new point, Best Low Interest. In the case portrayed in Figure A3.1, the proportion of money balances in his \$100 rises (look at the heavy arrow). That is to say, a fall in the interest rate has induced him to hold more money. The demand for money is sometimes called the demand for liquidity (or liquidity preference). What the diagram says is that because holding bonds is risky, a fall in the interest rate on bonds increases liquidity preference. A relationship between the interest rate and the amount of money demanded plays a large role in theories of inflation and unemployment: risk aversion is one justification for such a relationship. The justification was discovered by James Tobin, in his "Liquidity Preference as Behavior Towards Risk," *Review of Economic Studies* 25 (February 1958): 65–86, reprinted in many places, for example, M. G. Mueller, ed., *Readings in*

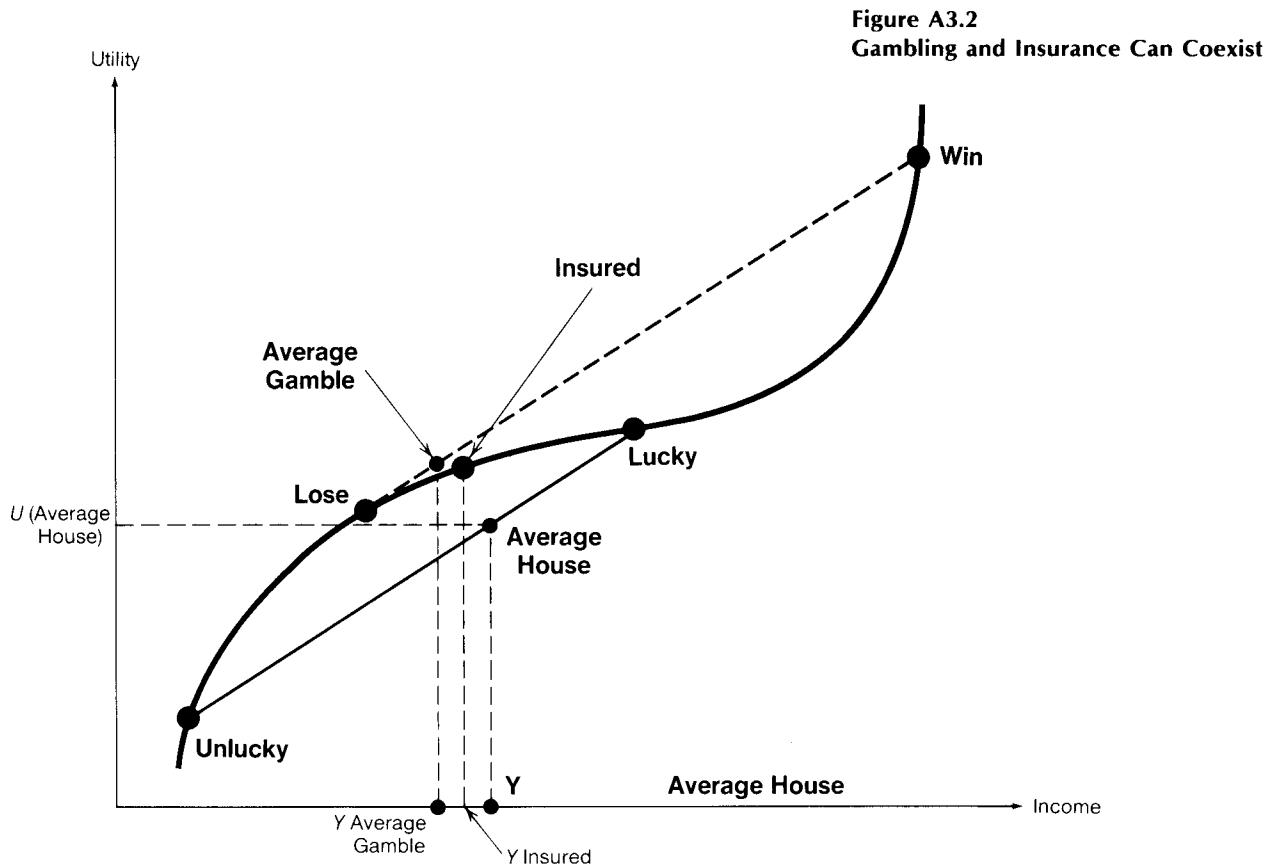
Figure A3.1
Liquidity Preference as Behavior Toward Risk



Macroeconomics, 2nd ed. (New York: Holt, Rinehart and Winston, 1971). The intuition behind the diagram is simple. Money is safe (ignoring the hazards of inflation) and a bond is risky. Safety is a good, risk a bad. But to get the safety of money one must give up another good, average return. The average return is the opportunity cost of more safety, that is, the price of holding money. By the Law of Demand, as the price goes down more is demanded.

COMMENT With other indifference curves the point Best Low Interest might be to the right of Best High Interest instead of to the left, in which case the theory would break down. As Tobin said, "The ambiguity is a familiar one in the theory of choice, and reflects the ubiquitous conflict between income and substitution [price] effects" [Mueller, ed., p. 185]. A fall of interest rates does reduce the price of safety facing the investors, but it also reduces his income (if bonds yield 1% rather than 10%, investors are obviously made poorer). If safety is a normal good he will demand less of it with a lower income. And the income effect, which leads him to demand less safe money balances, may offset the price effect, which leads him to demand more. Bear in mind this example of the significance of price and income effects: it will come up in various guises throughout the book (in, for example, the next chapter).

2. False. Draw the hint for David Bender in Figure A3.2. If Bender's house does not burn down he receives the income at Lucky, if it does burn down he receives the income at Unlucky; on average he receives Average House, which has a utility of only $U(\text{Average House})$, because it is risky. If the insurance company offers him unfair insurance, that is, the riskless income Y Insured (look at the income axis), he will take it, for its utility is higher. Now he is at the point Insured on his utility of income curve. Suppose that he is offered an unfair gamble between Lose and Win, with an average payoff of Average Gamble. If the utility of Average Gamble is higher than that of Insured, he will take the gamble, too.



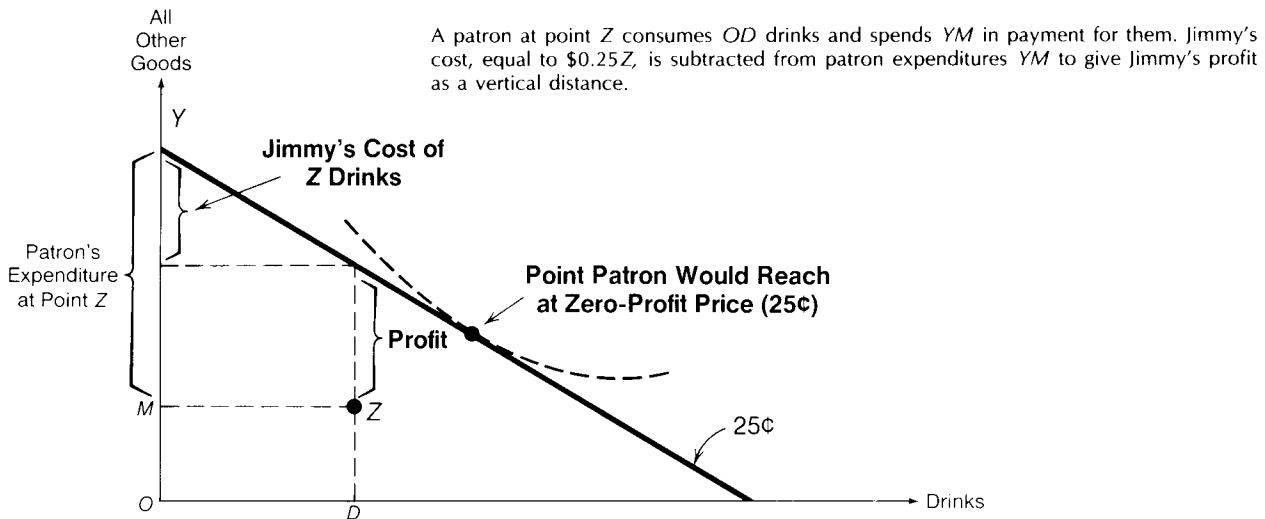
COMMENT Until 1948 many economists believed that the answer was “true”: that is, they believed that simultaneous gambling and insuring were impossible for a man who did not get utility from the acts of gambling or insuring themselves. Still other observations of how people behave in the face of risk can be accommodated by adding further regions of convexity and concavity.

CHAPTER 4

Exercises

1. Sure. Bankruptcy sells transport at a very low price in terms of all other goods. Wouldn't you like such a deal?
3. Yes, the railway could say: “You accept Fee Point or we're not going to do business with you.” Fogel would accept, if there was a cent of improvement in it.
5. Emphatically not. Although related one to the other, they are not even plotted within the same axes: one is within axes of quantity of Y against quantities of X ; the other is within axes of the relative price of X against quantities of X .
7. No. For each you can imagine a price at which you would not buy, or would buy more. Take the hardest case. If the price were your entire income forever (enslavement), you might take your chances with acute appendicitis: before the operation, after all, some people survived it. On the other

Figure A4.1



hand, at \$2.50 per operation and no cost of inconvenience you would have it when you felt the slightest pain in your stomach—why not? At such a price you might as well be safe.

Problems

3. a, b. Look at Figure A4.1.
- c. Jimmy cannot force the patron below the indifference curve through Income (see Figure A4.2), because if he attempted to do so the patron would simply not patronize his establishment (the patron would be better off with no drinks). The Point of Maximum Profit is at the tangency parallel to the cost-price budget line.
- d. The cover charge would be Fee in Figure A4.2 (equal, you can see, to Maximum Profit) with a

Figure A4.2

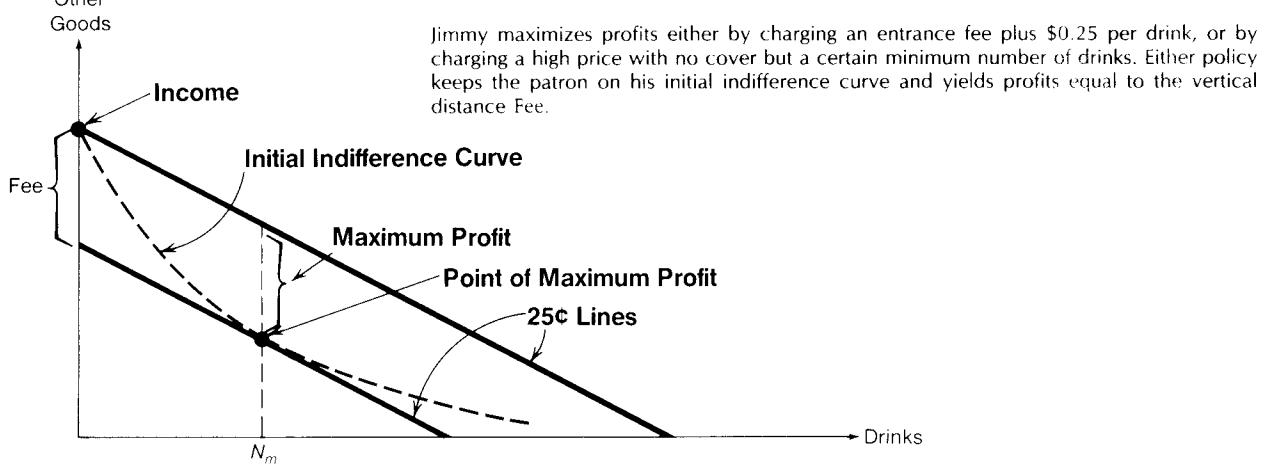
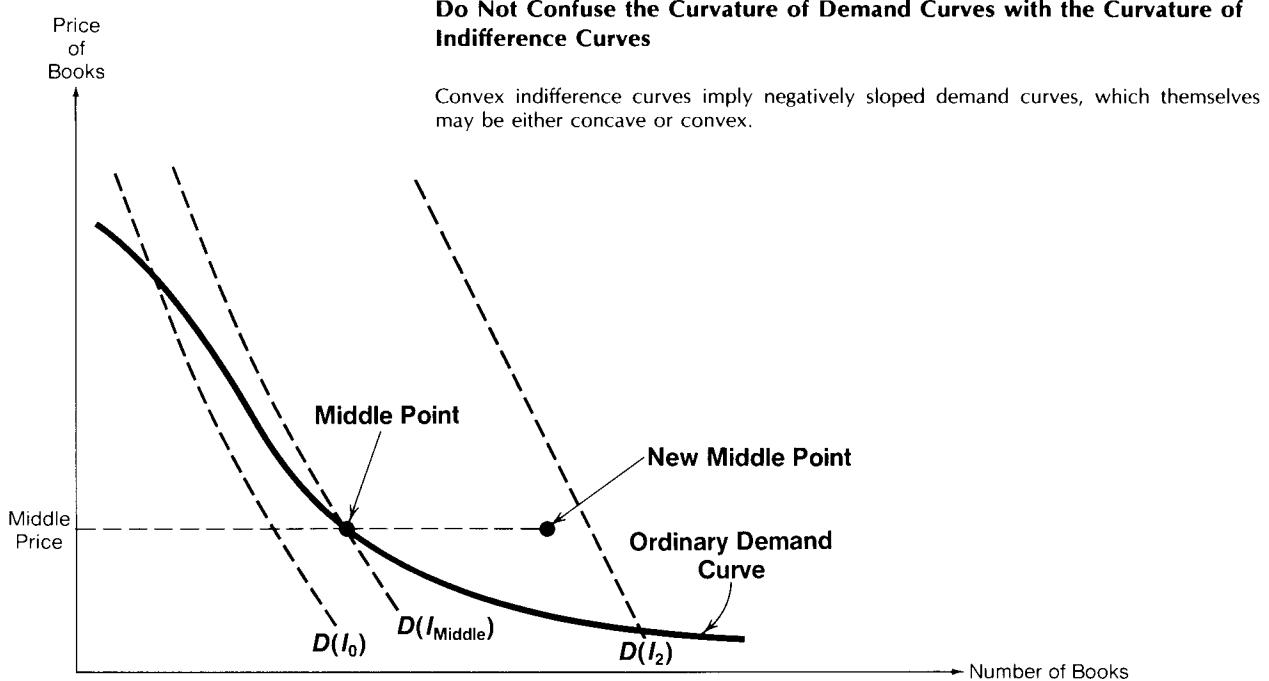


Figure A4.3
Do Not Confuse the Curvature of Demand Curves with the Curvature of Indifference Curves



price of 25 cents per drink. The patron would buy N_m drinks. Alternatively, Jimmy could get the same result by requiring the patron to buy at least N_m drinks at a price represented by the slope of the dashed straight line through Income and the Point of Maximum Profit.

e. Such fees or all-or-nothing offers, common in the retailing of liquor by the drink, break down if bars must compete with each other (assuming that the cover or minimum is not a charge for something other than drinks, such as entertainment). To entice business away from another bar Jimmy will reduce his fee; but so will the other bar. Competition will tend to drive the bars to sell drinks at 25 cents with no cover and no minimum.

5. See the discussion in Chapter 25 on the choice between work and leisure.

7. False. Suppose that the demand curve runs through the compensated demand curves in a convex fashion, as in Figure A4.3. Imagine (because they are not drawn here) the indifference curve corresponding to the compensated demand curve running through Middle Point. Slide it along the Middle Price budget line (remember: you are now in a Books/All Other Goods diagram, not the one here) toward the Books axis. This will increase the number of books bought at the Middle Price, that is, in the present diagram it will move Middle Point rightward to, say, New Middle Point. Connecting up New Middle Point with the two other (unchanged) points will produce a concave rather than a convex demand curve. The indifference curve has been changed in location but not changed in shape, that is, it is still convex. In other words, convex indifference curves are perfectly capable of implying concave demand curves.

8. True. Income-compensated means "along the same indifference curve." But along the same indifference curve no price change alters the rigid fixation of the consumer on the corner points. Look at the various budget lines running through the point Z in Figure A4.4. The first vertical (dashed) line in the bottom diagram is the demand curve for housing when income is compensated to keep the consumer on the indifference curve (L-shaped) through Z: His demand for housing is the same at any price. Even in this extreme case the ordinary (as distinct from the compensated) demand curve has an ordinary

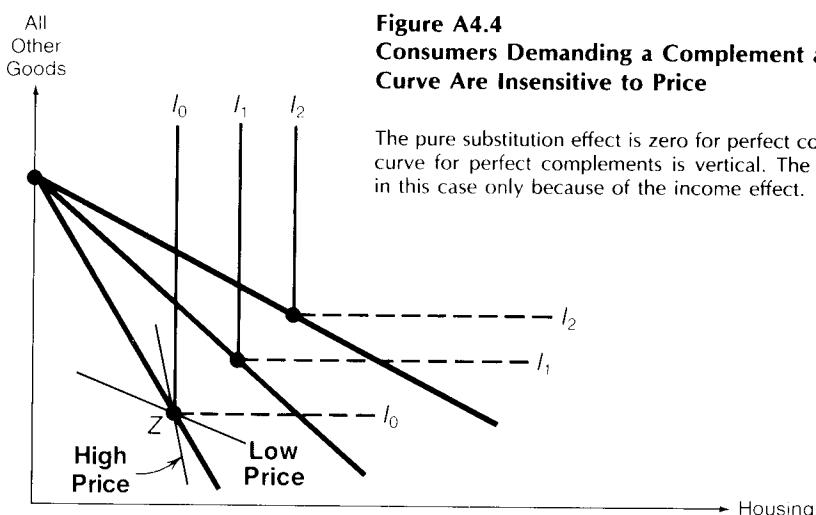
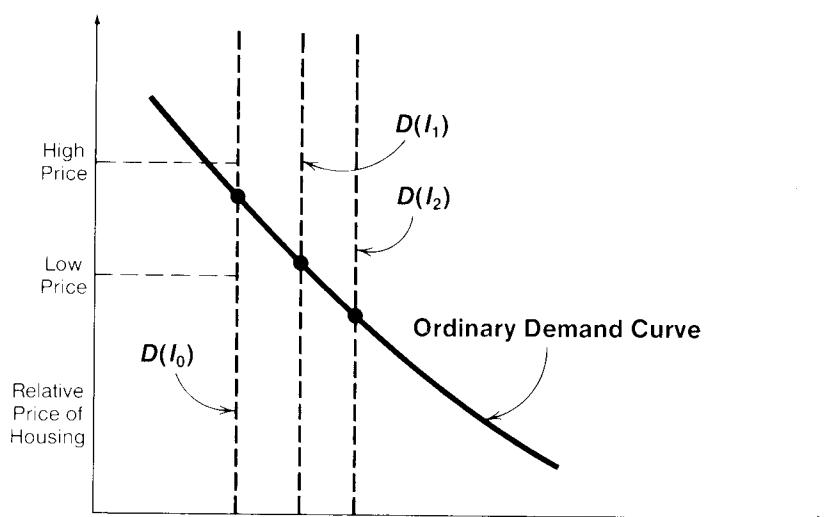


Figure A4.4
Consumers Demanding a Complement and Held to a Given Indifference Curve Are Insensitive to Price

The pure substitution effect is zero for perfect complements and the compensated demand curve for perfect complements is vertical. The ordinary demand curve is price sensitive in this case only because of the income effect.



shape. All its slope is an income effect since the substitution effect for perfect complements is zero (which is to say that the compensated demand curves are vertical).

CHAPTER 5

Exercises for Section 5.1

1. He would have to acquire wheat or computers or whatever else it was that the Italian sellers of paintings would want. Money economizes on this.
3. No. As the answer says, "Self-sufficiency is a matter of prices." He could have a zillion bushels of grain and 4.88965 pounds of wool and the same would hold: If his marginal valuation of grain was higher than the going market price even when he had a zillion bushels, he would still buy grain.

Problems for Section 5.1

2. If she bought wheat at 5 bushels of wheat per book she would get *before taxes* 5 bushels for each book given up. But after the government took its cut in taxes she would get only $2\frac{1}{2}$ bushels, not enough to compensate her for the sacrificed book (she demands 4 bushels per book). And a market price of 3 bushels for one book would be even worse. Only if the *before tax* price is *more than* 8 bushels per book will she care to sell books (that is, buy wheat). If she entered the market to buy books, on the other hand, she could get at most $\frac{1}{3}$ of a book for a single bushel (or worse at the other price = $\frac{1}{6}$ of a book). Since she demands $\frac{1}{4}$ of a book to compensate for the sacrifice of 1 bushel, the deal looks good. But again the government takes a cut, half of the $\frac{1}{3}$ of a book, leaving her only $\frac{1}{6}$ of a book—not enough to make the deal worth her while. Only if the 50% tax left her with more than $\frac{1}{4}$ of a book would she want to sell wheat (that is, buy books). Fifty percent of $\frac{1}{2}$ is $\frac{1}{4}$, so any number of books *more than* $\frac{1}{2}$ book per bushel of wheat would do the trick. In short, at a price of books *more than* 8 bushels per book (the earlier result) she will sell books. In between she does nothing: taxes, like transport costs, hinder trade.

5. a. Haines and McQuaid alone would clearly make one transaction; Haines, McQuaid, and Menard, three. Four people would give six transactions, as you can see by arranging four dots in a rough circle and counting how many lines it takes to hitch every dot with every other. By the same method, five gives 10 transactions, and six gives 15.

b. The total number of links for a society of N people is $(N - 1) + (N - 2) + (N - 3) + \dots + N - (N - 1)$: for 5, it is $4 + 3 + 2 + 1 = 10$; for 6, it is $5 + 4 + 3 + 2 + 1 = 15$. It can be shown that for very large N (say, 10,000 = a very small city), the number of links is approximately N^2 (100 million for $N = 10,000$). Those who have had calculus will be able to see the argument:

$$(N - 1) + (N - 2) + \dots + N - (N - 1) = N + N + N + \dots - 1 - 2 - 3 - \dots - (N - 1) \\ = (N - 1)(N) - \sum_{N=1}^{N-1} N. \text{ The limit of the last expression as } N \text{ gets very large is } N^2.$$

c. There are now N transactions, one for each person with the single General Store. That is, for a society of six people, six transactions with Gordon as against 15 without him. For a society of 10,000 people, 10,000 transactions as against 100,000,000 without him. Hurrah for Gordon! I don't know about you, but I feel just great about middlemen now.

Exercises for Section 5.2

1. The students do rent the rooms, in the way Pynchon's debtors took loans from him. Evidently, then, they think the three-bedroom apartment is worth the \$500 a month they give up for it. They enter the deal voluntarily with their eyes open. What's the beef?

Problems for Section 5.2

3. False in general, although at some point of great difference it becomes true. To see why, consider the Edgeworth box in Figure A5.1. The size of the lens-shapes coming out of a certain endowment can be taken for the moment as a (nonrigorous) measure of the potential gains from trade. One must have some measure of the difference in endowments, that is, some measure of the sameness in endowments. One possibility is the distance from the point Equal exactly in the middle of the box, the point of equal absolute endowments of pipes and gas in West Germany and the Soviet Union. Another is the distance from the dashed straight line bisecting the box, the line of equal proportional endowments. Whichever is used, it is true that associated with the endowment A is a larger lens than that associated with B , which is a more nearly same endowment than A . But for the endowments A' and B' , on the other side of the Contract Curve, the assertion is false. What is true is that the gain is larger the farther one is from the Contract Curve.

6. True. Look at the stories portrayed in Figure A5.2 (all from the initial endowment E).

Exercises for Section 5.3

1. No. There is no cheaply available alternative for either of you. Each of you can alter the price of water in terms of food by threatening not to bargain at all, or by whining, or by talking smoothly.

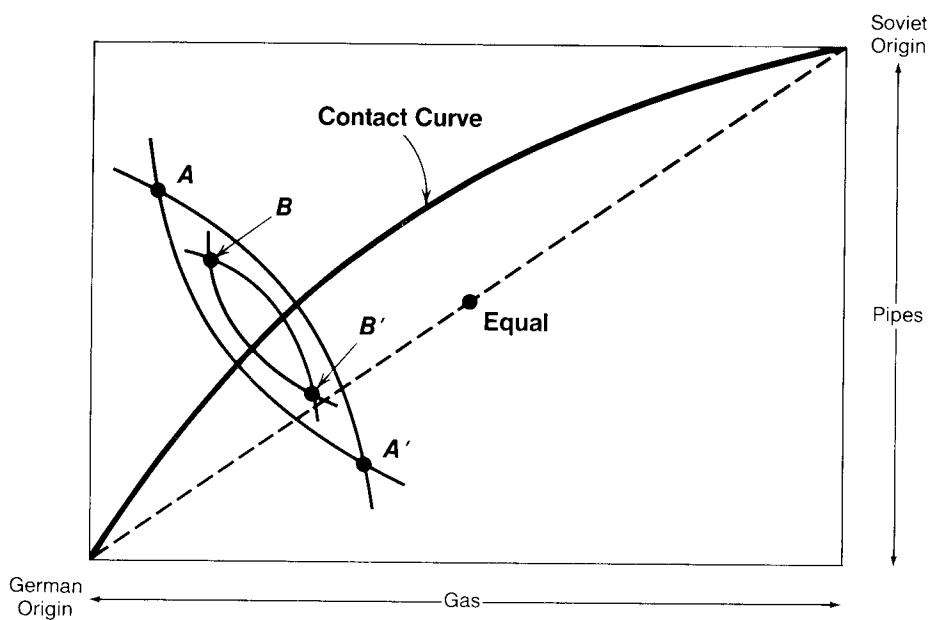


Figure A5.1
The Gains from Trade

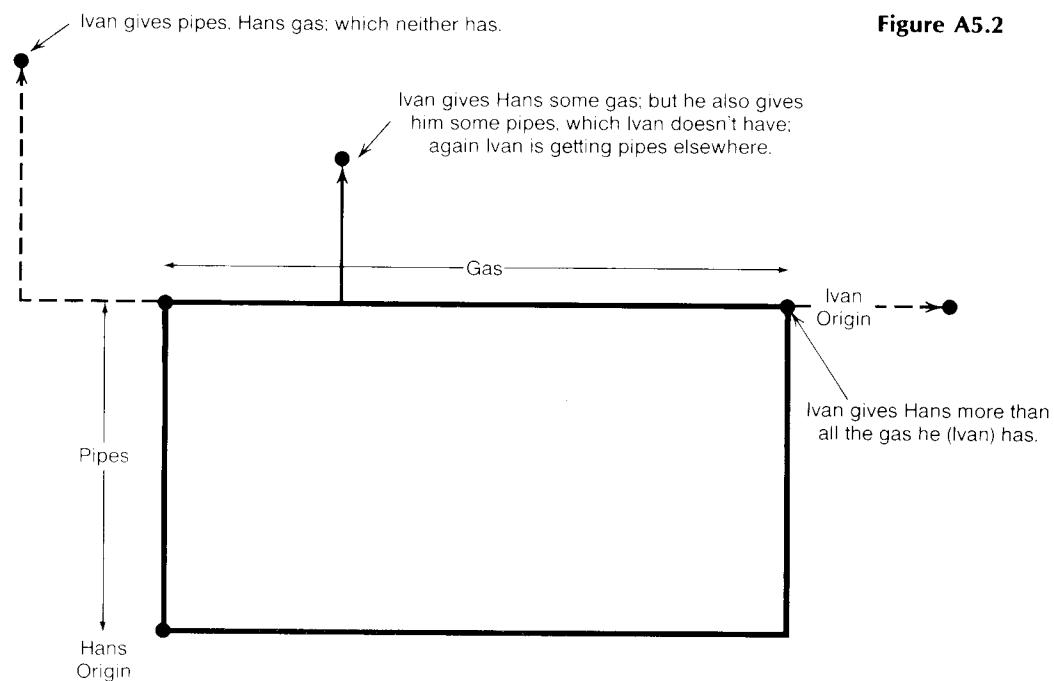


Figure A5.2

You do not face ten or a hundred alternative sources of food; and the food-bearing stranger does not face alternative sources of water. You must bargain.

3. True. One hundred thousandth of the supply is no big bargaining chip and gives no power to alter the price; but 60% is and does.

5. No. The price is no “gouge,” merely the result of supply and demand. It may be high, but it’s not unnatural.

Problems for Section 5.3

1. True. If Goldin wins or loses by more than one vote your trip to the polls has been pointless. Only in the very unlikely case of an exact tie does your vote matter. Therefore, the average advantage you expect to get from the trip is $(1/2000)(\$100) = \0.05 . If it costs more than this the game is not worth the candle.

COMMENT

You may be tempted to argue that one vote combined with the votes of other like-thinking people does influence the election. Therefore one vote does count. Resist the temptation, for the argument is irrelevant. One man, one toilet. The rueful thought common after a close election that a single vote *did* count, because “one vote per precinct” would have put Mr. Snerd in the Senate, entails the same fallacy. One voter deciding whether or not to go to the polls for Snerd does not in fact influence the others. The point is that in a crowded election or market one man’s vote or dollar is a substitute for every other’s and is not crucial.

Exercises for Section 5.4

1. False. No one can be made better off *without making someone else worse off*, that is, without taking something from someone else. Farmers can be made better off, but only by hurting consumers, and vice versa. At Equilibrium there are no more *mutually* advantageous deals to be made.

3. b. Send a check to owners of present allotments (as they are called), but then permit anyone to grow tobacco. The present owners would be just as well off, but consumers of tobacco better off (at least in the price they paid, if not in their circulatory system).

c. Send a check to American shipowners and seamen, then let the freight market rip. Same point.

d. Send a check to American carmakers, then let the market rip. Same point. And so on, through the thousands of volumes of federal regulations and taxes and restrictions that protect this or that interest.

Problems for Section 5.4

- 2. True. The operative words are “force” and “entice.” From the point of view of the individual citizen the taxation is hurtful but the loan is not. The government entices him into making the loan, that is, he makes it voluntarily and in his own judgment is made better off. He apparently feels that giving up some resources today in exchange for more resources tomorrow is desirable for him. But the government forces him to pay taxes, that is, he pays taxes involuntarily, and is made worse off. Taxes today to pay for today’s war, therefore, put the burden (the feeling of being made worse off) on today’s generation; borrowing today puts it on tomorrow’s generation, that is, on the generation that is taxed (to repay the loan). The question of the burden of government debt is among the oldest unsettled controversies in economics. The argument given here is due to James Buchanan, *Public Principles of Public Debt* (Homewood, Ill.: Irwin, 1958). It is controversial. A lucid summary of the state of play by 1962 is given by Carl S. Shoup, “Comment on the Burden of the Debt and Future Generations,” *Economic Journal* 72 (1962): 887–98, reprinted (with other articles) in R. W. Houghton, ed. *Public Finance: Selected Readings* (Harmondsworth, England: Penguin Books, 1970). Since then the controversy has continued. See Robert Barro, Martin Feldstein, and James Buchanan, *Journal of Political Economy* 84 (April 1976): 331–49 and works cited there.

CHAPTER 6

Exercises for Section 6.1

1. Clearly, 70,000 tickets is many more than you demand at any positive price. Since other people are willing to pay the price, you would sell to them. You would be a supplier, keeping in the end a few for your own use.

3. No. It produces less, so its excess supply has been changed into an excess demand.

5. Stealing or threats of violence hurts the victims. Standing in a long line wastes resources (time) in the very process of competing (stealing may, too: the cost of the gun, say, or the time spent sneaking around finding victims). Fraud (pretending to be a student) can be viewed as theft of a reputation, imposing on real students the cost of checking IDs to offset the fraud. Charming someone has no such costs unless one must spend resources to do it (wasting time being nice to someone you wouldn't want to be nice to anyway). Bribery has none at all.

7. The participants think it's swell, or else they wouldn't participate. The government officials think it's bad because it lets meat go to whoever values it most instead of to the people the government wants to favor.

9. The scalper doesn't look at bank accounts and award the tickets to the people with the biggest bank accounts. He couldn't care less what your bank account is: he cares only about what you offer. By contrast, in the beauty contest the "bank account" of beauty *does* matter. In fact, it's all that matters. The most beautiful person wins everything, all the time. Hardly seems fair. In any event, the nonbeautiful person never wins anything.

Exercises for Section 6.2

1. The question is answered by figuring out how each change affects the market supply and demand curves.

a. One less room, so quantity *falls* to 3 and price must *rise* to clear the market.

b. Two more rooms, so quantity *rises* to 6 and price must *fall* to clear the market.

c. No change in the number of rooms supplied, just in who owns them. Nothing happens to price and quantity.

d. At any given price Sutch demands more—exactly one more, in fact, because the intercept rises from 2.5 to 3.5. So the market demand increases and the price *rises*. The quantity doesn't rise, of course: it's stuck at 4 by the assumption that supply is inelastic.

e. At any given price Ransom demands less, because the coefficient on P is *more negative*: -0.003 instead of -0.002 . The price more strongly reduces his demand, so to speak. So the price falls.

Problems for Section 6.2

1. False. Read "exorbitant" as "higher than I think it should be." An effective boycott of meat will indeed bring its price down, but only by pushing back the demand curve. "Resuming their higher purchases" will push the curve back to its former level, and the price back up.

4. True under some circumstances, false under others. The problem illustrates the uses and the limitations of the law of one price. The wheat market is worldwide. A ban on all American exports (not only exports to South Africa) would reduce world supply outside the United States, raise the price, and cause more people in Zaire to starve. But the ban contemplated is in fact only on exports to South Africa. To a first approximation the consequent reshuffling of wheat has no effect: American wheat that went to South Africa goes to, say, Russia instead. Argentina, which formerly sold to Russia, now sells to South Africa. The reshuffling does not change the world price and therefore does not hurt or help Zaire. To a second approximation it has some small effect, because the former pattern of grain flows presumably had some reason for being (low transport costs, for example); the reshuffling distorts the flow and raises its price a little.

Exercises for Section 6.3

1. No, it doesn't necessarily. British exports were large because at the prevailing *world* price the British makers of wool cloth could do well supplying a lot. The question, like the Indian cotton problem, is what happens to world demand and supply. In fact, had America been outside the British empire (you'll recall that there was a little excitement about the matter in 1776), the demand of Americans for cloth would have been about the same.

3. True. Since General Motors and many of the other companies cannot reasonably take the price of autos as given to them and unalterable by each company's individual actions, there's no sense to "adding up each company's supply curve to make a market supply curve."

Problems for Section 6.3

1. Let the low-price market—the exporter, of course, because the transport cost is incurred moving televisions from the exporting to the importing country—slide up along the common price axis by the amount of the price differential, as in Figure A6.1. The shared point P_E on the price axis can in this way represent two prices, the higher American price and the low Japanese price (P_E being further from the zero price in America by the amount of the transport costs).

3. True. If the law of one price holds literally for autos then a 6% import of autos will have the same effect in bringing the domestic price up (or down) to the world price as does a 26% import. Domestic and imported autos must sell at the same price if they are perfectly fungible (as is supposed in the law of one price). So too must domestic and imported clothing, television sets, toys, steel plates, shoes, and so forth sell at the same price. The quantitative importance of the imports does not matter. Or, more accurately, it does not matter in proportion to the ratio of imports to domestic production: Britain is not more than four times more open to world prices than is America (notwithstanding that $0.26/0.06 = 4.3$).

7. False. In fact the proof would be strengthened. The proof as given in the text acknowledges only other bread demanders as a source of supply from a higher price; if it acknowledged that at a higher price bread suppliers would wish to produce more bread, their additional production would have to be added to the 20,000,000 additional loaves available to Lindert when he pushes up the market price.

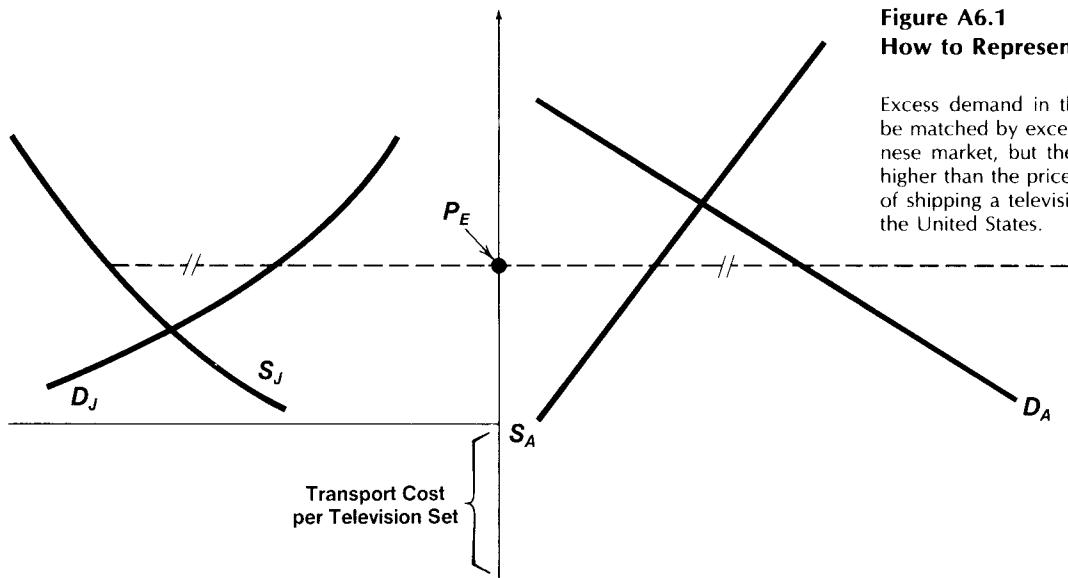


Figure A6.1
How to Represent Transport Costs

Excess demand in the U.S. market must be matched by excess supply in the Japanese market, but the U.S. price must be higher than the price in Japan by the cost of shipping a television set from Japan to the United States.

Exercises for Section 6.4

1. a. The outward movement of cotton supply causes, of course, its price to fall. In the market for wool, therefore, the *demand curve* for wool falls: cotton substitutes, since it's now cheaper. But the wool supply is flat: therefore, the *price* of wool doesn't change. So back in the cotton market the demand curve, which depends on the price of wool, also doesn't change. The feedback link is broken.

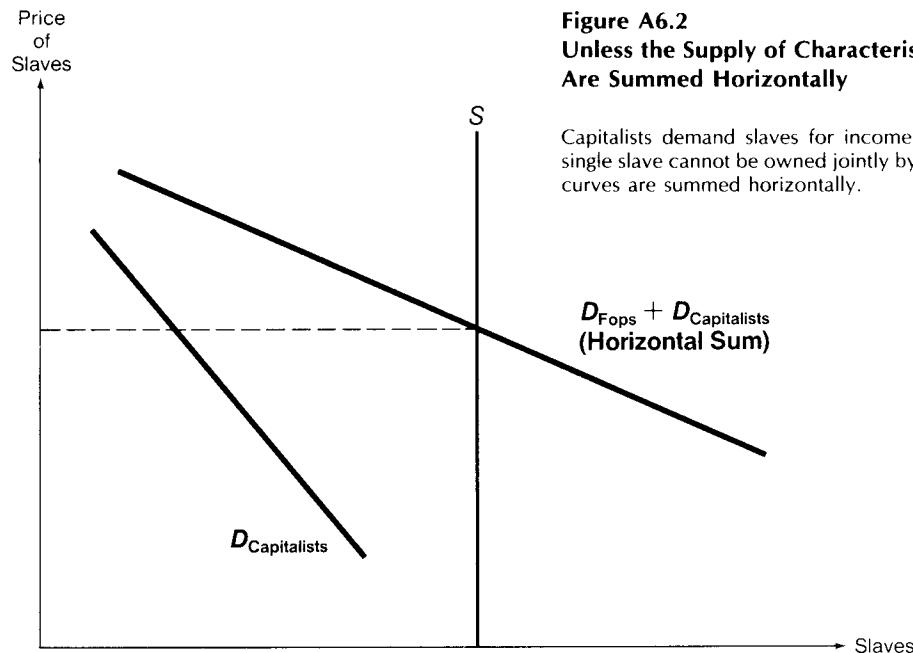
b. The story is the same as in the text. The only difference is that there's not even a slight fall in the quantity of wool exchanged: the vertical supply sees to that, since at any price the same quantity is supplied.

c. Again, the link is broken. The rise in the supply curve of cotton, and the drop in its price, now does *not* cause the demand for wool to fall. Nothing happens in the wool market. Therefore, the demand curve for cotton also stays put: there's no change in the price of a substitute to shift it.

3. 1 with E: As in text.
- 2 with F: Both are high school courses, jointly supplied.
- 3 with G: Both are supplied by the police. The whole value of the police is the sum.
- 4 with A: Big cars come with both.
- 5 with H: Both in this book.
- 6 with J: Both from roads.
- 7 with C: Both from American Foreign Policy.
- 8 with D: Both from some foods.
- 9 with B: Straw may be called a "by-product."
- 10 with I: Both from the newspaper.

Problems for Section 6.4

1. Since it is no longer the same slave that is providing income from business and pleasure from prestige, the demands of the two classes are summed horizontally, not vertically, as in Figure A6.2. If you took measurements of the value of slaves in money to Capitalists you would find the money value to be equal to the price—in contrast with what you would find if (as in the text example)



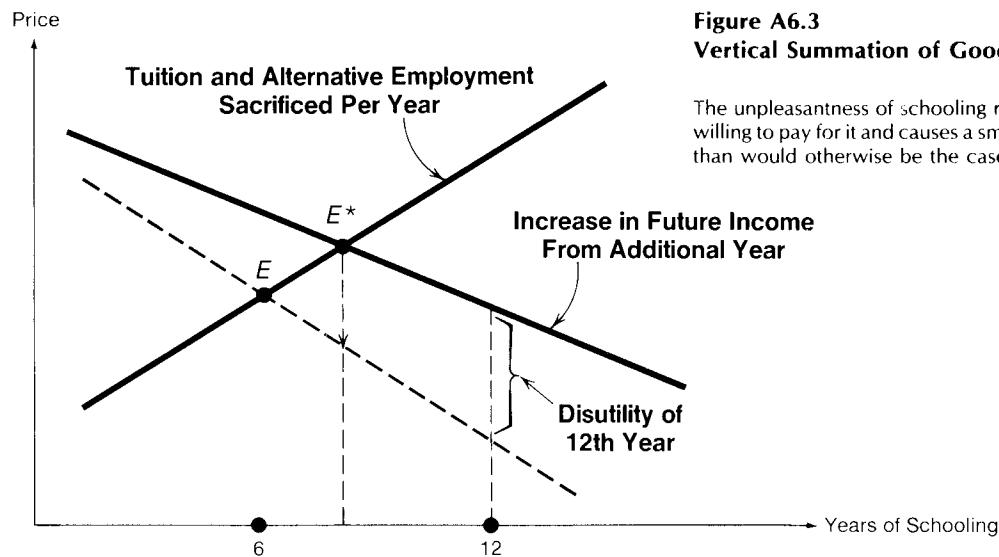


Figure A6.3
Vertical Summation of Goods and Bads

The unpleasantness of schooling reduces the price students are willing to pay for it and causes a smaller quantity to be consumed than would otherwise be the case.

Capitalists were somewhat foppish, namely, the money value lower than the price. Were Capitalists and Fops specialized in this way, therefore, a finding (on capitalist, record-keeping plantations) of no premium for prestige would shed no light on the prevalence of foppishness: Fops could be holding half the slaves, supporting its price far above what it would be had foppish motives not existed, yet would leave no trace in the records.

3. False. The assertion goes wrong at "and so on," with the implication that the prices fall indefinitely. The new solution need not be achieved by these steps, or, for that matter, by any steps. And the steps, should they occur, do not go on indefinitely: they get smaller as they get closer to their goal.

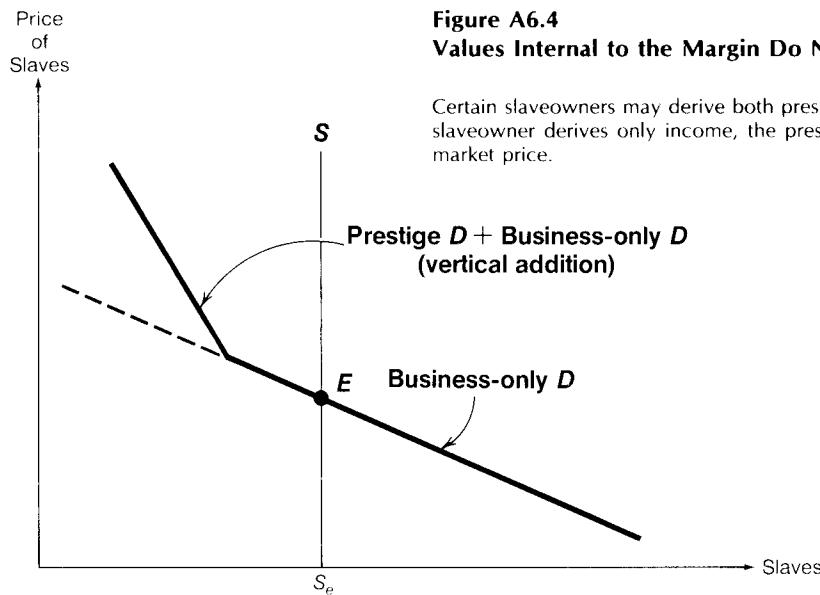


Figure A6.4
Values Internal to the Margin Do Not Affect Price

Certain slaveowners may derive both prestige and income from slaves, but if the *marginal* slaveowner derives only income, the prestige value of slaves will not be reflected in the market price.

COMMENT The idea of the “vicious circle” (and the tired witticism that travels with it, the “virtuous circle”) is common in ersatz economics and is not unknown in real economics. The “inflationary spiral” is a typical example, as are at a more sophisticated level certain theories of investment and economic growth. Many depend on an alleged inability of the market actors to anticipate equilibrium; the less sophisticated versions abandon the idea of equilibrium altogether. Compare Section 13.2.

7. True. It is obvious, really: by the meaning of the words, unpleasantness reduces the attractiveness of an activity. Think of the year as an investment good (increasing future income more than it costs) and simultaneously a consumption good, or, rather, a consumption bad (excruciatingly unpleasant). So the “demand curve” for additional years of schooling is the investment demand *minus* the negative consumption demand that the student has for the year. This demand curve is set against the “supply curve” (the costs of tuition and forgone employment) to determine the amount of schooling purchased, as in Figure A6.3. The actual equilibrium is E , below the E^* that would be observed if schooling were not unpleasant (this latter E^* being the correct amount of schooling “from the point of view of maximum money income” alone). See E. Lazear, “Education: Production or Consumption,” *Journal of Political Economy* 85 (June 1977): 569–597.
8. True. Look at Figure A6.4. The *marginal* prestige value of an additional slave is at equilibrium zero. Mathematically speaking, prestige and profit are in general both functions of the number of slaves, but it happens that at the existing number of slaves, S_e , the marginal utility of slaves in the prestige function is zero.

CHAPTER 7

Exercises for Section 7.1

1. In each case the income goes up 1%. At \$10,000 the wine consumed goes up more than 1% (by 2%), so the income elasticity is greater than 1.0 (namely, 2). At \$20,000 it goes up exactly 1%, or 1.0. At \$30,000 it goes up by more than 1% (by one-fifth of 1%), so the elasticity is less than 1.0.
3. True, as was said in the text. The reason is that along any such “ray” the magnitudes measured on the two axes move up in the same proportion. This is no matter what the slope is.
5. Not like crying. The case is the opposite of Exercise 4: their total revenue goes up, because the price will rise proportionately more than the quantity falls.

Problems for Section 7.1

1. True. People will spend all their grant, whatever the price. Why? Well, why not? After doing important copying they will, if they have any of the grant left, start copying shopping lists, newspaper ads, and any mildly interesting book. The total revenue, PQ , is therefore constant and equal to the amount granted. Therefore the price elasticity is 1.0: a doubling of the price per page would exactly halve the number of pages copied.

Exercises for Section 7.2

1. The 10% fall in P_A causes a rise in quantity of $(-200)(-10\%) = +20\%$. The 1% fall in the price of houses causes a fall of $(+1.5)(-1\%) = -1.5\%$. The 5% rise in income causes a rise of $(+0.5)(+5\%) = 2.5\%$. If they all happen together they cause $+20\% - 1.5\% + 2.5\%$, or a 21.5% rise in the quantity demanded. You merely add up the separate effects.
3. It's got to be somewhere between the two, 0.25 and 0.75. Where? Use the formula in the text developed in the paragraph beginning “The last fact. . . .” With 50-50 weights the answer is 0.50.

Problems for Section 7.2

3. a. $Q = D_u P_u^{-\eta_u} + Q_r P_r^{-\eta_r}$ is the equilibrium condition, and since there is one market in slaves there is only one price ($P_u = P_r$).

b. Evidently, there were better substitutes for slave labor in the cities than in the countryside; that is, η_u was greater than η_r .

c. The simplest way to see this is to suppose for a moment that the demand curve in city and country were stable. The same rise in price faced by both city and country would evidently cause the city quantity demanded to fall off more sharply, because that is what one means by saying that the elasticity of demand was higher in the city. That is, the city quantity demanded of slaves is more easily squeezed by the higher price than the rural quantity demanded. Likewise, a fall in the slave price would cause city demand to rebound more violently than rural demand.

d. Notice that when prices are falling (or rising very slowly, as in 1820–1830) the slave population of St. Louis is rising sharply. So the logic of (c) can be applied: the oscillation in the quantity of slaves in St. Louis can be explained by an oscillation in prices (caused by a nationwide oscillation in the rate of growth of the slave population) interacting with high price elasticities of demand in St. Louis and low elasticities in the countryside. As Claudia Goldin argued, the rates of shift of the city demand curve reflecting the allegedly hostile environment for slavery in the cities need have nothing to do with the story.

6. True. Start with the equilibrium condition: $Q_{\text{Brazil}}^D = Q_{\text{world}}^D - Q_{\text{others}}^S$. Leap to the elasticity equation (measuring elasticities of demand, η^D , positive):

$$\eta_{\text{Brazil}} = \left[\frac{Q_{\text{world}}^D}{Q_B^D} \right] \eta_{\text{world}} + \left[\frac{Q_{\text{others}}^S}{Q_B^D} \right] \epsilon_{\text{others}}$$

Then insert the information given: $\eta_B = (3/1)(0.1) + (2/1)(\epsilon_0)$. The one remaining unknown term, the elasticity of supply, ϵ_0 , must be at least zero, meaning that the elasticity of demand facing Brazil must be at least $(3/1)(0.1) = 0.3$, as asserted.

9. This apparently impossible problem is worked out as follows. A rise in the price of transport (P_t) will raise the price of steel in the provinces (call that price P): in fact, if transport costs are 50% of the price of steel in the provinces, clearly a 1% rise in the price of transport will increase the price of steel by $\frac{1}{2}\%$; and in general the rise in the price of steel in the provinces from each 1% rise in the price of transport will be 1% multiplied by the share of railway services in the final price of steel. Now this increase in the price of steel will reduce the quantity demanded of steel, by an amount depending on the elasticity of demand. If the elasticity of demand for steel were, say, 3, the $\frac{1}{2}\%$ increase in the price of steel derived in the particular example given above would result in a $1\frac{1}{2}\%$ fall in the quantity of steel demanded. Now steel and railway services are demanded in fixed proportions. So the $1\frac{1}{2}\%$ fall in the quantity demanded of steel resulting from a 1% increase in the cost of transport makes the quantity demanded of transport services fall by $1\frac{1}{2}\%$ as well. So the elasticity of demand for transport services is the percentage fall in the quantity of transport services divided by the percentage rise in the price of transport, that is, $(1\frac{1}{2}\%)/(1\%) = 1\frac{1}{2}$; and in general it is the elasticity of demand for steel multiplied by the share of transport costs in the final market price of steel.

Mathematically: $P = P_0 + (P_t)(Q_t)/Q$. That is, the price of steel in the provinces is the price in Pittsburgh plus the average cost of transporting steel to the provinces, which is, in turn, the price of transport multiplied by the quantity, used, divided by the quantity of steel shipped. Signifying rates of change with an asterisk and applying the rate-of-change formulas:

$$P^* = \frac{P_0}{P} P_0^* + \frac{P_t Q_t / Q}{P} \left(P_t^* + \frac{Q_t^* - Q^*}{Q} \right)$$

0 by the assumption 0 by the assumption that Q and Q_t
of elastic supply change in fixed proportions
at Pittsburgh

So

$$P^* = \left[\frac{P_t Q_t}{PQ} \right] P_t^*$$

↓
share of transport in value of steel in the provinces (call it s_t)

Now

$$Q_t^* = Q^* = \eta_{Q_P} P^* = (\eta_{Q_P} s_t) P_t^*$$

So

$$\frac{Q_t^*}{P_t^*} = \eta_{Q_P} s_t = (\eta_{Q_P})(s_t), \text{ which was to be shown.}$$

COMMENT

The result is a reason for believing that the price elasticity of demand for transportation (or mail or electricity or other products used to make other products that have few good substitutes and are small percentages of the total costs of the products they make) is lower than that of products with good substitutes, products important in the costs of the products in which they are used, or products consumed directly by final consumers.

CHAPTER 8

Exercises for Section 8.1

1. a. Trevor will specialize in body fixing, Thomas in auto fixing. If Trevor fixed his own auto he would have less time to attend to bodies. An hour spent by him fixing his auto would have a higher cost in body fixing given up than would an hour spent by Thomas. So more is going to get done—more auto and body fixing in total—if each specializes to some degree in what he does best.
- b. With a limited team—25 people, say—an intelligent manager will obviously have Clark specialize in pitching and Robert in hitting. The combination will produce more victories for the team.
- c. Here Gallman has an *absolute* advantage in everything. But still he should probably be put in right field (or maybe now left field, with his new arm). The team may give up less by assigning Gallman to fielding (where he can play every day) than to pitching. A historical case in point is Babe Ruth, who pitched very well for the Red Sox. He was bought by the Yankees and put in the outfield, so that he could play (and bat) every day.
- d. There's nothing inevitable or biological about it, but it's a good bet that Mrs. Hogendorn has a comparative advantage in curtain making, even if she is also a whiz at chair painting. Mr. Hogendorn may be lousy at both, but in a world of scarce time he is better employed at botching the chair painting than really botching the curtains.
- e. That the United States can do both better is irrelevant. What matters is comparative advantage. More of both get made if the United States specializes in soybeans.

Problems for Section 8.1

2. Look at the construction of the world's production possibility curve in Figure A8.1. The point $E + E'$ is on the curve. If no one has a comparative advantage, in other words, nothing is served by specialization and trade.
3. a. The maximum possible amounts given in the table are clearly the intercepts of the (straight-line) production possibility curves on the Food and the Machine axes. When Japan, for example, gets 20 million Machines it must sacrifice 6 million tons of Food. The price of a Machine in Japan, therefore, is $6/20 = 0.3$ tons of Food per Machine. In like fashion, the price of a Machine in Britain is $15/20 = 0.75$ and in America $30/30 = 1.0$. The low-cost producer of Machines is Japan, even though America is more productive per unit of resources in producing machines: so unproductive is Japanese compared with American farming that building a Machine in Japan sacrifices less of the world's Food.
- b. Following the instructions, each country's consumption point is the point along its production possibility curve at which tons of Food and numbers of Machines are equal. Because you know the intercepts from the table, you can use the prices derived in (a) (which are slopes) to write down without calculation the three production possibility curves (use the units in which things are measured to guide the writing):

$$\begin{aligned} \text{Japan: } F_J &= 6 - 0.3M_J \\ \text{Britain: } F_B &= 15 - 0.75M_B \\ \text{America: } F_A &= 30 - 1.0M_A \end{aligned}$$

The points of $F_J = M_J$ and so forth are

$$F_J = M_J = \frac{6}{1.3} = 4.6$$

$$F_B = M_B = \frac{15}{1.75} = 8.6$$

$$F_A = M_A = \frac{30}{2.0} = 15.0$$

The world consumes $4.6 + 8.6 + 15.0 = 28.2$ million tons of Food and 28.2 million Machines.

c. Start with the whole world producing only Food, that is, $6 + 15 + 30 = 51$ million tons. This is the Food intercept. Since Japan is the least – cost producer of Machines, if some are to be produced Japan will do the producing; if more are to be produced Britain will also; and only if still more are to be produced will America. Successively adding each country's line in approximately the correct scale yields Figure A8.2. From the diagram it is clear that Britain will be producing both Food and Machines (look at where the 45° line intersects). Japan will specialize in Machinery, America in Food.

d. Taking the broad hint, the Equilibrium is the solution of $F = M$ and $F = [(20)(0.75) + 45] = 0.75M$; or $F = 61/1.75 = 34.9$ million tons and 34.9 million machines. Consumption has increased by a factor of $34.7/28.2 = 1.23$ by the opening of trade and the consequent specialization. Japan has gotten out of producing what it is comparatively bad at producing, Food; and America out of Machines.

6. False; in some economic respects, but not all. The trips by train, car, bus, and plane have opportunity costs. Society must sacrifice the output of shoes, houses, and so forth that the engineers, stewardesses, autoworkers, and so forth could produce were the trips not demanded. The tax, by contrast, corresponds

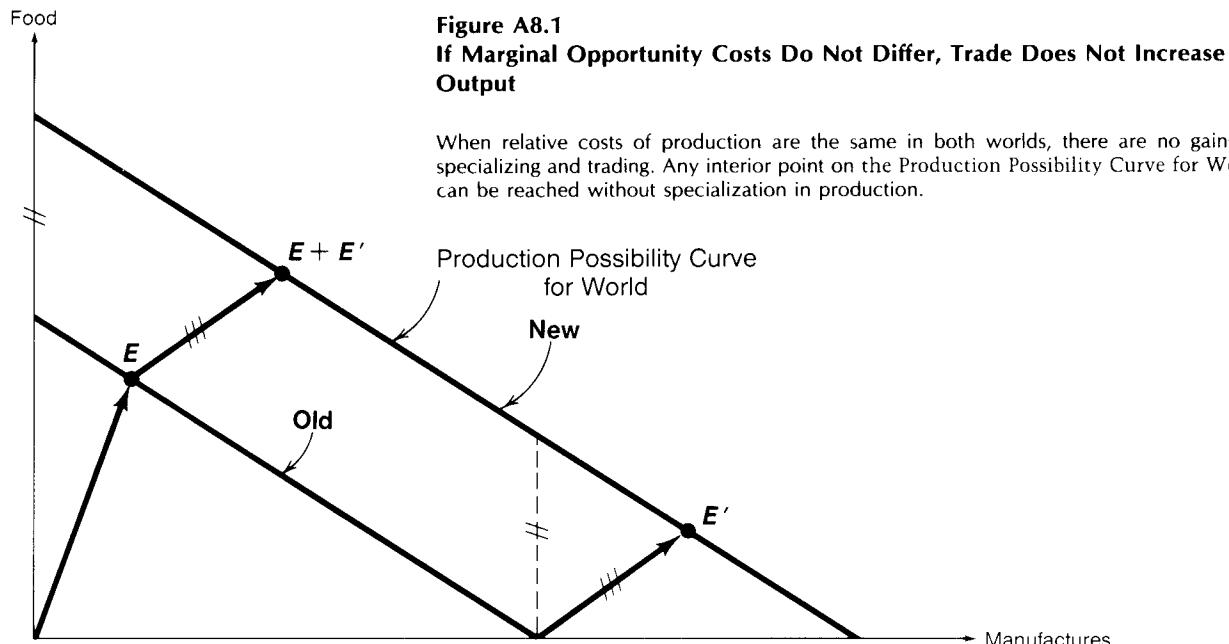
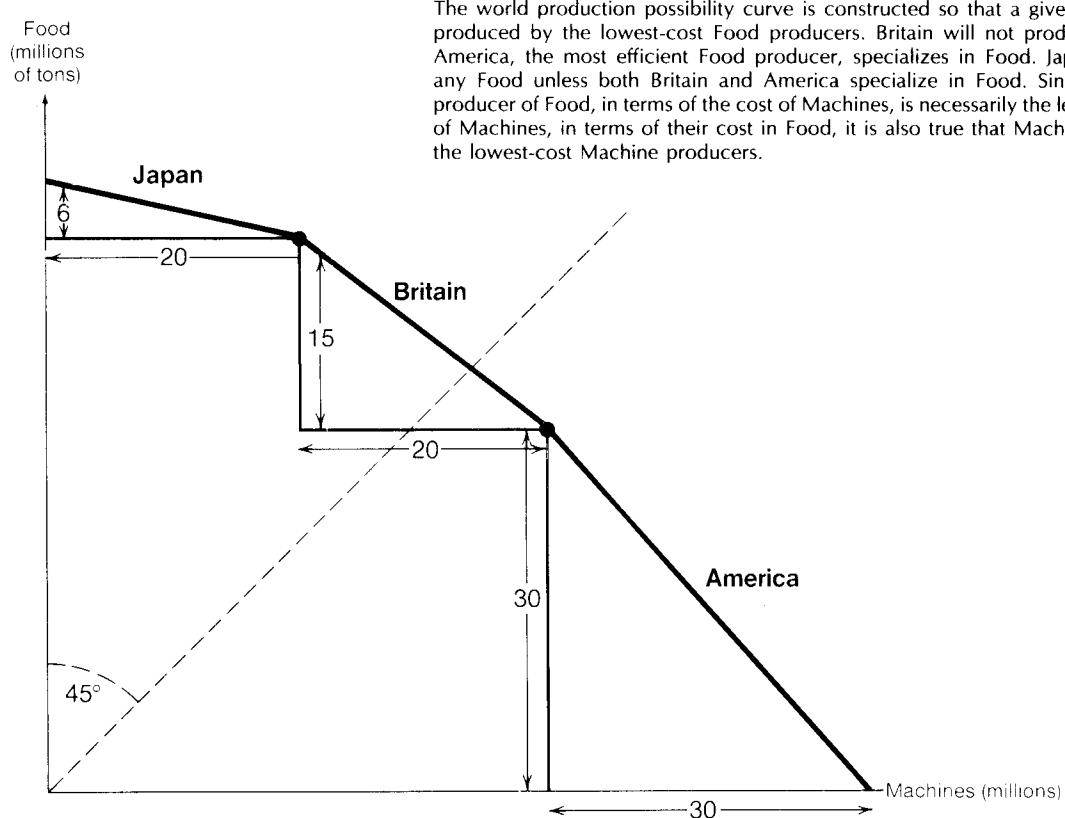


Figure A8.2
Production Possibility Curves Are Drawn by Adding Successively the Lowest-to the Highest-Cost Producers



The world production possibility curve is constructed so that a given quantity of Food is produced by the lowest-cost Food producers. Britain will not produce any Food unless America, the most efficient Food producer, specializes in Food. Japan will not produce any Food unless both Britain and America specialize in Food. Since the most efficient producer of Food, in terms of the cost of Machines, is necessarily the least efficient producer of Machines, in terms of their cost in Food, it is also true that Machines are produced by the lowest-cost Machine producers.

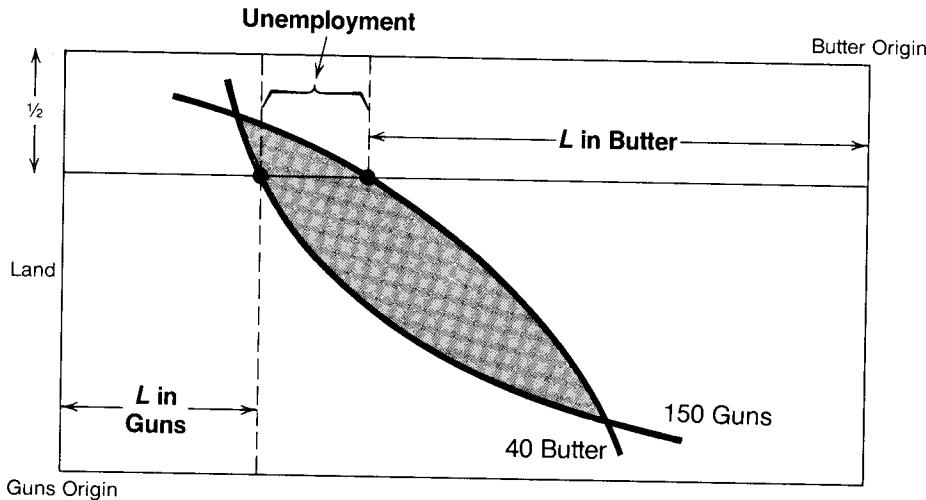
to no opportunity costs. The traveler is indifferent between paying his \$100 to Eastern Airlines or to the Office of Taxed Movement. But the rest of the society is not indifferent between putting transport workers to work at Eastern Airlines or leaving them free to work at making shoes, houses, and other things. Economists arguing, as they often do, that transport or pollution or sharecropping is from the social point of view "just like a tax" commit this fallacy of misplaced opportunity costs.

Exercises for Section 8.2

1. a. Forests, lumberjacks, caterpillar tractors, saws, sawmill operators, financing of inventories, transportation.
- b. Paper, ink, presses, advertisers, draftspeople, artists, presspeople, copy editors, acquisition editors, editors, more editors, and (incidentally) authors. (Note the people-and-skill-intensive nature of the product.)
- c. Postpeople, post offices, land under post offices, roads, automatic mail sorters and other transport media, mailboxes, addresses.
- d. Police, guns, cars, stations, jails, courts, trust of citizens. (Note the last.)
- e. Lumber, nails, appliances, plumbers, carpenters, electricians, land, cement, bricks, glass, finance from bank. (Note the last, and that lumber goes into houses.)

Figure A8.3
Unemployment Leaves a Gap of Unexploited Opportunities for Exchange

By employing all the labor that is available, the economy could move to a higher Gun and a higher Butter isoquant. Thus unemployment of labor is inefficient.



- f. New houses (which become old once built), maintenance services, furniture, replacement of appliances. (Note the new houses which use lumber which use Caterpillar tractors which use steel which uses labor which use old houses which use new houses: the circular flow of economic life.)

Problems for Section 8.2

3. The three shapes in Problem 2 all imply that only one of the inputs would be used. That is, choosing the high contour subject to a straight-line budget constraint would lead to a corner solution. But the purpose of a production function is to exhibit recipes in which more than one input is used—bread made with flour and water, not flour alone. The point is analogous to that of utility contours, which in any of the shapes in Problem 2 imply specialization in consumption, that is, man living by bread alone.

5. The 50–50 division of Land, fully employed, means that the economy will be somewhere on the light horizontal line in Figure A8.3. Since Labor is not fully employed, however, the economy's position is not one point but two: one for the Labor employed to make (as drawn) 150 million guns and another for the Labor employed to make 40 billion pounds of butter, with a gap equal to the unemployed Labor. The two points are not unique, because the same amount of unemployment could be represented by any gap of the same size between any two isoquants. More output can be attained by moving to a point inside the shaded lens shape. Unemployment, then, implies that the economy is inside its production possibility curve, that is, that more of both Guns and Butter could be attained were the unemployment eliminated (note the coordinates of the Unemployment Point in Figure A8.4). The dashed curve gives all the other outputs attainable with this degree of unemployment. In such a world there is a free lunch—the free lunch from setting people to work and moving out to the production possibility curve.

6. The production contours have corners at the required proportions. Set up the Edgeworth box as in Figure A8.5. The mysterious-sounding conditions in the problem ("the proportions being different . . . and lying on opposite sides . . . in the nation as a whole") merely prevent the diagram from reducing to a special case (such as a case of no point of intersection between the dashed lines out of the corners). Only at the point Full (which corresponds to a kink in the production possibility

Millions
of Guns

Figure A8.4
Unemployment Puts the Economy Inside the Production Possibilities Curve

Unemployment is inefficient; inefficiency in production corresponds to a point inside the True Production Possibility Curve.

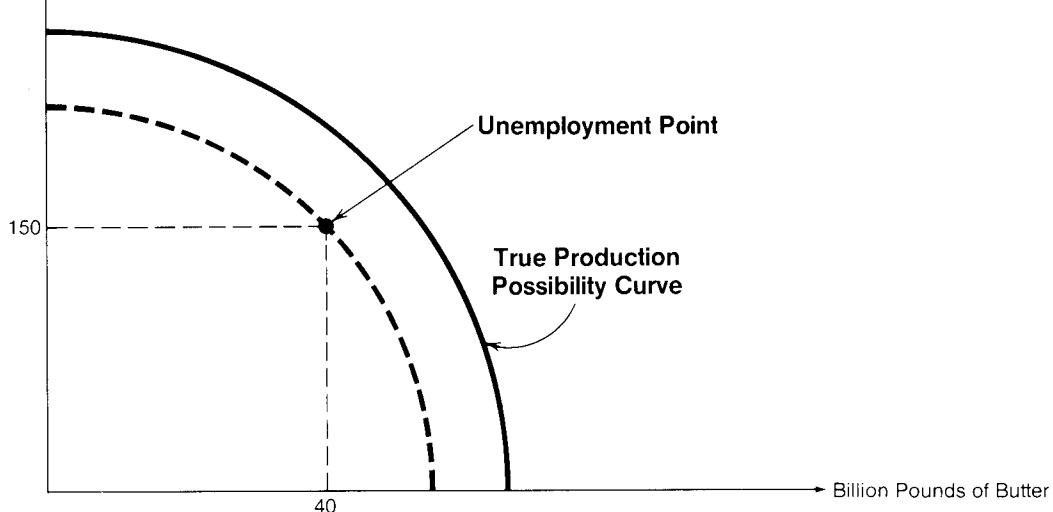
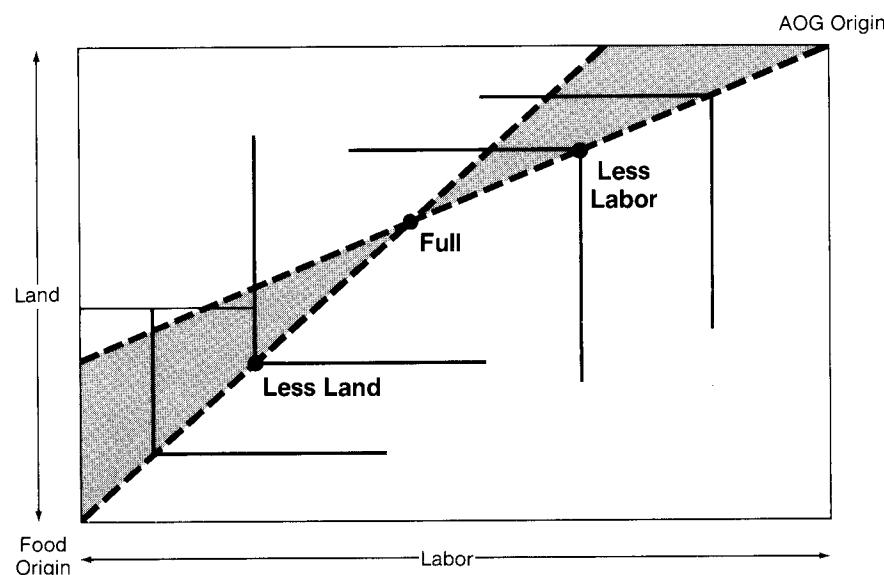


Figure A8.5
The Factor Proportions Problem

At point Less Labor, some labor is unemployed; at point Less Land, some land is unemployed. The point Full, and only that point, is consistent with full employment of both factors.



curve) is there full employment. At other points, such as Less Labor, there is less than full employment. The point Less Labor is efficient, as is any point in or on the shaded region, because at all such points two isoquants are just touching (in fact, overlapping). But at Less Labor not all the economy's labor is required to produce the Food and All Other Goods actually produced. Labor is unemployed, or, at best, uselessly employed. On the other side of Full it is land that is unemployed.

Notice that the economy in Figure A8.5 resembles a typical poor country, such as India: labor is abundant relative to land (the longer is the horizontal, labor axis the larger is the upper, labor-excess portion of the contract region); Food is relatively land-using (All Other Goods—such as postal service or light bulbs—use a lower ratio of land to labor than does Food); the social prestige of land-holding and the high demand for Food relative to All Other Goods in a poor country keeps the land in agriculture (and therefore keeps the possible equilibria in the labor-excess portion above Full in the diagram). If the techniques of producing Food and All Other Goods were in fact fixed in their input proportions, the argument would explain the apparent fact of high unemployment (or "disguised unemployment," that is, putting people to work at useless tasks) in poor countries. See Richard S. Eckaus, "The Factor Proportions Problem in Underdeveloped Areas," *American Economic Review* 45 (September 1955): 539–565. There is disagreement about whether the fact is a fact.

Exercises for Section 8.3.

1. There's no "answer" that wouldn't just repeat the book. The answer goes on in your brain, as you think through the argument with another example concretely in mind.

Problems for Section 8.3

1. The result depends on a shift in the economy's production possibility curve between Children and All Other Goods similar to the shift discussed in the text, namely, a shift (due to technological change) that raises the relative price of children at every quantity produced, as in Figure A8.6. A

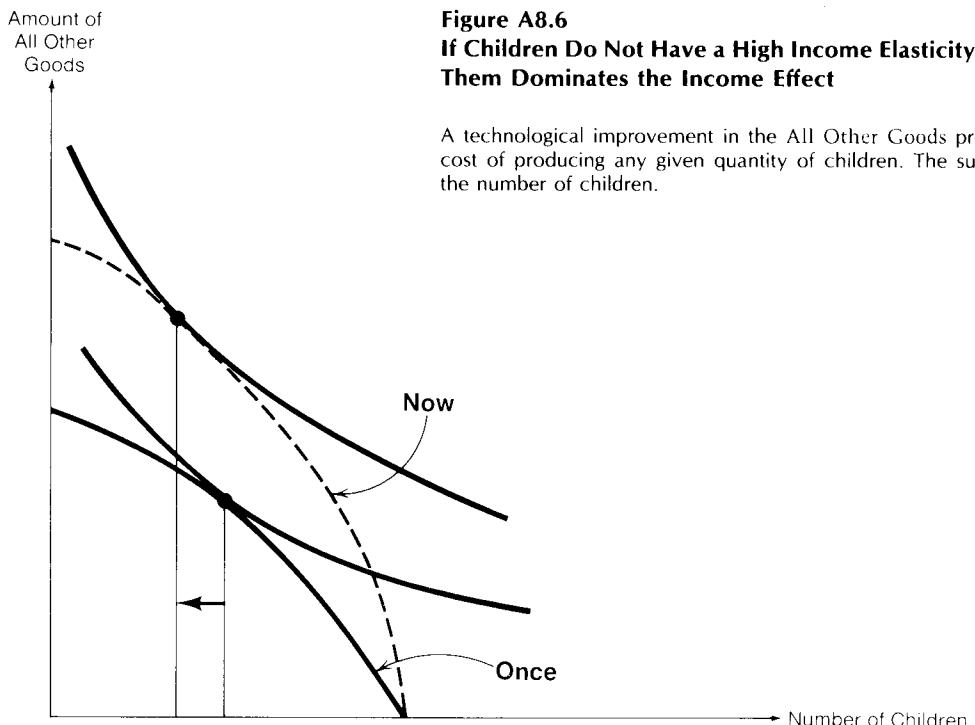


Figure A8.6
If Children Do Not Have a High Income Elasticity, the Rising Price of Raising Them Dominates the Income Effect

A technological improvement in the All Other Goods production increases the marginal cost of producing any given quantity of children. The substitution effect causes a fall in the number of children.

more pessimistic scenario would be that the income elasticity of demand for children is very high, in which case each enrichment through technological change is offset by more hands pressing against resources. The optimistic scenario (that is, low income elasticity for sheer numbers of children) appears to fit experience in industrial countries over the past century better.

3. False, it is less likely. Openness to a world market with a given price of manufacturers relative to agriculture would insulate the decision about what to produce from the decision about what to consume. The positive income effect on the demand for manufactures of the good harvest would not affect production. Production of manufactures would depend entirely on the price effect, which is negative: British manufactures were made relatively more expensive by the cheapening of agriculture, and in the face of a given world price of manufactures relative to agriculture, British manufactures would have to contract to preserve the optimum composition of production.

CHAPTER 9

Exercises for Section 9.1

1. a. Normative, because entirely about what should be.
 - b. Positive, because entirely about what is.
 - c. Positive.
 - d. Mixed, with heavy doses of normative. For instance, what's good depends on who is made better off. Deciding that making a person better off is "good" is normative.
 - e. Positive, largely, although the exact meaning of "applies" will depend on artistic grounds: when is a statistical test, say, "sufficiently strong"?
 - f. Normative: matters of beauty, as of morality, are normative.
 - g. Mixed: both artistic and factual matters are involved. When is an argument persuasive? (Or true?) The answer depends on artistic judgments.
3. False. They want the income to be achieved by free exchange. This may or may not get the highest possible income (though in such discussions it is often assumed for the sake of argument that it does).

Problems for Section 9.1

1. To take away some Rockefeller wealth in the name of justice is to apply an "end-state" criterion of justice. It is by no means obvious that the end-state criterion is superior to the "entitlement" criterion, according to which a just transfer of wealth earned justly is itself just.

Exercises for Section 9.2

1. Money/Real/Happiness. The money income is calculated simply by adding up all his expenditures, without paying attention to whether relative prices have changed from 1983 to 1984. The real income does pay attention to it. The utility change requires much more information: you need Galenson's cardinal utility function, not merely the prices he faced and the quantities he bought.
3. All of them, because in each the relative prices faced by each half of the pair differ.

Problems for Section 9.2

3. Manufacturing grew over the period from \$100 to \$140 (a 40% increase, as reported) in the prices of 1859, while agriculture shrunk from \$168 to \$155 (an 8% decrease). The total outputs in the two years (both years expressed in prices of 1859) are $\$100 + \$168 = \$268$ and $\$140 + \$155 = \$295$, giving a $(\$295/\$268) - 1 = 10\%$ increase. Note that the result lies as it should between the percentage changes for manufacturing (+40%) and agriculture (-8%) alone.

A slightly different way of doing the calculation gives the same result but mimics better what one does on a diagram of manufactures and agriculture (or Food and All Other Goods). Decide to express the increase in total output in manufacturing units, that is, choose the manufacturing axis as the one along which output as a whole is to be measured (compare Figure 9.6 showing Galenson's situation; note the "change of income measured in all other goods"). Since the price of a unit of manufacturing

relative to a unit of agriculture is here 1.0, the initial output measured in manufactures is $100 + (1.0)(168) = 268$ or, in words, the 1859 output of manufacturing plus the value in terms of manufactures of the 1859 output of agriculture. The increase in output from 1859 to 1874 is, likewise, the increase in manufactures (40) plus the increase (that is, decrease) in agriculture expressed in terms of manufacturing ($155 - 168 = -13$), or $40 + (1.0)(-13) = +27$. The increase divided by the 1859 level, therefore, is $27/268$, or 10%—as before.

4. The \$7,000 expressed in dollars of 1967 was $\$7000/0.945 = \7407 ; the $\$13,000/1.612 = \8499 . So the percentage rise was $8499/7407 = 1.15$, or 15%, or about 1.5% on average in each of the ten years. Notice that the money increase is much larger: $13,700/7000 = 1.96$, nearly a doubling instead of a mere 15% increase. Although the money increase is very relevant for the government's tax tables—taxpayers were pushed into much higher tax brackets by the rapid inflation of the 1960s and 1970s—what is relevant to real “purchasing power” is income deflated by an index of prices.

5. True. The underlying reasoning is monotonously familiar, namely, that giving a country the option of trading can do no harm and can possibly do good. Enlarging people's opportunities cannot hurt them. The diagram of national income might be Figure A9.1. Expressed in before-trade prices national income rises in proportion to the dashed budget lines. Expressed in terms of importables, for example, income rises by the amount of More Income. Notice that you could use the after-trade price (the solid budget line through After Trade) to arrive at the same conclusion. The magnitude of the rise in income would be different—quite naturally, since the prices used to evaluate a bundle are different—but the rise would still be a rise.

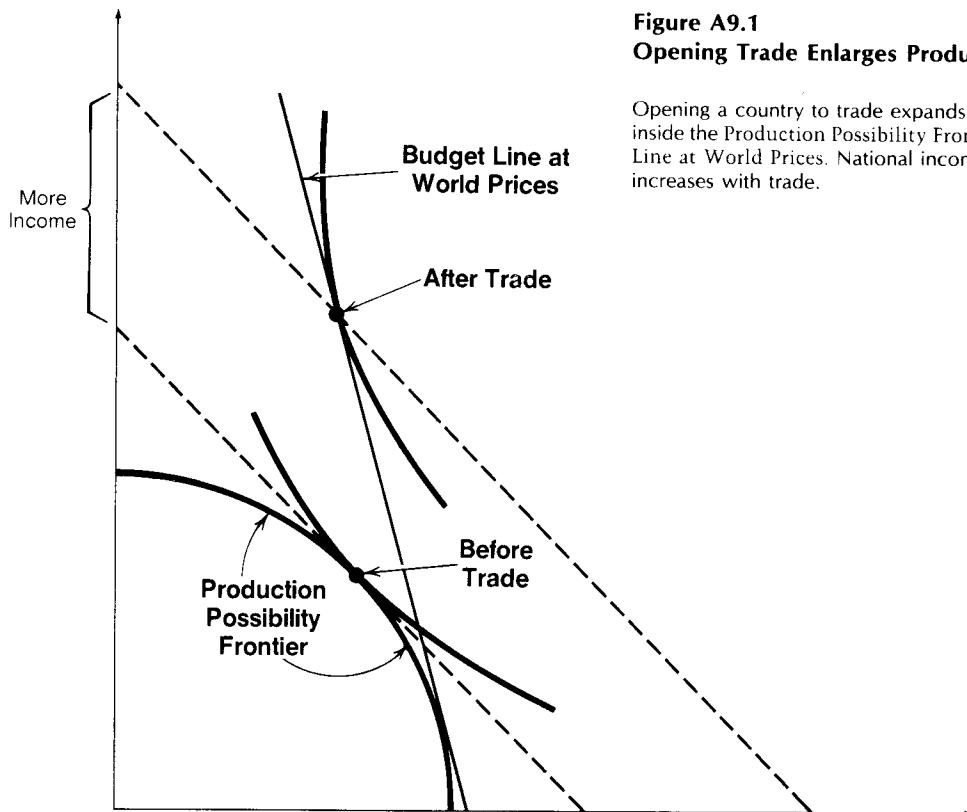


Figure A9.1
Opening Trade Enlarges Production Possibilities

Opening a country to trade expands its opportunity set from the area inside the Production Possibility Frontier to the area inside the Budget Line at World Prices. National income, measured in pre-trade prices, increases with trade.

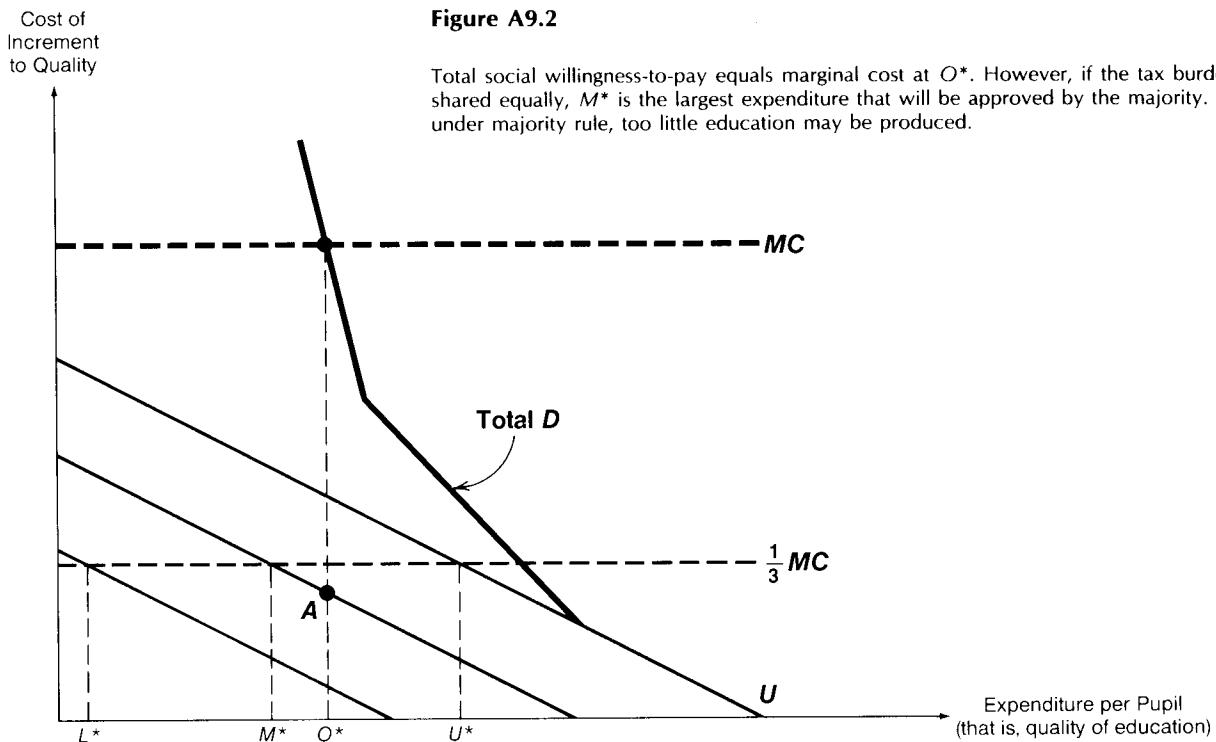
Exercises for Section 9.3

1. The building owners—and indirectly their customers or tenants—wish to pay to violate the building code. Each side sees its opportunity and takes it.
3. True. The bigger population gets more kicks in total out of the road, since all get some, and kicks can be added.

Problems for Section 9.3

1. True. If the bargaining starts with a preliminary agreement to give equal shares to all three, then any two—Ripoff and Incognito, say—gang up against the third, enriching each of them at the expense of the third. With Ripoff and Incognito each getting half, however, Highwayman has an incentive to offer his vote to one of the two in exchange for some (small) share of the pie. The coalition of Ripoff and Incognito, in other words, will break down. For the same reason so will any other coalition. Any division of the pie, in short, is better than any other, including itself. Such difficulty in agreeing how to divide the spoils is one obstacle to monopoly. The situation is one of pure bargaining, with all the unpredictability of that state.

2. The social willingness-to-pay is the *vertical* sum of the demand curves (see Figure A9.2). The key is understanding that each member of the community must consume the same quantity of education or, better, that they consume the same commodity, just as you are now consuming, whether you like it or not, a certain amount of American national defense. For a private good like candy, the marginal candy bar is being consumed by one person alone, so that what matters for achieving a social optimum is only *his* willingness-to-pay. But the marginal unit of educational quality or national defense is consumed by everyone in the community, so what matters is the total willingness-to-pay of the community, that is, the sum of all individual willingness-to-pay. The optimum is clearly, then, where the opportunity cost of an additional unit of educational quality equals the social willingness-



to-pay, namely, O^* along the expenditure (quality) axis in the diagram. If the marginal cost is shared equally among the three classes, each class will face $\frac{1}{3}$ of the total marginal cost when it collectively enters the voting booth. L will vote against any proposal to spend more than L^* per pupil, M will vote against more than M^* , and U against U^* or beyond. In the particular configuration of the curves given in the diagram, only expenditures less than or equal to M^* receive a majority (namely, the votes of U and M). Notice how the *median* voter runs the show: this is a feature of majority voting systems. This expenditure, less than or equal to M^* , is below the social optimum.

Clearly, though, you can make the system give the social optimum by manipulating the burden of taxes. Note that the question here is one of efficiency, not fairness. One could devise "fair" systems that get the requisite majority for exactly the optimal amount, but not all fair systems would get it, nor is it conceivable that unfair systems would achieve it, too. For example, in the diagram, one could lower M 's share of the burden of taxes and raise L 's: L is going to vote against the expenditure anyway (not always, but this is one way of putting it), so making him more unhappy does not affect the outcome. You could also raise U 's burden, taking care not to raise it so much that you turn him into an opponent of the optimal amount, O^* . You lower M 's burden until his burden line (analogous to $\frac{1}{3}MC$ in the diagram) crosses his demand curve at exactly the optimal amount (point A). There will then be a majority for the optimal amount, no more or less. See R. Barlow, "Efficiency Aspects of Local School Finance," *Journal of Political Economy* (September–October 1970), for a full treatment and further references.

CHAPTER 10

Exercises for Section 10.1

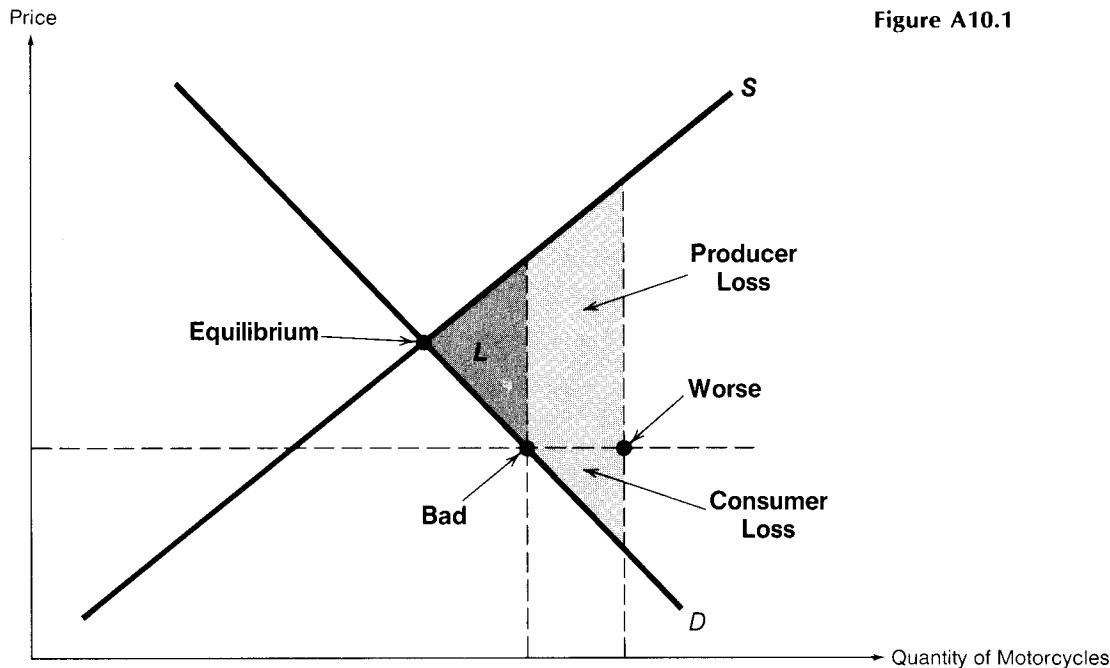
1. The actual prices paid and willing to pay would vary from person to person. For William Barker they are:

- a. \$10 at most, because he can easily wait until dinner to eat; he actually paid \$5. Surplus = $10 - 5 = 5$.
- b. \$5000, since he wouldn't want to go shoeless. Actually, he paid \$200 in total, so the surplus is \$4800.
- c. He would pay at most \$60, actually paid \$40, so his surplus is \$20.
- d. He would have paid \$500,000. He actually paid \$50,000. His consumers' surplus is \$450,000. The cost is realistic, incidentally: that is what you give up in earnings foregone.
- e. He would have paid only \$60,000—because there are plenty of other colleges—and actually paid \$50,000. The surplus is \$10,000.

3. Suppose that each of 100,000 houses sold for \$1000 each before the fire. The price at Supply Before the Fire is \$1000, the quantity 100,000, and the area $C + E$ is $(\$1000)(100,000)$, or \$100,000,000. To make the "value" after the fire greater than this the price must rise more than the proportion in which quantity falls. Quantity falls by half. So price must more than double. Suppose, for instance, it trebles, going from \$1000 to \$3000 per house. The area $B + C$ will be $(\$3000)(50,000) = \$150,000,000$. That's the paradox. But if price rises less than quantity falls, there will be no apparent paradox. For instance, if the price went from \$1000 to only \$1200, then the total "value" (in exchange) would fall, not rise. In other words, the "paradox" arises only when the demand curve has a price elasticity of less than 1.0.

Problems for Section 10.1

3. One could answer by going through the gains and losses from the beginning, but a simpler and more illuminating argument builds on the argument in the text, as in Figure A10.1. The point Worse is the one to be evaluated relative to Equilibrium. The point Bad, however, stands between them, and is on the demand curve. The argument in the text shows that going from Equilibrium to Bad costs society on balance the shaded area L . What, then, is the subsequent loss of moving further, to Worse? Moving off both curves in the Worse direction makes both consumers and producers worse off: No one gains, unlike the earlier move from Equilibrium to Bad (in which consumers gain and producers lose). In particular, consumers lose the area Consumer Loss, which is the excess of the



rectangle of their payment for the additional quantity over the trapezoid of their additional unwillingness-to-pay. Likewise, producers lose the area Producer Loss, the excess of the trapezoid of their additional costs over the rectangle of their receipts. The whole loss is the sum, the lightly shaded area. And the loss from the entire move from Equilibrium through Bad to Worse is the entire lightly and heavily shaded triangle. In other words, as was to be shown, it is bad on balance for society to stray from Equilibrium.

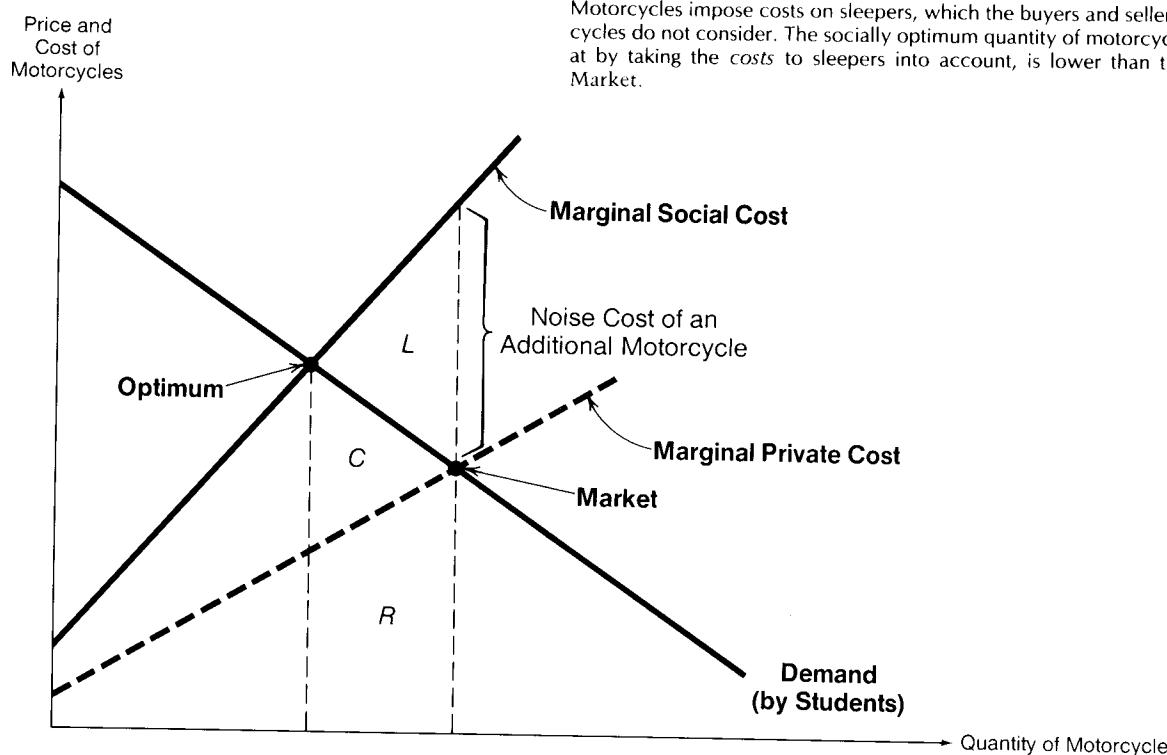
An even simpler argument follows from later developments of the idea of surpluses. Think of motorcyclists and sellers as one person, Society in Columbus. This person will incur larger costs than benefits if he produces and consumes beyond Equilibrium. The net loss will be in fact the big shaded triangle just identified in the diagram, since the triangle is the difference between the area under the cost curve and the area under the demand curve. And since society as a whole is worse off, any benefits to one segment must be more than offset by harm to some other, which was to be demonstrated.

5. The true marginal cost of a motorcycle, including every alternative sacrificed, is above the private cost perceived by the students. The students decide how many to buy on the basis of the low private marginal cost (the dashed curve in Figure A10.2); but if society as a whole had the decision it would decide how many to buy on the basis of the high, social marginal cost.

The optimum is Optimum, where marginal valuation equals (every and all) marginal cost. The market, however, leads to Market, where marginal valuation equals the cost faced by the students only. The excess number of motorcycles hurts society by the triangle L of the excess of social cost ($L + C + R$) over social benefit ($C + R$). Motorcyclists and their suppliers are better off by getting to Market, but by less than sleepers are made worse off.

8. True. Look back at Figure A10.2. The consumers and producers, who face the marginal *private* cost, would clearly desire as a group to move to Market, gaining the area C in mutual advantage. But each additional motorcycle beyond the low Optimum number puts costs on sleepers, in the summed account $C + L$ (which is the vertical distance between private and social marginal cost). So the hurt to sleepers exceeds the joint gain to producers and consumers by L . As one would expect, the loss from moving from the Optimum to the Market is borne by the third party, sleepers, who do not bring dollar votes themselves into the decision of how many motorcycles Columbus should have.

Figure A10.2
Sleepers of Columbus, Lend Me Your Ears; Or, Why Too Many Motorcycles Exist



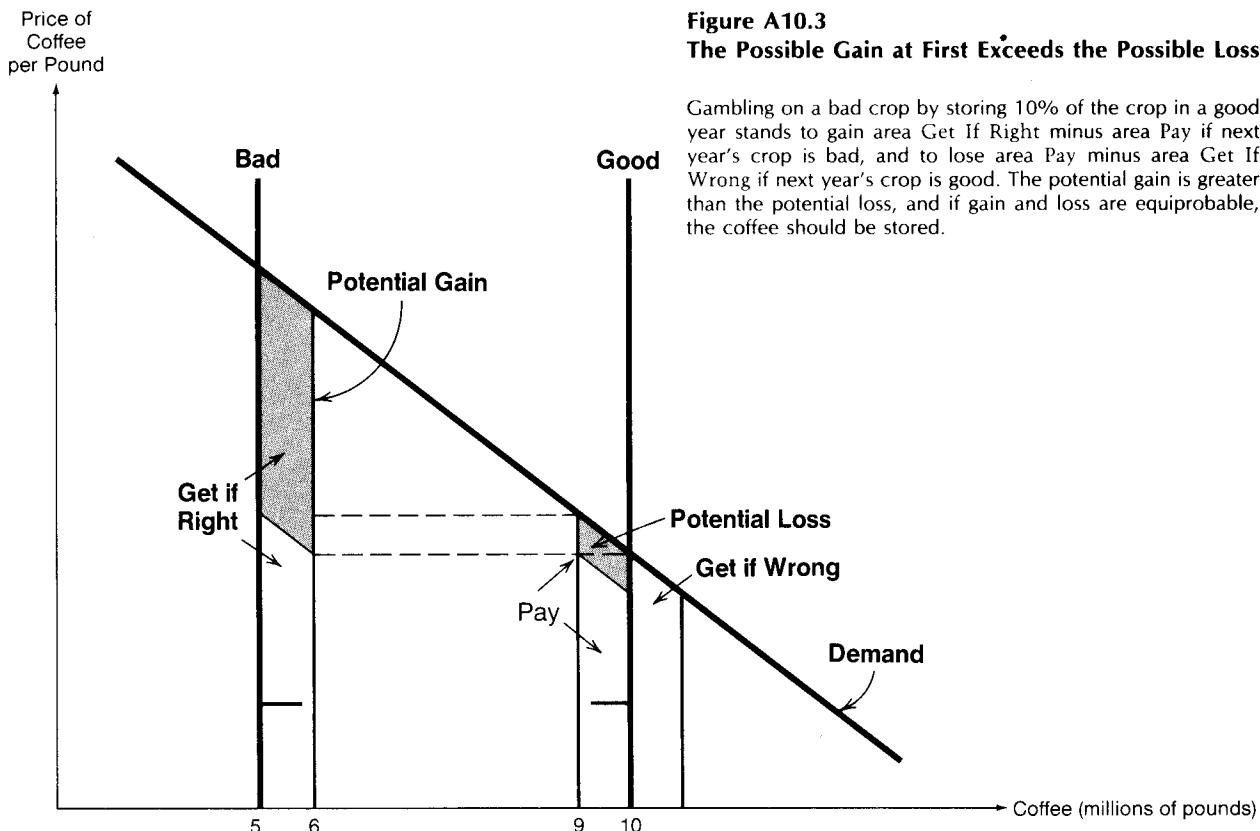
Motorcycles impose costs on sleepers, which the buyers and sellers of motorcycles do not consider. The socially optimum quantity of motorcycles, arrived at by taking the costs to sleepers into account, is lower than the quantity Market.

Exercises for Section 10.2

1. Obviously, he'll store coconuts for a rainy day. In a sense, the off-season coconuts are expensive, and Crusoe, like Baker, is "transporting" coconuts by storage from the low-price time to the high-price time. Crusoe is the same person in both times. So he wants the same amount to eat. Like the analysis of Baker and coffee speculation, Crusoe stores until the quantities consumed at the two times are equal. So too is the "price"—namely, his marginal valuation of coconuts at different times.
3. All three buy low and sell high. Each case is identical to the coffee problem. The deals are made voluntarily. Society is better off, at least if society is not pained by the very thought of people buying basketball tickets, tires, and oil as they will.

Problems for Section 10.2

1. a. You make the left shaded area of Potential Gain in Figure A10.3, which is the entire tall trapezoid on the left (called Get If Right) minus the shorter trapezoid on the right (called Pay: note that Pay includes the little shaded area). The unshaded portion of the tall one, you see, is equal to Pay. You pay Pay for the 1 million pounds in the good year, selling it at a high price in the bad year.
- b. If you are wrong you must sell the rotting coffee in a good year, that is, at an even lower price than the price at which you bought it—in fact, at Get If Wrong. You lose the area Pay minus Get If Wrong, the small parallelogram called Potential Loss.
- c. The relevant areas will be similar, but for larger quantities transferred. The area of loss grows,



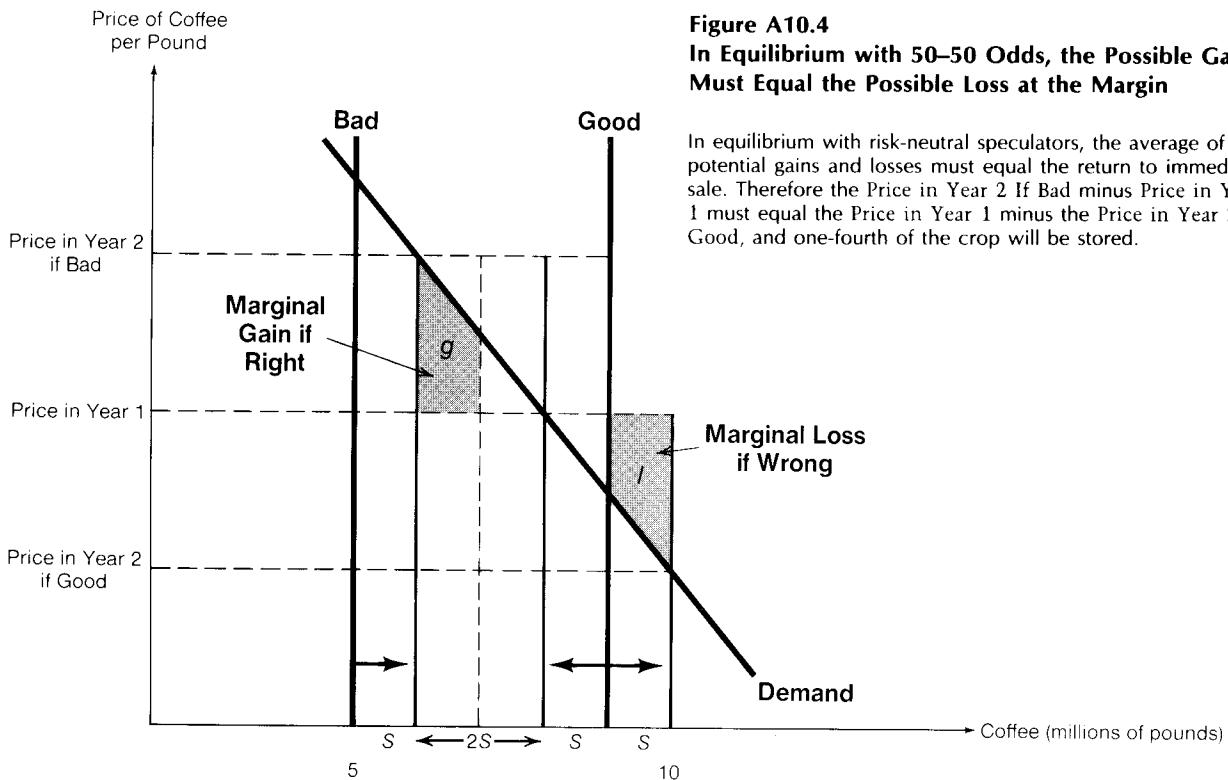
because the price at which you have to sell the surplus coffee is smaller for larger surpluses. The area of gain grows and then finally declines, because the higher price at which you have to purchase coffee to store as you buy more finally offsets the rising revenue from selling it to a coffee-scarce future.

d. From the point of view of each of many competitors, none of whom has himself a noticeable impact on the prices of coffee, the transfer of coffee remains profitable until the marginal pound of coffee will yield a profit if you are right (half the time, by assumption) exactly equal to the loss if you are wrong. Figure A10.4 illustrates such an equilibrium of marginal cost and benefit for the last pound put in storage. Regardless of the slope of the (straight-line) demand curve, exactly one-fourth of the difference between Good and Bad is stored. *Proof:* All the S's under the arrows must be equal because the arrows indicate the increments and decrements to the stored coffee stocks. Further, the distance marked "2S" must indeed be $2S$ because only then (with straight-line demand) will the marginal gain equal the marginal loss (that is, only then will the shaded trapezoid g be identical to the shaded trapezoid l). So the distance between 5 and 10 million pounds is split into fourths.

3. True. The diagram, similar to the one in the text, is Figure A10.5. The gross gain from the storage to consumers in the bad year is $E + F + G$, the gross loss to consumers in the good year is $H + I$. But if the demand curve is a straight line, G equals I and E equals H . So F is what is left over, the net gain. And F is the product of half the two differences, as asserted.

4. False. Use Figure A10.5, labeling outputs of 1.00 and 0.90 with prices of 1.0 and 1.20 and perform the calculations: $[(1/2)(0.10)][(1/2)(0.20)] = (1/4)(0.02) = 0.005$, not 0.02.

8. Well, sure they will: After all, the differential in price is the incentive to store and resell. The



incentive remains until it is eliminated, that is, until the differential is zero. One cannot easily use a single diagram to make the argument, but two side by side will do, as in Figure A10.6.

Problems for Section 10.3

1. Yes. To put the question another way, is the trapezoidal area shaded in Figure A10.7 a correct measure of the consumers' surplus accruing to the Western demanders alone? That is, is area D equal to area Δ ? Yes, it is. The area of a trapezoid (which is what both D and Δ are) is its height multiplied by the average of its top and bottom widths. Although they are placed differently relative to each other, the top and bottom widths in D and Δ are the same, because it is by adding the widths on to the Eastern demand curve that the curve Market Demand is constructed. Since the widths are the same their average is the same (the heights are certainly the same), and Δ is the same as D in area.

By approximating a curved demand with a succession of little straight lines, applying the proof in the text to each segment, and (as they say in first-year calculus) taking the limit one can show that the proof does not depend on the simplifying assumption of straight-line demand curves. In short, you can work with one rather than many diagrams in dividing surpluses into parts.

3. True. In Figure A10.8 the old sum of consumers' and producers' surplus is the diagonally shaded triangle. The new sum is the old area plus the vertically shaded area. The vertically shaded area is the difference between the two and therefore is the gain expressed in units of All Other Goods.

A fuller argument would get the sums themselves by starting with areas of society's willingness-to-pay (in the after-railways case it is the trapezoid C plus the shading) and subtracting out the cost in all other goods sacrificed of providing that much transport (in the after-railways case it is the area C).

As before, the shaded area of net social benefit can be broken into two parts, the area of Cost

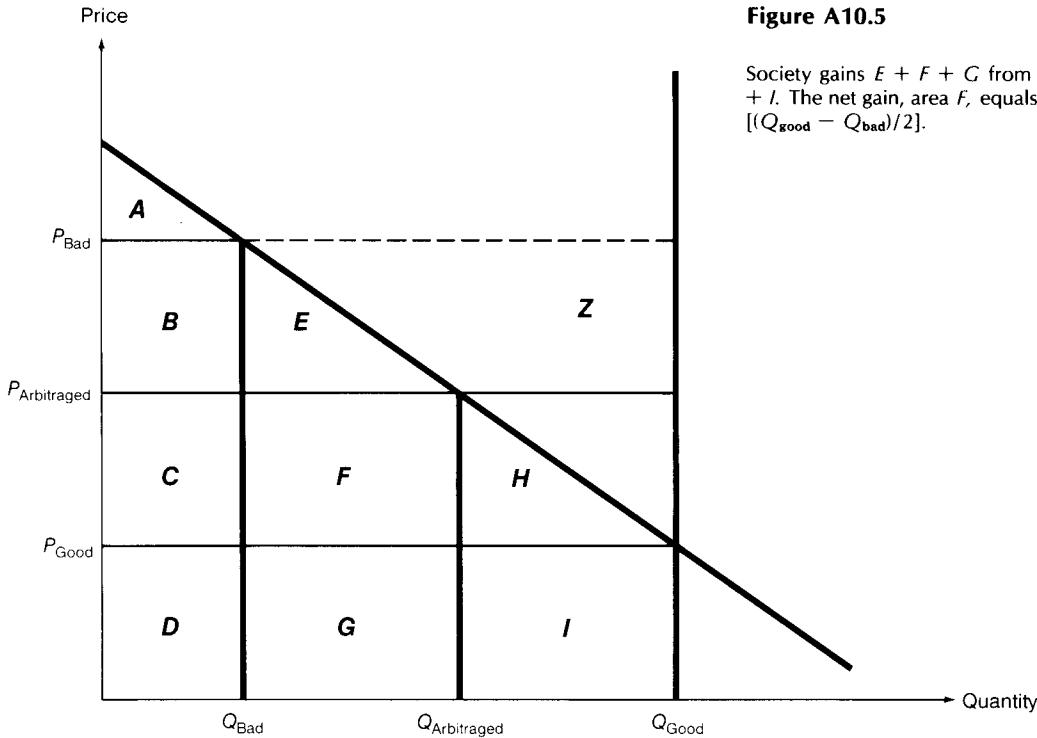


Figure A10.5

Society gains $E + F + G$ from arbitrage at a cost of $H + I$. The net gain, area F , equals $[(P_{bad} - P_{good})/2 \text{ times}] [(Q_{good} - Q_{bad})/2]$.

**Figure A10.6
Equal Prices End Arbitrage**

Arbitrage continues until prices are equalized across time. The result that prices must equalize holds irrespective of the shapes of the demand curves in the various periods.

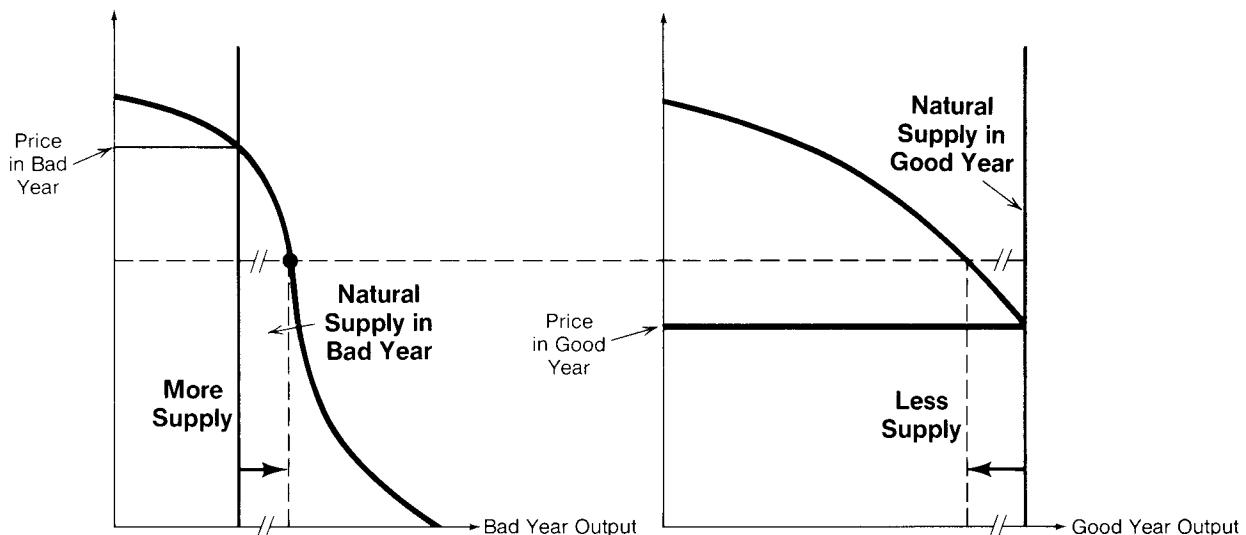


Figure A10.7
Areas Between Horizontally Added Demand Curves Are Correct Measures

Area D and Area Δ are both measures of Western consumers' surplus. Horizontal addition of demand curves implies that the Western Demand curve is to the price axis as the Market Demand curve is to the Eastern Demand curve

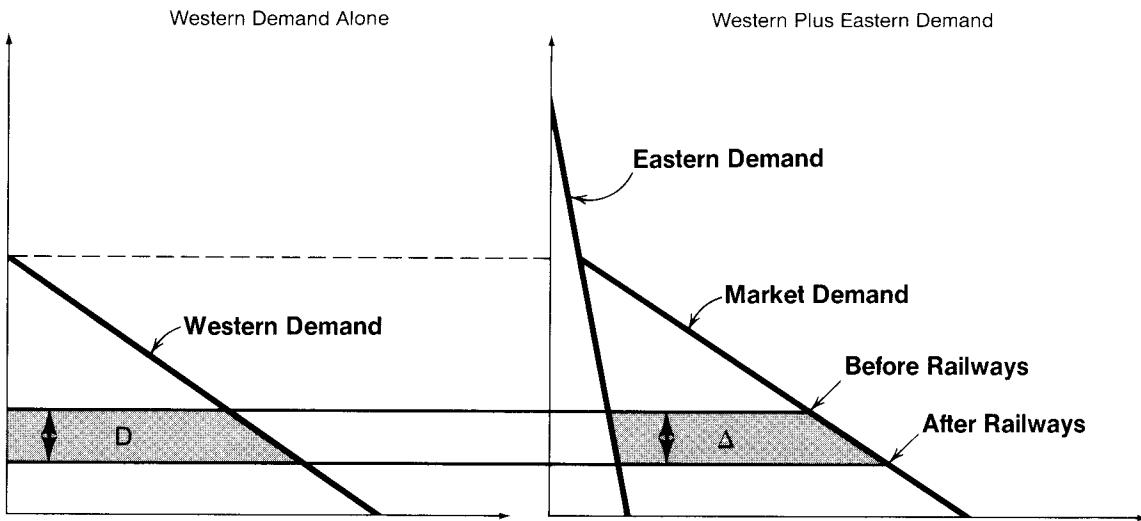
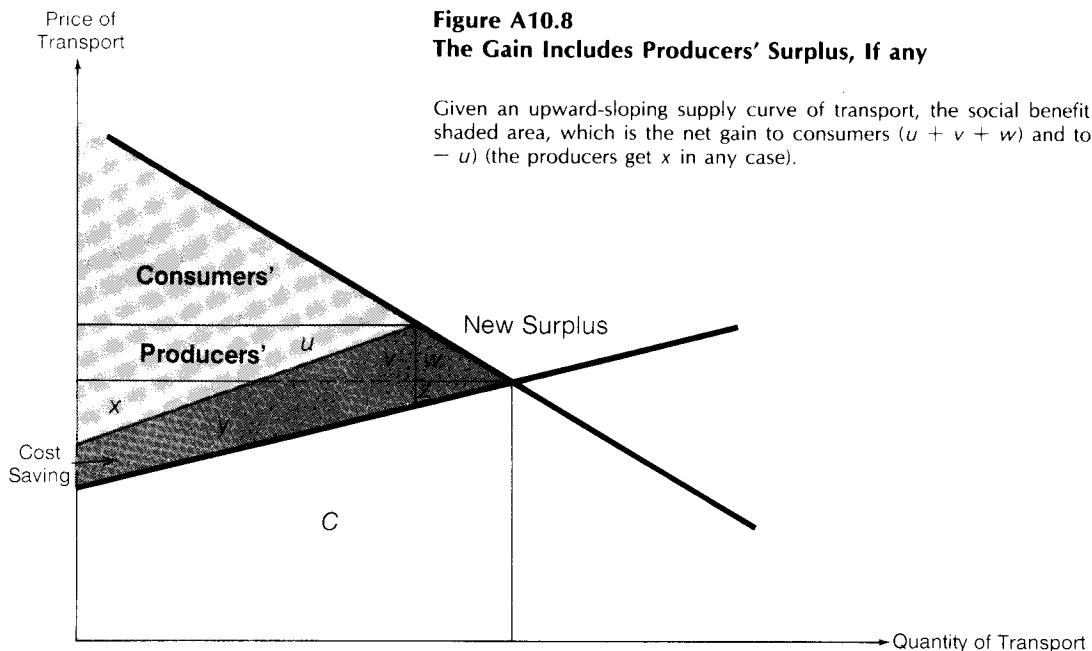


Figure A10.8
The Gain Includes Producers' Surplus, If any

Given an upward-sloping supply curve of transport, the social benefit of railways is the shaded area, which is the net gain to consumers ($u + v + w$) and to producers ($z + y - u$) (the producers get x in any case).



Saving on the old output ($v + y$) and the area of new transport ($w + z$). And notice too that if you wanted to divide up the social benefit into its distribution between consumers and producers you would add up the areas marked u , v , w , x , y , and z :

Distribution of the Spoils

Increase in Producers' Surplus		Increase in Consumers' Surplus
New Producers' Surplus = $+z + y + x$		
<i>minus</i>		
Old Producers' Surplus = $-u - x$		
		$+v + y + w$
Increase in Sum of Consumers' and Producers' Surplus $= z + y + v + w$ (which is the vertically shaded area)		

The moral, to repeat, is that consumers' and producers' surplus can be measured up or measured across. The two will be the same if measured correctly.

CHAPTER 11

Exercises for Section 11.1

1. The richness of the answer will depend on how concretely you can think about the wheel of wealth:
 - a. Staggs the lumberjack for wood, Redfield the rubber plantation owner for the rubber for the eraser, and so forth.
 - b. Likewise: the rancher, the wheat farmer, the sesame seed shipper, the waitress, the trucker.
 - c. The copy editor of this book, the acquisitions editor, the author, your teacher, your janitor, your college president, and so forth.

Problems for Section 11.1

1. For the diagrams themselves, look into the macroeconomics section of your first-year textbook in economics. The essential point is to create entities (for example, the Foreign Sector) that buy and sell from consumers.
3. He does not pay two "profit margins." He pays whatever is necessary to compensate the firms for the book finding and storing services they provide. The firms are not pointless middlemen tacking on an arbitrary profit margin to each transaction. As we shall see later, competition would not allow such useless firms to survive. Mr. Murray-Smith would like to provide his own book binding service. But doing so in the specialized columns of *The Clique* would reduce the value of the magazine to its subscribers, professional bookmen. As usual in economics, the point is clearer if carried to an extreme. Why should not all the Mr. Murray-Smiths (all 100,000 of them) be invited to inspect the library of a recently deceased collector in Berkshire, eliminating middlemen such as lawyers, auctioneers, book dealers, bookstores? Why in the case of books by authors still writing should he and his companions not each get a book in handwritten form directly from the author in Kent, eliminating middlemen such as typists, literary agents, publishers, printers, book wholesalers, and yes, bookstores? He could, but specialization of these functions in separate firms makes the final bundle of goods and services called a book cheaper.
4. True. By the Wheel of Wealth the one is equal to the other. Firms produced in 1975 about \$1.2 trillion worth of final goods and services; which is to say that people earned \$1.2 trillion from the operation of these firms, as their employees, landlords, creditors, or owners. The balance of the bookkeeping makes it possible to check national "product" against national "income."

One could estimate it for England a century ago, say, either by valuing and adding the ships, sealing wax, cabbages, royal services, and so forth bought or by valuing and adding up the seamen, wax, cabbage patches, kings, and so forth hired. The two are not exactly the same because the registration of ships, agricultural surveys, income tax returns, expense accounts of the royal household, and so forth on which they are based are not perfectly accurate. But they come out close to equal: in the United Kingdom in 1870 they differ by only 14%; in 1920, 8%; and in 1965, 7%.

Exercises for Section 11.2

1. A rich farmer can achieve these other things. Money buys land, buys the ability to follow in one's father's footsteps, buys leisure, buys independence. It doesn't buy everything, but it does buy a whole lot.

3. Take the equation for the difference in the cost of Lincolnshire iron and ask what cost of coke (C) would make it zero:

$$\begin{aligned} & (-0.5 \text{ shilling})(3.5 \text{ tons of ore/ton iron}) \\ & + (C \text{ shillings})(1.5 \text{ tons of coke/ton iron}) \\ & = 0 \end{aligned}$$

Solving for C , $C = (0.5)(3.5)/1.5$, or 1.1667. So at a cost difference of coke of $1\frac{1}{2}$ shillings per ton or lower (as against the actual 2.0 shillings), Lincolnshire would have been a good location. It soon became one, in fact.

Exercises for Section 11.3

1. Carstensen at 10,000 bushels earns \$20,000 of revenue. His costs are \$18,762, so his profit is \$1238. Spechler earns the same \$20,000 but spends \$32,500, for a loss of \$12,500. In like fashion:

At:	Profits (or Losses, Negative) for:		
	Carstensen	Gregory	Spechler
10,000 bushels	\$ 1,238	\$ 2,500	\$ -12,500
30,000 bushels	6,250	12,500	-22,500
60,000 bushels	-5,000	-10,000	0

3. At 30,000 bushels for Carstensen it is \$53,750 — \$53,748 for the 1 bushel between 29,999 and 30,000. So it is \$2.00 exactly. In like fashion, the table is

At:	Marginal Cost for:		
	Carstensen	Gregory	Spechler
10,000 bushels	\$1.50	\$1.00	\$3.00
30,000 bushels	2.00	2.00	2.00
60,000 bushels	2.75	3.50	0.50

5. Spechler's marginal cost *declines*, from \$3.00 to \$2.00 to \$0.50. It would be irrational for him to stop at 30,000 bushels, even though price is equal to marginal cost there. At higher outputs, say 60,000 bushels, his marginal cost falls *below* marginal revenue (always \$2.00). His marginal costs

and benefits are signaling him to produce more. And more. And more. If his marginal cost at no output rises he will continue searching for Just Right at higher and higher outputs.

7. Carstensen incurs losses \$6250 plus \$5000 if he moves from 30,000 bushels to 60,000: a total area of Losses Incurred of \$11,250 (see Figure 11.3). Gregory—again, it is his steeper marginal cost curve that is causing the difficulty—loses much more, \$12,500 + \$10,000, or \$22,500.

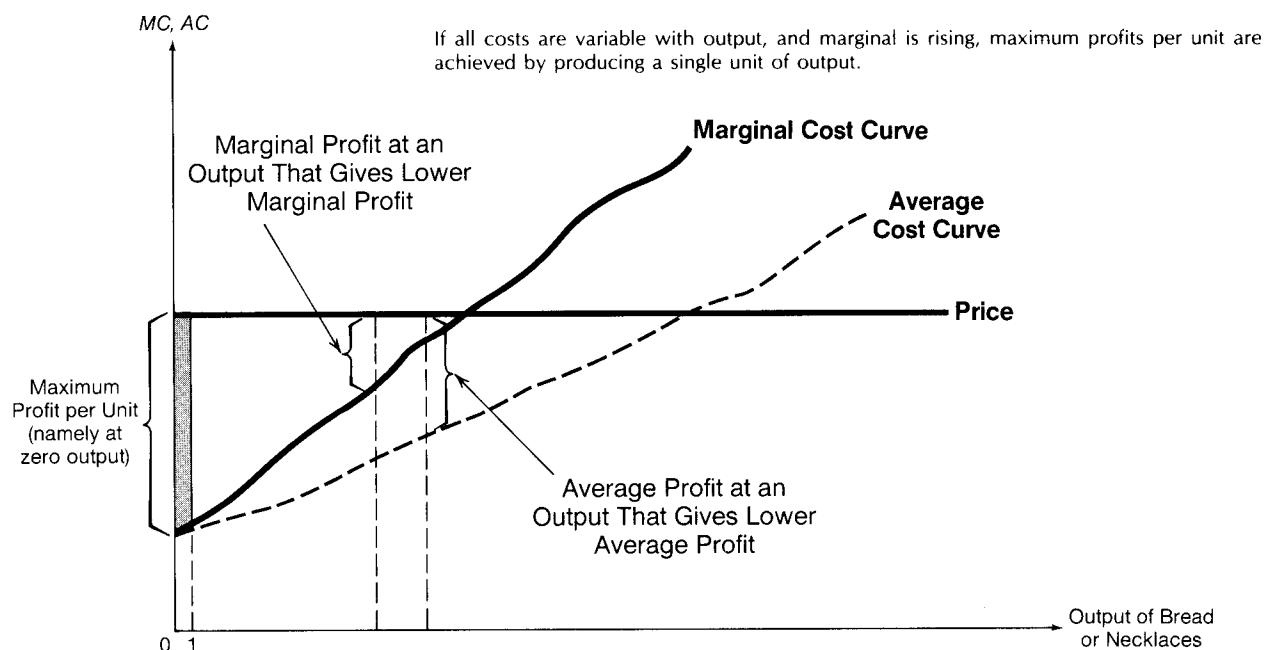
Problems for Section 11.3

2. True, and absurd, which suggests that there is something peculiar in the reasonable-sounding formula that traditional businessmen would maximize profit per unit of sale. The little vertical shaded area in Figure A11.1 is the amount earned. At the lowest point of the curve of say 1000 loaves per year, the average profit per unit is as high as possible (the marginal profit, though, is still as high as possible at 1 loaf). The baker, in short could have produced more than 1 loaf (in this case 1000) even though he was following a rule of maximum average profit. What makes this interpretation inconsistent with the historical argument, by the way, is precisely the implication that businessmen minimize average cost. On the contrary, the very point of the historical argument is to assert that French businessmen cared little for narrowly rational matters such as costs. In any event, the first loaf or necklace would make the most per unit. It is not merely a small amount but a virtually zero amount that is implied by maximizing profit per unit.

COMMENT

Although inconsistent with the historical argument here, there is a way of rescuing the economic logic. If over some early output the cost curves fall, then the per-unit maximization does not yield patently absurd results. A baker, for example, might have had a large and expensive oven. His average "variable" costs of good flour and good workmen might well have risen with each new loaf, but his average "fixed" costs of paying for the oven declined. The discussion in the next section will explore

Figure A11.1
Maximum Profit per Unit Occurs at Zero Output if Costs Are Rising



these terminological wonders. For present purposes the only result needed is that an average cost curve under such conditions is shaped roughly like a U, that is, at first falling as the fixed cost of the oven is spread over larger output and at last rising as the spread-out fixed costs come to be dominated by rising variable costs, as in Figure A11.2.

Exercises for Section 11.4

1. Equalize the marginal intellectual value of each hour of Latin to that of Economics. For instance, if your productivity in Economics falls off quickly, study more Latin.
- b. Equalize the marginal (not average) costs of a kilowatt-hour from each plant, shifting output around until they are equal.
- c. Equalize the marginal cost of borrowing from each bank, instead of borrowing everything from one bank, which feels it must charge a high rate for bearing the entire risk.
- d. Equalize the marginal value of the soldier on the hill to one in the valley.

Problems for Section 11.4

1. The dashed line in Figure A11.3 is total total cost (yes, two "totals"). Draw the diagram now if you did not in attempting to answer the questions: you will find it more difficult to produce than the

Figure A11.2
Maximum Profit per Unit with Fixed Costs Occurs at Less Than Profit-Maximizing Output

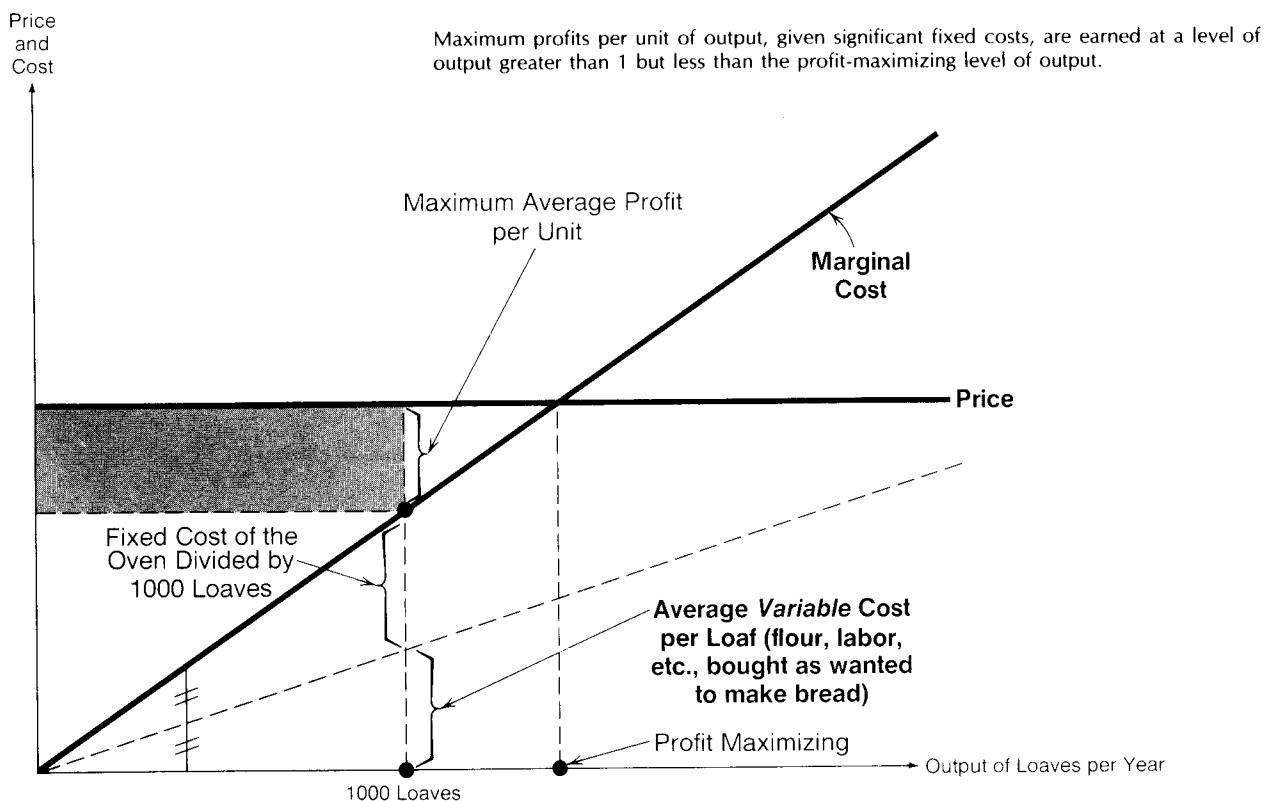
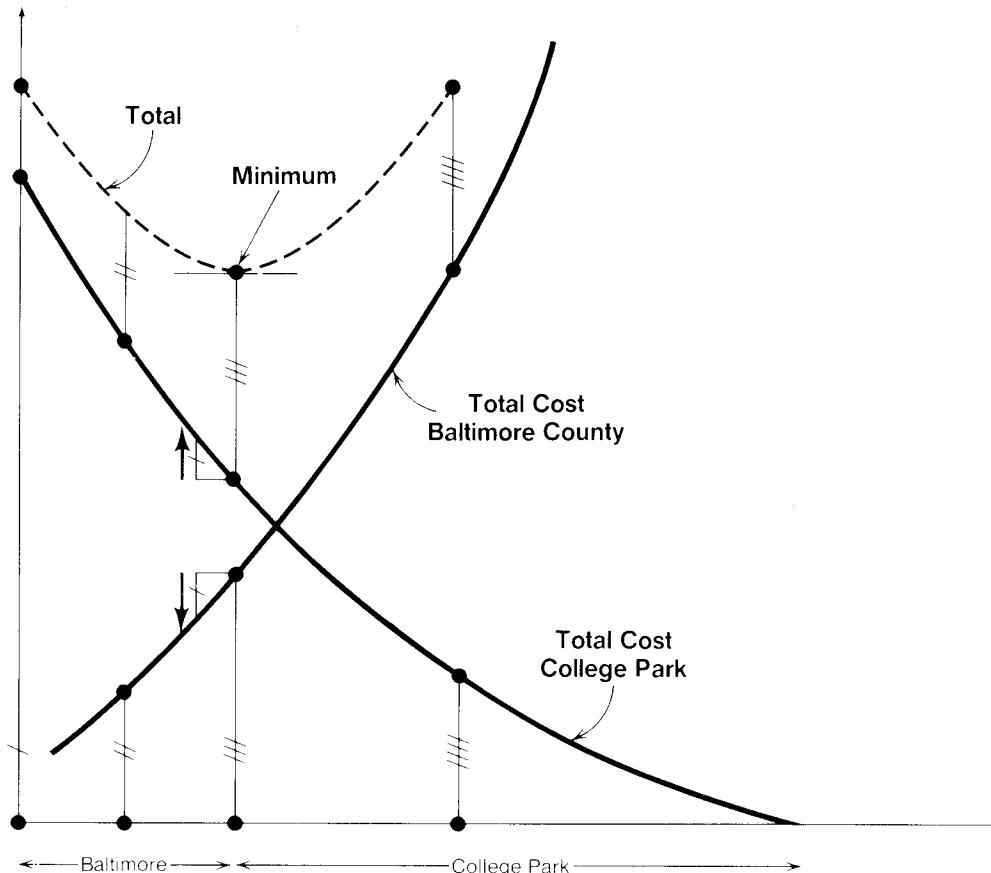


Figure A11.3
Minimum Cost Using Total Cost

Minimum total cost is reached at a point where the marginal cost of assigning a student to Baltimore County equals the marginal reduction in cost achieved by removing the student from College Park.



corresponding diagram using marginal magnitudes. The point marked Minimum has a flat slope, which is to say that a small reallocation toward College Park causes a rise in costs there just equal to the fall in costs at Baltimore County: the two little triangles of slope just offset each other, leaving total total cost unaltered. The point of minimum total total cost, therefore, is the same as the point of equal marginal costs, as one would hope. Notice the analogy to the inventory diagram. The allocation having equal average cost has equal slopes of rays from the origins. It is a difficult geometric trick to find it, emphasizing again the clumsiness of the diagram relative to a diagram of marginal (and on the same axes average) cost. In fact, it is easy to draw total cost curves for the two schools that have no allocation between them with equal average cost.

CHAPTER 12

Exercises for Section 12.1

1. None are acceptable. Each has increasing returns to scale. For instance, in (b), when L doubles, Q quadruples. More and more labor produces many more times more output, without end. None of the three functions has a section of diminishing returns.
3. Minimum average cost is at about 500 gallons, where average and marginal cost are equal. If he gets \$16 a gallon he produces at 675 gallons, this being where marginal benefit equals marginal cost.
5. At \$1.00 he sets the marginal benefit equal to marginal cost by producing 800 gallons, where marginal cost is also \$1.00. At \$3.20, by the same reasoning, he produces 1575 gallons. At \$8.00, 1800 gallons. In other words, *his supply curve is his marginal cost curve*: he moves up and down as price rises or falls.

Problems for Section 12.1

1. Since output is proportional to the cubic volume, it will be true that $Q = Kh^3$, where K is the constant of proportionality. Since cost is proportional to the surface area, and a cube has six sides of area h^2 each, it will also be true that cost = $g6h^2$, where g is another constant. Taking the hint, solve for the height, h , from the equation for output to get

$$h = \left[\frac{1}{K} Q \right]^{1/3}$$

Then substitute this expression into the equation for cost (that is, into cost = $g6h^2$) to get

$$\text{Cost} = g 6 \left[\frac{1}{K} Q^{1/3} \right]^2$$

This can be written more simply as cost $cQ^{2/3}$, in which c stands for the collection of constants. The exponent on Q is $\frac{2}{3}$, as desired. Notice that since the elasticity, $\frac{2}{3}$, is less than 1.0, the function becomes flatter as Q increases, that is, marginal costs decline.

COMMENT

The problem is typical of others concerning the cost of output of pieces of equipment, especially in industries such as chemicals and metals that handle bulk inputs with vats, tubs, tanks, furnaces, and other things having volumes and surface areas. For example, a tank in the shape of a sphere (such as a natural gas tank) also gives a $\frac{2}{3}$ exponent on output. Engineers call the result the "sixth-tenths factor" (0.6 is close to $\frac{2}{3}$), and find it true even for total costs (as against mere construction and maintenance costs of the plant's shell). It is one rationale for economies of scale.

Exercises for Section 12.2

1. No. The line through the points would be biased up, and maybe biased in slope, too. The points aren't on the Cost Function. The better estimate would be the lower envelope of points. Actually, people do estimate cost curves by running lines through such points. Not very smart.
3. False. Like the sewing machine problem, Sears is roughly indifferent between leasing and owning: It doesn't "save rent" by owning, since it faces the opportunity cost of not getting rent from someone else if it chooses to put its stores on the land.

Problems for Section 12.2

1. a. No. The vertical gap between the total cost measured by the Old and the New curve is not enough to offset the Transaction Costs. At an even higher output the gap widens and the Transaction Costs are overcome. But not at New Output. Despite the new output you stick with the Old cost curve.
- b. You will adjust to the new fleet when the Old cost curve has risen enough to offset Transaction Costs. The rise in cost is the effect of wearing out, or physical "depreciation" of the old fleet.

c. Same as (b), except that the gap widens because the New is falling instead of the Old rising. The fall in cost is the effect of obsolescence of the Old rising. The fall in cost is the effect of obsolescence of the Old fleet, obsolescence being part of full *economic* depreciation (physical wearing out being the other part).

d. Many: investment would be high because for all three reasons a given amount of transaction costs would be overcome frequently.

Exercises for Section 12.3

1. In the nature of the exercise the book cannot give an answer, since you have supplied all the data. But your diagrams should look a lot like Figure 12.11, though not of course the same in exact shape.

3. Again, no answer.

5. All except (g) have elements of fixity. The heating, untenured faculty, and dining room fees are fixed (somewhat) and avoidable: if the college shut down tomorrow these costs would fall to zero (as would of course the variable cost, the purchases of exam booklets: no students, no exams). The payments by Wretched to the college and the college to tenured faculty are fixed and unavoidable, although one can quarrel about both (under recent court decisions, for instance, a college can fire "tenured" faculty if it closes down an entire section, such as the whole Department of Economics). Whether or not the incorporation fees are avoidable depends how extreme the closing down is: if the college not only has no students but turns in its badge, so to speak, then they are avoidable (unless they were a one-time payment, in which case they are unavoidable—or irrelevant—because bygone).

Problems for Section 12.3

1. The area $A + B$ is total revenue (average revenue—which is price—times quantity) minus the area under the marginal cost curve. The area $A + D$ is total revenue minus the area of average revenue times quantity. The areas F and G are irrelevant.

3. a. The total revenue curve is an inverted U, with average revenue steadily declining. The inverted U is at a peak when the elasticity of demand is zero, that is, when total revenue along the demand curve is at a maximum.

b. It would set its output to maximize the distance between total revenue and total cost. No, price is not equal to marginal cost in this equilibrium. It is above marginal cost: a monopolist such as Saudi Arabia is led as by an invisible hand to produce the *wrong* output, namely, an output that does not equalize the marginal cost of production to the marginal valuation consumers put on the production.

CHAPTER 13

Exercises for Section 13.1

1. Since Kruskal is "typical," he's obviously going to supply one hundredth of the total supplied by the hundred. So Kruskal produces 1 brick and the industry produces 100. If Kruskal increases output to 1.33, a big increase from his point of view, the increase of output for the industry is only 0.33 out of 100, or only $\frac{1}{3}$ of 1%. One out of many little suppliers can't affect the output of the industry much.

3. Taking a price as given means supposing a flat demand curve (these are suppliers, so they face a demand curve; they *have* a supply curve). But a flat demand curve is a curve with a high price elasticity. Kruskal, being a small part of the whole, faces a demand curve with a high elasticity, as was shown in Exercise 2. So Kruskal is a price taker. He takes the price as (approximately) given, because the elasticity of demand he faces is so high.

Problems for Section 13.1

2. a. The Rule of Rational life in its various forms, described in Chapter 12, tells you that marginal cost in each plant must be equal to the marginal cost in every other plant. If not, reallocating output

among the plants could lower the cost of a given output. Likewise, to maximize profit the marginal benefit (that is, price) must equal the marginal cost. The condition is that marginal cost = \$1.

b. You would need to know their marginal cost functions and the market price, that is, the information needed to satisfy the conditions of Problem 1. The three marginal costs are

$$\frac{4}{1000} q, \quad \frac{4}{1000} q, \quad \text{and } \frac{2}{1000} q$$

Setting each marginal cost equal to the (same) marginal benefit of \$1 per pen implies (\$1) (1000/4 pens per \$) = 250 pens from both the Craven and the Farrington plant (notice that the difference in fixed costs of \$100 makes no difference); and 500 pens from the Paul plant, with its less sharply rising marginal cost curve.

c. The results of the game will of course be identical to the central planning of part (b). Now each manager will himself maximize profits in view of "his" marginal benefit (\$1) and his plant's marginal cost. In fact, if the managers had time to experiment each could *discover* the marginal cost, not knowing it initially.

d. The decentralized system saves the cost of transmitting the knowledge of cost curves up from the branch plants to the central office. It relies on the knowledge in the place it is most likely to be accurate, namely, in the branch plant itself.

4. The three smallest sizes (up to 49 slaves) were the least successful, losing among them 3.6% of the whole agricultural slave population, or 4.6% of their base population. The 3.6% went to the other sizes, and represented a 17% increase from their (smaller) 1850 base. The most growth comes in the 50–199 sizes; the over-200 sizes are a little less successful but still gain. Since the decade after 1850 brought a relative contraction of farms with under 50 slaves, it would seem reasonable to suppose that these were in 1850 below the best size. That is, one would draw an average cost curve that sloped downward until it reached 50 slaves and then leveled off: over 50 slaves there was not much difference in the survival qualities of different sizes. One would conclude, as did Robert Fogel and Stanley Engerman in their study of slavery, that there were in fact economies of scale (50 slaves is an unusually large plantation, incidentally, although our image of American slavery portrays such plantations as common). [*Time on the Cross: Evidence and Methods—A Supplement* (Boston: Little, Brown, 1974), p. 144.]

The general principle is:

Whether or not the advantages obtained by operating on a large scale preponderate in any particular case over the more watchful attention, and greater regard to minor gains and losses, usually found in small establishments, can be ascertained . . . by an unfailing test. Wherever there are large and small establishments in the same business, that one of the two which in existing circumstances carries on the production at greatest advantage, will be able to undersell the other. [J. S. Mill, *Principles of Political Economy* (various eds., 1848–1871), Book I, Chapter IX, Section 1]

It has been reinvented for moderns by George Stigler (a famous student of the history of economic thought who once wrote an essay asserting that the study of such history was useless for modern economics) in his "The Economics of Scale," *Journal of Law and Economics* 1 (October 1958): 54–71, and is now known as the "survivorship principle."

Exercises for Section 13.2

1. If oil were specialized in gasoline production, then clearly it would be forced up in price as the demand for gasoline rose. The oil has nowhere to come from, no other use. The same would hold true, though to a lesser extent, if it were (as it is during the summer) heavily but not completely specialized. All the demand must be worked out in a rise in the price of oil.

3. No. A low share means that the input doesn't contribute much to total cost. Take an extreme case: if the input accounted for only 1% of costs, then even a doubling of its price would cause the cost to rise only 2%.

5. a. Beef cattle: specialized (completely), relatively inelastically supplied (though not in the long run), and a large input (about a third of the cost).

- b. Dairy cattle: same story.
- c. Stonecutters, although the coming of sandblasting and power tools has reduced the skill in this, and therefore its specialization and inelasticity.
- d. Wood, pulp mills.
- e. Economists—though not inelastic in the long run. Classics professors are more specialized in teaching (well over half of the economists work for industry and government).
- f. Quarterbacks; runners with good hands able to run the 40-yard dash in 4.7.

Problems for Section 13.2

2. True. The number K will be the amount of fluctuation of the demand curve at a certain price. If the supply curve were upward sloping, output would fluctuate *less* than K . If, however, economies of industry scale were great enough to make the supply curve slope *down*, output would fluctuate *more* than K . If you observe such an industry, therefore, you can rightly suspect that it may be experiencing strong economies of industry scale.

CHAPTER 14

Exercises for Section 14.1

1. a. Yes: the “feature-length film” industry. All are close substitutes in the minds of consumers. All can profitably supply something that consumers think of as virtually one product.
- b. Yes, to the extent that students choose across universities. Not every student needs to consider seriously each university: as long as many customers consider two or three each there will be enough overlap.
- c. No, although a comparison with (b) shows how arbitrary such decisions are. The technology is the same (clerks, cornflakes, cash registers, and so forth) but the customers do not overlap. Well, maybe they do as much as do customers of universities. Each customer considers two or three each, after all.
- d. No, because the definition would exclude foreign producers. For lobbying in Congress to get restrictions on bicycle imports the American industry alone is a meaningful entity. But for competing to get customers it is not, and this is the key.
3. The curve will have a flat stretch at \$5 an hour, which is by far the most common entry price.
5. Economic rent, supernormal profit, producer’s surplus.

Problems for Section 14.1

2. Such an attempt to stop “profiteering” will in fact increase it. The manufacturers operating before 1914 would obviously make larger profits if the new firms were cut out, which is one way of seeing the point. Another is that the slope of the supply curve must rise if the new firms are removed. Since demand is inelastic the area of profit behind the supply curve must therefore rise.
3. a. No, it is not, no more than the low apprehension rate of thieves is evidence that the criminal law has no effect on the behavior of thieves; or than is the low rate of business failure evidence that the threat of failure has no effect on businessmen. The question is what would slave imports have been without the threat of seizure? It would have been *much* larger than 2 million.
- b. The answer depends, then, on the effect of the threat of seizure and loss of cargo on the slave trader’s costs, that is, on the supply curve of slaves transported to the New World. Without the British policy the S curve would be lower on two counts: (1) the 0.08 probability that the cargo would be lost (confiscated) would be eliminated, plus any allowance for higher variability (as distinct from a lower average) of yields of trips; (2) the costs actually incurred to evade capture (using smaller, faster ships; spending as little time as possible picking up a cargo) would be eliminated (without these evasive measures the probability of confiscation would have been higher than 0.08).
4. The Bureau’s subsidy—certain and transferrable as it is (bankers in the West will make loans with the Bureau’s grazing rights as collateral)—has been capitalized into the value of the land. When the land changes hands it does so at the fancy price warranted by the Bureau’s subsidy, and the *second* owner will earn only a normal rate of return on *his* investment.

Exercises for Section 14.2

1. a. It's the diamond case. The expense of the product is irrelevant.
- b. It's the cab case. The file clerks earn what file clerks normally earn, since there are lots of them, not specialized in furriers. They get no part of the rent.
- c. It's the If-you're-so-smart case. If the economist saying this were so smart, she would be rich. And why would she tell you?
3. True. The threat of entry by potential competitors provides an incentive for firms to produce safe products. Entry and exit is a substitute for government regulation.

Problems for Section 14.2

1. a. Large firms buy the innovation because they save their large output multiplied by the saving on average cost if they buy it. For the same cost of buying it, however, the small firms save only their *small* output multiplied by the saving.
- b. There must be obstacles to entering the industry at a large scale. If the initial situation was one of an equilibrium in which the small firms were just barely profitable, an innovation biased in this way toward helping large firms would make small firms unprofitable, and drive them all out if entry to large firms was free. If, however, the small firms were initially earning rents ("more profitable") this would not be necessary: they would suffer a decline in rents, but would stay in the industry until the rents were exhausted.
- c. If the innovation could be rented, *any* scale of firm could take advantage of it, at the fixed rental (cf. rental trucks, computer time sharing). You have to assume that there are transactions costs that make renting more expensive than owning. It would be very expensive for a brewer, for example, to hire rather than own his own malting vats.
- d. Economies of scale at the level of the firm are ignored essentially because we assume that firms will expand until the economies are exhausted. David's argument is one of several that attempts to bring economies of scale at the level of the firm back to life.
4. The consumer groups are asking that grocers bear the cost of unanticipated inflation. The grocers will try to avoid this by stocking their shelves only at the last minute (with the high price), understocking their shelves relative to the present position. If they aren't able to do this they will cut back on their provision of services, raising their prices until the returns on their capital are again normal. So, true.
8. If he and others made more than normal returns he and others would be willing to pay the original owner more, which would itself drive up the price until the license was no longer a bargain. So, true.

CHAPTER 15

Exercises

1. a. Thus

$$Q = 20 - 2P$$

$$Q = 2P$$

Eliminate Q by setting the two equal:

$$2P = 20 - 2P$$

Collect terms:

$$4P = 20$$

Solve for P :

$$P = \frac{20}{4} = \$5 \text{ thousand}$$

So the quantity, using $Q = 2P$, is 10 million.

The diagram is easy. The supply curve bisects the demand curve (note that in the algebra both supply and demand have the same slopes, though of opposite sign).

b. The tax moves the supply curve up, because new suppliers must get the money to pay the tax as well as to pay their costs. To be exact, it moves it up in parallel fashion, by \$2 thousand at each output. Using the hint: The quantity distance between the old equilibrium and the intercept of demand is 10 million minus zero, or just 10 million, and one-fifth of this is 2 million. So quantity falls 2 million, from 10 million to 8 million. Likewise, the demand price (here make sure to follow along on your diagram) will be higher by one-fifth of the price distance between the old equilibrium and the intercept, or one-fifth of the difference between \$10 thousand and \$5 thousand, or \$1 thousand. So the demand price rises from \$5 thousand to \$6 thousand. The supply price—different from the demand price because of the tax—is just the demand price minus the tax, or \$6 thousand minus \$2 thousand = \$4 thousand.

$$\text{c. Demand: } P = \frac{20}{2} - \frac{1}{2} Q$$

$$\text{Supply: } P = \frac{1}{2} Q$$

Set the two equal:

$$10 - \frac{1}{2} Q = \frac{1}{2} Q$$

So $Q = 10$ million. Substituting into the supply curve, $P = \frac{1}{2}(10) = \$5$ thousand, as before.

d. The new system is:

$$\text{Demand: } P = 10 - \frac{1}{2} Q$$

$$\text{Supply: } P = \frac{1}{2} Q + 2$$

Solving as usual:

$$10 - \frac{1}{2} Q = \frac{1}{2} Q + 2$$

Or, $18 = Q$, the same result as the geometry.

Using $Q = 8$ to solve for P gives $P = 10 - \frac{1}{2}(8) = \6 thousand, again the same as the geometry.

e. The tax is \$2 thousand on 8 million automobiles, so the revenue from it is \$16 thousand million (\$16 billion).

f. Viewed from the side (to get a base and altitude easy to measure), the triangle has a base of \$2 thousand and an altitude of 2 million automobiles (10 million, the quantity at the old equilibrium—which is the apex of the triangle—minus 8 million, the new equilibrium, which is the base. So the area, half the altitude times the base is $\frac{1}{2}(\$2\text{ thousand})(2\text{ million}) = \2 billion .

g. Total expenditure on automobiles is their new, after-tax price of \$6 thousand multiplied by their new quantity, 8 million. Therefore, it is \$48 million, and the social loss is only $\frac{1}{24}$ of the expenditure, less than 4%. It will not be surprising, then, to find that triangles of social loss from taxes and other distortions amount to pretty small shares of national income.

Problems

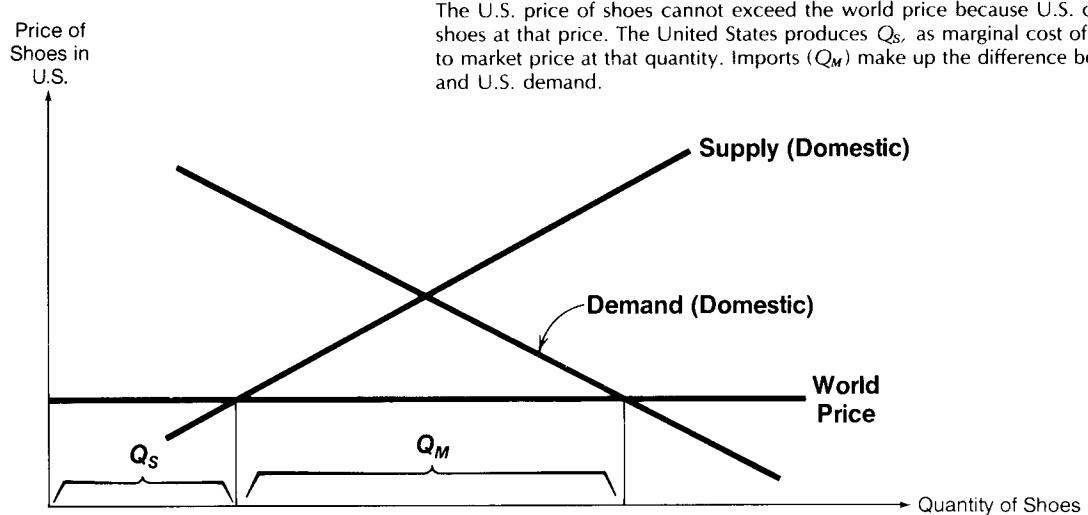
1. a. Nothing to it: it is simply a straight horizontal line, a supply of shoes from foreigners that is perfectly elastic at the world price. Notice that if the U.S. exports rather than imports shoes the same line is the demand curve. The United States can buy or sell any amount of shoes at an unchanged world price, a condition known in this context as the “small country assumption.”

b. See Figure A15.1. The world price of shoes determines (that is, is equal to) the price of shoes in the United States. The quantity of imports is Q_M (the usual sign for imports is “M” and for exports “X,” leaving I and E for other jobs). The quantity demanded by Americans is equal to the quantity supplied by American firms (Q_S) plus the quantity imported.

c. Look at Figure A15.2. Notice that if the tariff was high enough to bring the domestic price to point Z or above it would be prohibitive and imports would cease. The foreign supply curve of shoes rises by the extent of the tariff, up to World Price + Tariff. Naturally, the American price rises to meet the world price plus tariff. The American quantity supplied rises (encouraged by the higher price) by the amount ΔQ_S . American demand falls by ΔQ_D . The imports fall therefore by $\Delta Q_S + \Delta Q_D$.

d. The increase in producers’ surplus is A . The loss in consumers’ surplus is $A + B + D + C$, from which one might mistakenly infer that $(A + B + D + C) - A = B + C + D$ was the social

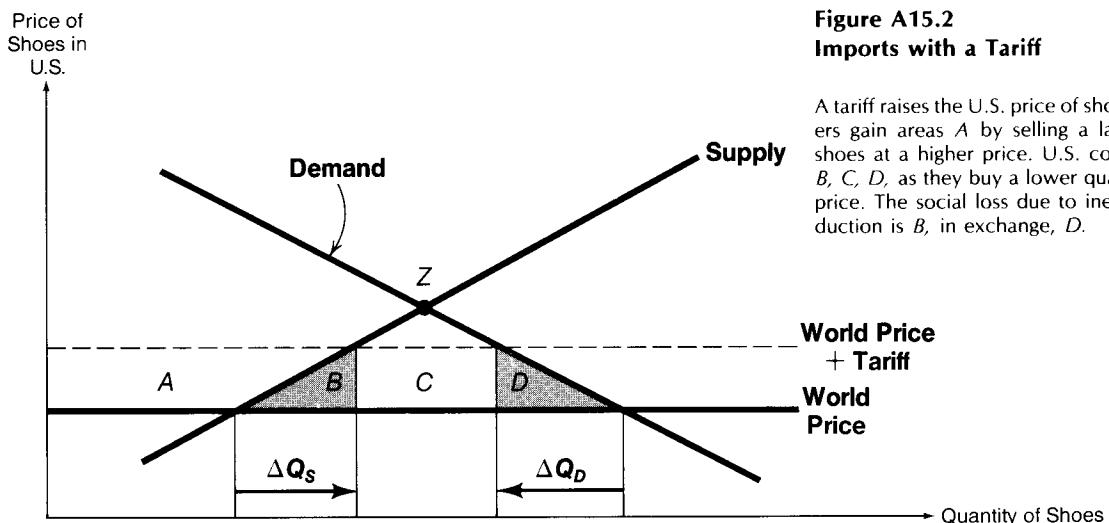
Figure A15.1
Imports Without a Tariff



loss. But area C is not a net social loss to Americans, at least not on the usual assumptions as to how tax revenues are treated; it is a mere transfer from importers to general taxpayers, that is, from one citizen to another. The net loss is $[A + B + C + D] - [A] - [C] = B + D$, the shaded triangles in the diagram. The import tariff has reduced national income by the amount $B + D$.

5. a. Since Britain is described in the question as a small country, the world price is given and the usual diagram applies, given in Figure A15.3. As before, the government revenue is the area $bgef$, imports are reduced to the length bg , rents to British producers are raised to area $abco$, and consumption of autos is reduced to the length ag .

b. The simplest way to see the equivalence is to reflect that the £100 tax is imposed on British



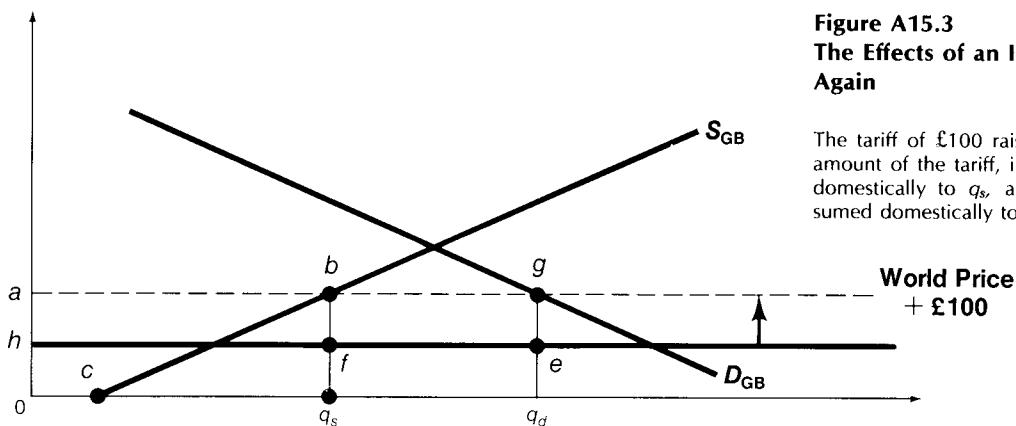


Figure A15.3
The Effects of an Import Tariff Once Again

The tariff of £100 raises domestic price by the amount of the tariff, increases quantity supplied domestically to q_s , and reduces quantity consumed domestically to q_d .

and foreign cars consumed in Britain, but the tax on British cars is offset by the £100 subsidy; that is the same as charging £100 on every foreign car consumed, that is, having a £100 import tariff. Diagrammatically speaking, the tax drives down the British demand curve, and the subsidy drives down the British supply curve, as in Figure A15.4. The effects are equivalent to the tariff. For example, the revenue to the government from the £100 tax on consumption is the area $ageh$, but the area $abfh$ goes back to producers as a subsidy, leaving a net additional revenue of $bgef$, exactly the same as the import tariff. Again, the rents of British producers of autos are raised to $abco$, just as under the tariff, because the total value of sales including subsidy is $abq_s o$, to gain which the manufacturers spend $c b q_s$ in costs.

c. Evidently, "free trade" does not consist only of the abolition of explicit tariffs on imports. The European Common Market does not have free trade in automobiles.

7. True, as can be seen by drawing an inelastic supply curve and the corresponding production possibility curve (namely, a rectangle), as in Figure A15.5. Notice that to avoid confusion here the tax must be viewed as being imposed on the demand curve. The supply price will now bear all the adjustment of the tax, falling until it just induces the demanders to demand the unchanged quantity of Bread supplied despite the tax. Although the tax might be viewed as hurting bakers and wheat growers (not bread eaters, who pay less per loaf by the amount of the tax), it does not change the amounts of Bread and Other Goods consumed. The tax has no effects on efficiency.

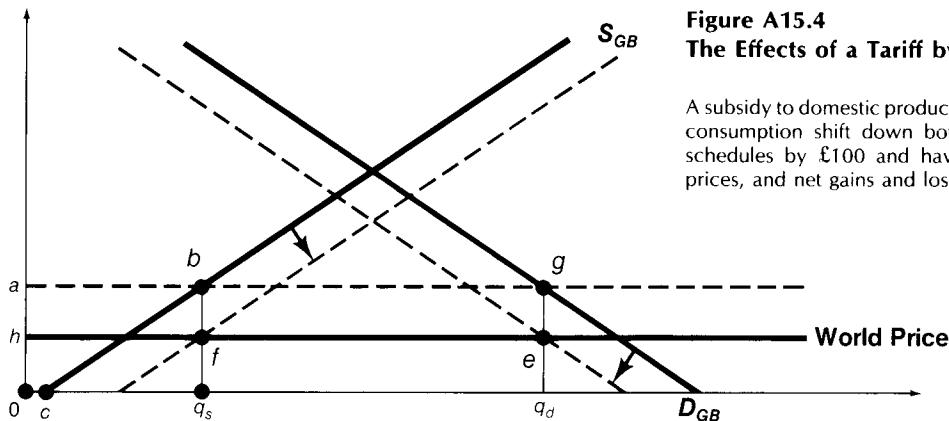
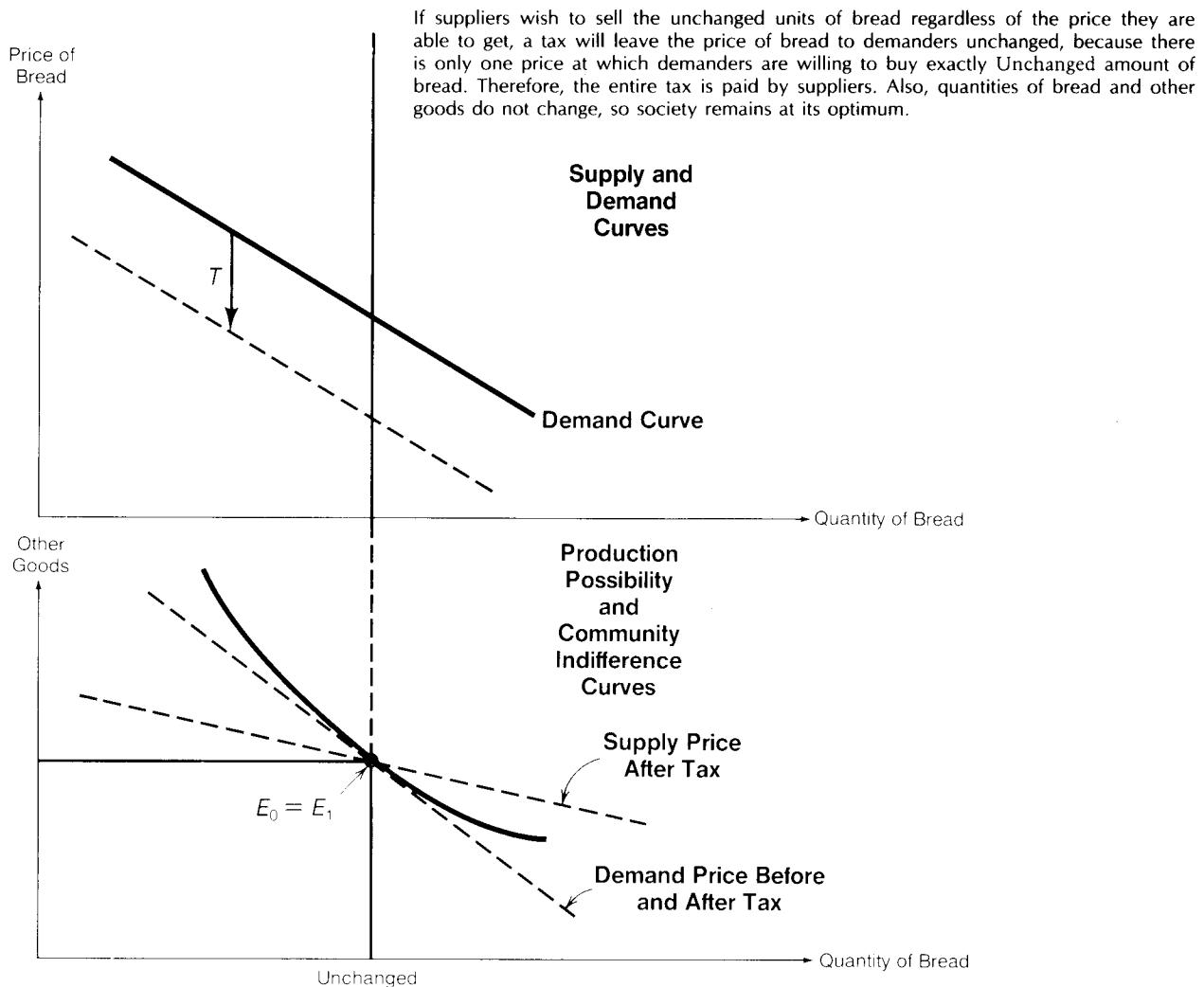


Figure A15.4
The Effects of a Tariff by Any Other Name

A subsidy to domestic production of £100 and a tax on domestic consumption shift down both domestic supply and demand schedules by £100 and have identical effects on quantities, prices, and net gains and losses as does a tariff of £100.

Figure A15.5
Taxing Inelastically Supplied Goods Has No Social Cost



CHAPTER 16

Exercises for Section 16.1

1. a. At \$50 a ton, the Q is $= 50 - \frac{1}{2}(50) = 25$ million tons.
- b. At only 12.5 million tons imported, using the quantity-dependent version of the demand curve, the price would be $P = \$100 - 2Q = \$100 - 2(12.5) = \$75$ a ton. A tariff of \$25 a ton (\$75 minus the world price, \$50) would achieve the same result.
- c. The profit is the revenue from the equivalent tariff, $(\$25)(12.5 \text{ million}) = \312.5 million. The present holders of licenses would be willing to use the entire sum—or the entire sum less one dollar—to protect it.

Problems for Section 16.1

3. Each of the 6 gallons cost 30 cents in waiting time plus 50 cents in cash. In other words, a rise in price from 50 to 80 cents a gallon sufficed to reduce the quantity demanded by 25 percent (if it had not sufficed the queues would have lengthened until it did). Evidently the elasticity was not zero. Depending on which of the two points along the demand curve is taken as the base in the arithmetic, then, the elasticity is $(25/100)/(30/50) = 0.42$ or $(25/75)/(30/80) = 0.88$, for an average of $(0.42 + 0.88)/2 = 0.65$. More elegantly, consider a constant elasticity demand curve:

$$Q_0 = DP_0^{-\epsilon} \quad \text{and} \quad Q_1 = DP_1^{-\epsilon}$$

Inserting the two observed pairs of Q and P (let $Q_0 = 100$) and dividing one equation by the other yields:

$$\frac{100}{75} = \frac{D(50)^{-\epsilon}}{D(80)^{-\epsilon}}$$

from which, taking logarithms and solving for ϵ , one finds that $\epsilon = 0.61$. An elasticity of 0.61 or 0.65 is not the same as an elasticity of zero. Don't believe everything you read in the newspapers.

5. True. The excess willingness-to-pay created by the low price must somehow be choked off. One way to choke it off is to let people stand in sweltering lines for 6 hours. By contrast, if the museum gives them queue tickets the last person to join the line will not experience as much pain and suffering as he would without the queue tickets. In consequence, more people will join the line, until the last person to join is just barely satisfied with the deal. The result is that the total inconvenience will be the same as before. Instead of standing in line the people will be able to tour the city, but the touring must itself be tiring and boring enough to exactly equal in inconvenience the time saved from standing in line. To put it another way, the number of slots passing by the exhibit cases is a scarcity that must somehow be allocated. The full price (money plus inconvenience) must rise to equal the marginal willingness-to-pay at that quantity.

6. Since there was no evidence of excess demand, the government regulations were apparently not holding down the price. Therefore the removal of the regulations would have no effect. It is possible, indeed, that the regulations merely caused a misallocation of the oil supply, raising rather than lowering prices. There may well be some line of reasoning that favors the more conventional assumption, but it is not simply a matter of the government sitting on oil prices.

8. a. True. The price of coupons will be equal to the difference between the money price and the marginal valuation consumers put on gasoline when its supply is 100 gallons, for coupons bestow the right to buy a gallon, and are therefore worth what the marginal consumer will give for them. Queues will not develop, because the money price plus the value of the coupon will discourage demand to exactly the amount 100 gallons: nothing is gained by standing in line; time wasted in lines is not necessary to ration the gasoline.

b. True. Suppose the government issues coupons for only 50 gallons. Then only 50 can be sold, for a coupon is necessary to buy a gallon. Suppliers will find that they can't sell the 50 additional gallons they would be willing to offer at the controlled price. So they offer less. So the supply price necessary to get them to supply the smaller amount falls. Meanwhile, with a smaller amount purchasable, consumers value the marginal gallon more. The marginal valuation increases, that is, consumers move up their demand curves. To maintain the equality,

$$P_{\text{money}} + P_{\text{coupon}} = P_{\text{marginal valuation}}$$

the price of coupons must rise to the difference between the supply price (at the lower, 50-gallon supply) and the marginal valuation.

c. True. Issuing enormous numbers of coupons drives their price to zero. That is, they are not a constant constraint on gasoline purchases. In consequence, they no longer ration the available quantity, and time must do it. Answer the following question: what happens when the issue of coupons is only a little above the current amount? Coupons are evidently not literally valueless in this case; but they do not make it *certain* that one gets a gallon, for there are more coupons than gallons.

Exercises for Section 16.2

1. a. When the peace of the night air is owned by no one it may be broken by anyone.
- b. Were the elephants owned, their owners would protect them for their ivory, or their grandeur.
- c. The pasture itself could become owned, no longer common; or the right to graze could become a piece of property to be bought and sold, limited in total to the amount that permitted grass to regrow. Both were in fact adopted in English villages: so much for hypothetical history learned at one's mother's knee.
- d. Were the Channel owned it would be allocated by price, like a toll on the Dan Ryan Freeway. Now, like the Dan Ryan, it is a freeway, and has the congestion of one.
- e. Properly, perhaps, no one owns the property right to their view. Without such a right, if some architect crazed by a passion for flat roofs and wide glass is given the commission to wreck the view up Park Avenue, he may with impunity.
- f. Were places in the line bought and sold, it would help. Still better, Easton could (and should) charge more.

Problems for Section 16.2

2. a. The diagram is the usual one of allocation to two demands or two supplies, such as students to two campuses or labor to two employments: the whole quantity of oranges available is put on the horizontal axis, and any point along the axis represents an allocation of the oranges between two places. An allocation far to the left, where the Florida demand curve starts its downward descent, represents shipping most of the oranges out of Florida (leaving the Floridians with a high marginal valuation of oranges). An allocation far to the right, where the New York demand curve (turned to face the opposite way) begins its descent, represents keeping most of the oranges in Florida, making them high priced in New York. The oranges will actually move until the gap between the high New York price and the low Florida price just barely equals the cost of transportation. Since the two prices are not pushed all the way to equality, there is a triangle of "mutually advantageous exchange" that is "lost" because "too few" oranges get to New York. But the quotation marks emphasize that it is a loss only from the imaginary perspective of *no* transport costs. It would not be desirable to subsidize orange transport to bring the market to the point of equal prices, for that would entail from the social point of view the using up of scarce resources in transport to an extent in excess of value (their value being the price differential).
- b. The analogy is obvious. It is all transport, all costly.
- c. The assumption is that the government can perform the "relocation" of the property rights or capital or oranges or whatever perfectly costlessly. An ideal government bureaucracy is being compared with an actual market, a common error (in both directions: ideal capitalism is often compared favorably with actual socialism). Market failure is not a very useful terminology, though widespread.
3. True. The point is simply that these nice things are not owned by anyone. Therefore they will be overused; that is, California will be overpopulated with people moving there to take advantage of them.

CHAPTER 17

Exercises for Section 17.1

1. a. Total Revenue (Price Times Quantity)	Marginal Revenue (Change in Total Revenue)
\$599,794	
599,898.5	+\$104.50
600,000	+ 101.50
600,098.5	+ 98.50
600,194	+ 95.50

The marginal revenue is positive, though of course declining.

b. Total Revenue	Marginal Revenue
\$600,392	
600,198	-\$194
600,000	- 198
599,798	- 202
599,592	- 206

Marginal revenue is negative (and again declining, that is, getting more negative).
 c. Close enough, total revenue is constant: output rises in the same proportion as price falls.

Crossings	Total Revenue	Marginal Revenue
98,039	\$100,000	
99,010	100,000	0
100,000	100,000	0
101,010	100,000	0
102,041	100,000	0

Therefore, marginal revenue is zero.

d. Total Revenue	Marginal Revenue
\$149,994	
149,997	\$3
150,000	3
150,003	3
150,006	3

Clearly, Austen is a price taker, and therefore his marginal revenue is his average revenue, that is, the going price.

e. Total Revenue (millions)	Marginal Revenue
\$149,898	
149,949.5	\$51.5 millions/100 millions = \$0.515 per bushel
150,000	50.5/100 = 0.505
150,049.5	49.5/100 = 0.495
150,098	48.5/100 = 0.485

Marginal revenue is positive at around 50 cents a bushel.

3. His marginal revenue is negative. Evidently, things have gone too far: No matter what his marginal cost is (as long as it is positive) his marginal revenue is less than it. He loses money on every additional exhibition. So he wants to cut back, to allow a (much) smaller number of exhibitions.

5. No, because he can't get a higher price. He's too small a part of the market, as is apparent in the fixed price he gets regardless of whether his output is 49,800 bushels a year or 50,200. Price takers have no incentive to allow for their effect on price, because they don't have any.

Problems for Section 17.1

1. Most of the answer is high school algebra. Two points on the curve are 219 units of money (relative to 100 as the world price) and 5 million pounds and 112.5 units of money and 16 million pounds. The demand curve, if straight, takes the general form $Q = D - bP$. These points imply that the slope, b , was $(16 - 5)/(219 - 112.5) = 0.1033$. One must keep this many digits to keep the rounding error small. Taking either pair of numbers yields the intercept, D , by $D = Q + bP = 16 + (0.10)(112.5) = 27.62$. The demand curve in its usual form is therefore $Q = 27.62 - 0.1033 P$. Check by inserting one Q, P pair and showing consistency: does 16 equal $27.62 - 0.1033(112.5)$? Yes. The marginal revenue can be written down directly after one has the demand curve expressed as price depending on quantity, instead of the usual form. From the usual form one can deduce that $P = 27.62/0.1033 - 1/0.1033Q = 267.4 - 9.682Q$. The orders of magnitude make sense. The intercept on the price axis, for example, could easily be 267.4 units of money, since the price at a low quantity is as high as 219. By the geometry of straight-line demand curves, the marginal revenue has the same price intercept (267.4) and twice the (negative) slope. It is therefore $MR = 267.4 - 2(9.682)Q = 267.4 - 19.364Q$. The world price is 100 units of money. The question is, what is the quantity that brings marginal revenue down to the world price? The answer is the Q such that $267.4 - 19.364Q = (100 - 267.4) (-19.364) = 13.24$ million pounds. Returning to the demand curve, 13.24 million pounds corresponds to a price after tax of $P = 267.4 - 9.682(13.24) = 139$. The revenue-maximizing tariff is 39%.

Check: Does a price of 139 units correspond to an output of 13.24 million pounds? Yes: $27.62 = 0.1033(139) = 13.26$, close enough to be explicable as rounding error. At the 1784 level of 119% the tariff was too high but at the 1785 level of 12.5% the tariff was too low to maximize government revenue.

COMMENT

The argument is precarious. The problem is the specification of a straight-line demand curve. One could with equal justice specify a constant-elasticity curve, in which case the revenue-maximizing tariff is different. In fact the elasticity under the constant-elasticity specification (that is, using the general form $Q = DP^{-\epsilon}$, or $\ln Q = \ln D - \epsilon \ln P$) is 1.746, which implies a tariff of 134%, *higher* than the 1784 level, not lower.

7. It will simplify your thinking to read the problem carefully, and notice that the problem is about the movement of the whole curve (not the intersection of the curve with some marginal cost curve) and that it specifies that the taxes are equal in dollar amount at the *pretax* output. For straight-line demand curves the diagram is Figure A17.1. The Unit Elastic Point, incidentally, is misdrawn (why?). Notice that to the left of $MR = 0$ (the Monopolist has some positive marginal cost) and to the right of Z (which must necessarily be to the left of Q_0) the after-tax marginal revenue for the specific (t cents) tax is always lower than the same for the ad valorem (t percent) tax. So, true.

Having worked through the simple case of straight-line demand curves one can now discard them and see through to the essential reason for the truth of the proposition. Suppose the monopolist increases output. His marginal revenue is lower on account of the tax (he gets the average revenue after tax). But with the ad valorem tax, the *amount of the tax falls* (for the price is lower), whereas it does *not* with the specific tax. So marginal revenue must be larger for the ad valorem tax than for the specific tax.

Exercises for Section 17.2

1. False. The point at which marginal cost equals marginal revenue is the point of maximum profits. The economist's model is merely another way of saying "Firms seek maximum profits." That businesspeople do not say it this way is not relevant.
3. Yes. A Ceiling above the Old Fare would not, but it would be ineffective: the Ceiling would be higher anyway than the monopolist wanted to charge. Any Ceiling below the Old Fare makes the old marginal revenue curve irrelevant: it is now the Ceiling that is the marginal revenue curve.

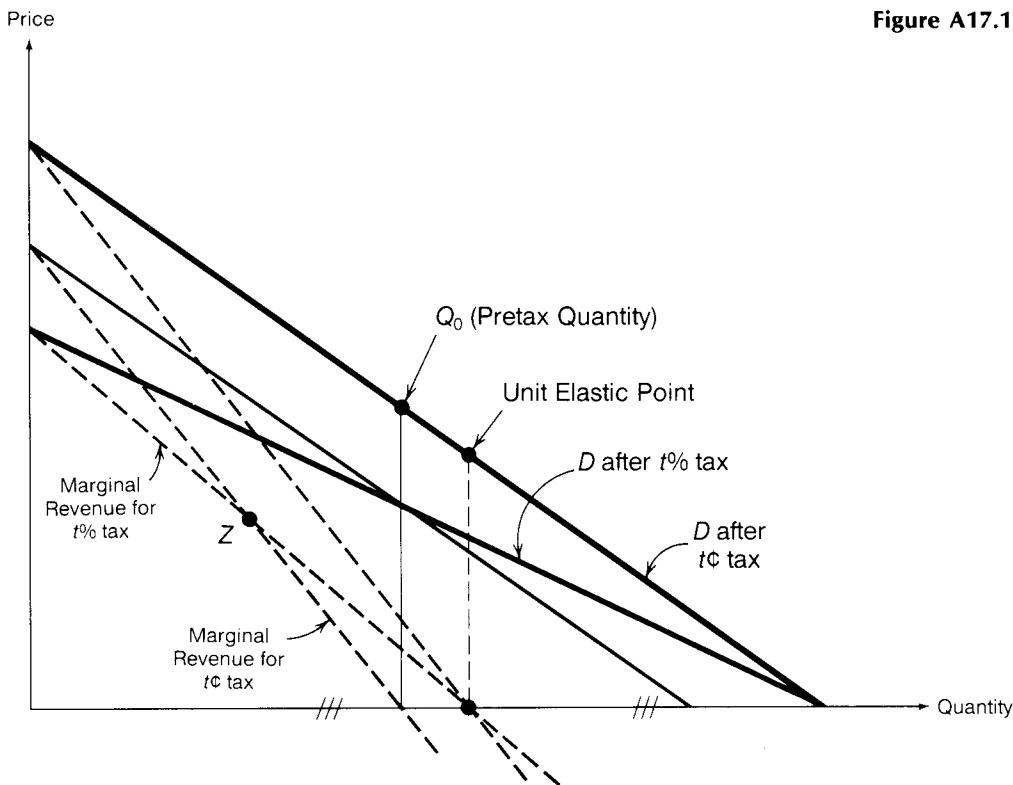


Figure A17.1

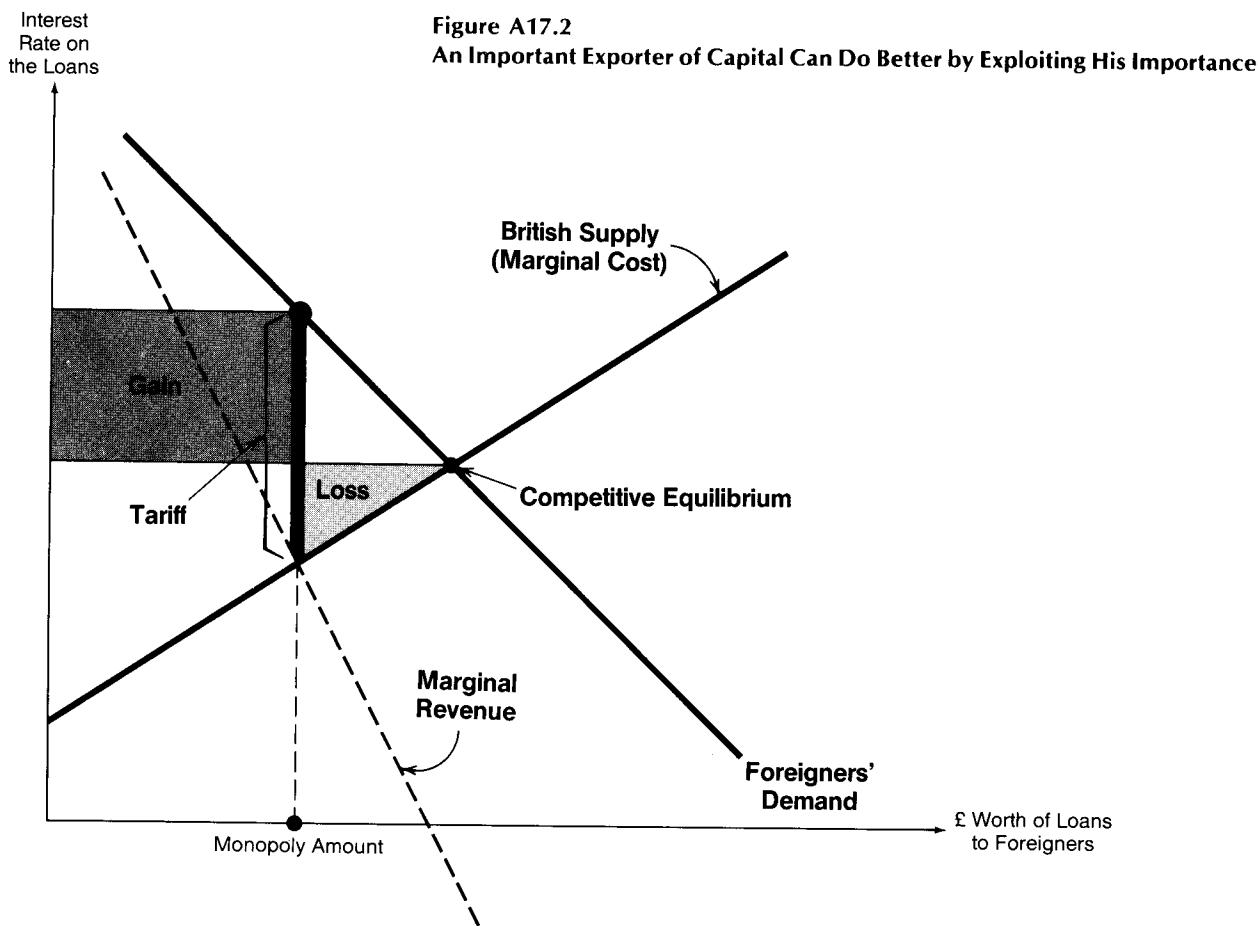
Problems for Section 17.2

1. a. The Competitive Equilibrium is the point at which the British supply—the sum of atomistic decisions by each British supplier—equals the foreigners' demand, as in Figure A17.2.

b. Britain was to some degree a monopolist, that is, facing a downward-sloping demand of its victims. By restricting the amount supplied to Monopoly Amount it could do better, Gain being larger than Loss of revenue over the marginal opportunity cost of the funds. The government could impose an export Tariff (marked so) on loans to bring about this optimal exploitation of foreigners. In actual fact Britain did nothing of the kind, allowing virtually free exportation of capital funds in 1900. Only when Britain's monopoly power in world capital markets was gone did Britain impose exchange controls limiting investment abroad.

4. True (roughly: see the Comment). The highest bidder would be the company that believed it had the lowest costs, and was willing therefore to charge the lowest fare. The field of competition is shifted from the trolley line itself (where only one firm can compete and there is no incentive to offer low prices) to acquiring the franchise itself (where many can compete and the rules of acquisition provide the incentive for low prices). French railways in the early nineteenth century were regulated this way, and the scheme would be applicable nowadays to cable TV or to gasoline stations on tollways. Since mere consumers are the beneficiaries of this simple scheme to solve the problem of natural monopoly, and not important aldermen, bureaucrats, and businessmen, it is no surprise that it is rarely proposed.

COMMENT The scheme shares a difficulty, incidentally, with direct regulation to assure average cost pricing. The difficulty is that at the truly optimal point, where the demand curve cuts the marginal cost curve, the monopoly earns more or less than the normal



return, depending on whether average cost including normal return is rising or falling at that output. If you sketch two diagrams, one with U-shaped average cost and another with forever falling average cost, you will see the problem. To put it another way, the bidding to charge the lowest price will induce companies to offer service along their average cost curves. But setting average cost equal to price is not socially optimal.

CHAPTER 18

Exercises

1. Most likely none will earn more than normal returns.

a. Harley is buying into a monopoly—the \$20,000 cost for a mere right to operate shows that liquor stores have been made artificially scarce. But the seller has no reason to charge him less than every last cent the monopoly is expected to earn. Harley just makes what one makes in any competitive investment. He runs the liquor store as a monopoly, but all the advantage goes to the original owner. The profit is hidden in the price.

b. Easton is in a similar position. The sellers of the stock sell it at its full, monopoly value. Easton gets none of the net benefits: he *pays* for the benefit, which is again reflected (or hidden) in the price.

c. If the training is lengthy and expensive enough, Lewis will dissipate all the gain to being an electrician: compare the dissipation of monopoly rent in medical education. The profits are hidden in the fees and opportunity cost of training.

3. a. In the European market, which is probably competitive so far as the Pittsburgh maker is concerned.

b. When the conspiracy collapses.

5. Use the Fundamental Equation of Monopoly:

$$1 - \frac{1}{2} = \frac{1}{2} = \frac{MC}{P}$$

Marginal cost would be half of price.

7. a. The senior citizens apparently have in the judgment of the theater owners a higher elasticity of demand. They therefore pay a lower price, to maximize the revenue from a showing of *Superman 12*.

b. Big families shop more carefully for detergent, since they use a lot of it and are therefore more likely to care about its price. So big families have a more elastic demand. Boxes aimed at big families therefore have a lower per-ounce price.

c. This is the luggage trick. Those who already own a television set have a higher elasticity of demand for a new one. Having them turn it in allows the seller to identify the high-elasticity buyers to whom one must charge a lower price.

Problems

3. The optimal ratio of price to marginal cost for a monopolist is $P/MC = \epsilon / (\epsilon - 1)$, where ϵ is the elasticity of demand. If this ratio is greater than the ratio of the old price to the new marginal cost, the monopolist sets the price a penny below the old price, since he cannot set it above. For example, if the elasticity, ϵ , is 2, then the optimal ratio is $2/(2 - 1) = 2$, and the royalty should be 100% of the post-invention cost. But if the pre-invention cost is in fact only 50% higher than the post-invention cost the patent owner obviously cannot charge a 100% royalty. He will charge a 49.999% royalty. The conclusion is that the cost saving has to be very large for the inventor *not* to rake off all the cost saving for himself. If the old method were only 10% more expensive than the new, for example, the elasticity of demand would have to be such that $1.10 = \epsilon / \epsilon - 1$, or $1.10 \epsilon - 1.10 = \epsilon$, or (using your very cheap new hand calculator) $\epsilon = 11$. It is a most unusual product whose quantity demanded increases by 110% or more when its price falls only 10%. But only for such a product would the owner of the patent take less than the whole social gain for himself. What would do the trick, of course, is a substitute patent in the hands of a competitive fringe.

4. a. Sony's marginal revenue curve slopes down at home but is flat abroad at the price abroad. It will never take up a position on its home curve below the price abroad. The entire marginal revenue is therefore the emphasized, kinked line in Figure A18.1.

b. The marginal cost curve is as drawn. Note the Equilibrium. The total revenue is the rectangular area out to Equilibrium under the kinked line. The total variable cost is the area out to Equilibrium under the Marginal Cost curve. The profit is therefore the shaded area.

c. By the first definition, yes, Sony dumps (that is, price discriminates). Sony may be typical in this connection, however: a thorough study of 76 American producers in 1940 found that 45 sold at export prices below domestic prices [Milton Gilbert, "A Sample Study of Differences Between Domestic and Export Pricing Policy of United States Corporations," U.S. Temporary National Economic Committee, *Export Prices and Export Cartels*, Monograph No. 6, Part I (Washington, 1940), Chapters 5 and 6]. By the second definition it does not, for it would be irrational for it to sell anything at a price less than marginal cost.

d. As long as the marginal cost curve goes through the minimum point of the average curve and as long as areas such as *A* equal areas such as *B*, the average cost curve can be anywhere. As drawn, for example, it is indeed above the Price Abroad. At any output the area under the Marginal Cost would equal the height of the Average (Variable) Cost multiplied by the output. At the Equilibrium output in particular the area of cost calculated this way is evidently below the area of total revenue; that is, Sony is making money. Only for a competitor is it irrational to sell at below average variable

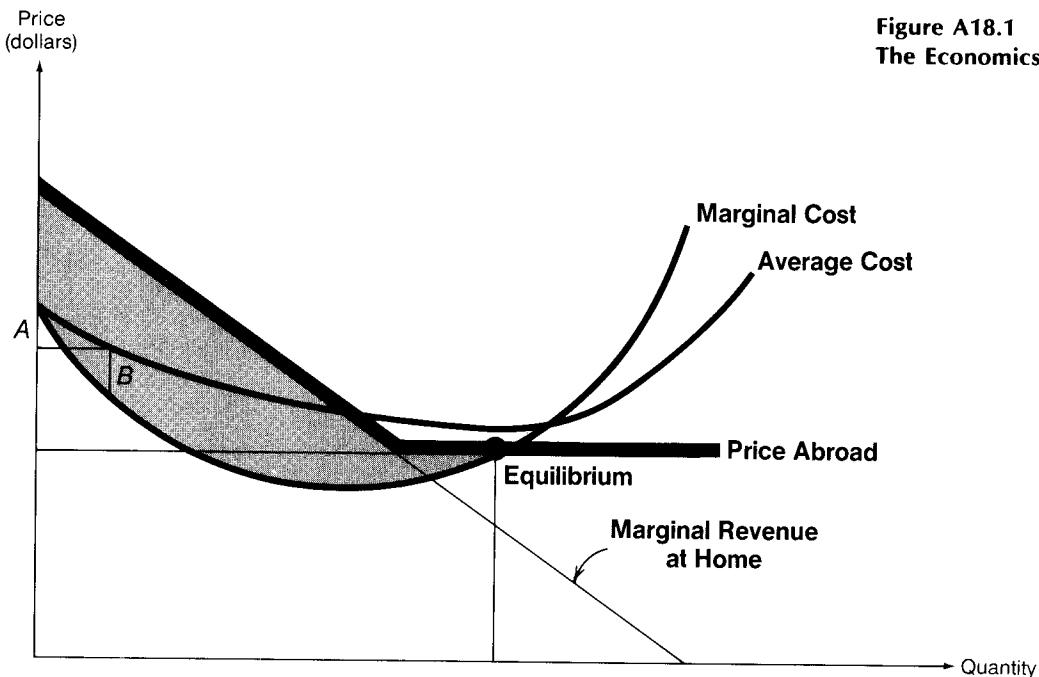


Figure A18.1
The Economics of Dumping

cost, because a competitor does not have, so to speak, the “earlier” area under the declining marginal revenue at home.

12. False, unless he has no marginal costs. He will charge the price that brings marginal revenue equal to marginal cost, that is, that maximizes receipts *net of costs* (profits).

14. Double true. In Figure A18.2, the quantity demanded at some P along D_2 is just some multiple, say δ , of the quantity along D_1 (and the same multiple for every price). So if a fall in price from P to P' of ΔP causes an increase of ΔQ_1 along D_1 it will cause an increase of $\delta \Delta Q_1$ along D_2 , because

$$\Delta Q_2 = Q'_2 - Q_2 = \delta Q'_1 - \delta Q_1 = \delta(Q'_1 - Q_1) = \delta \Delta Q_1$$

So the elasticities are

$$\frac{\Delta Q_1/Q_1}{\Delta P/P} \stackrel{?}{=} \frac{\Delta Q_2/Q_2}{\Delta P/P} = \frac{\delta \Delta Q_1/\delta Q_1}{\Delta P/P}$$

and the δ 's cancel, affirming the first equality. So it would be inadvisable to use such demand curves to determine how a discriminating monopolist sets prices, because at equal marginal cost of production an equality of marginal revenues in the two markets would result in equal prices [since $MR/P = (1 - [1/\epsilon])$, defined positive]. This wouldn't give a lot of insight into unequal prices. Such demand curves would be to the point if the consumers in the two markets were different only in numbers, not in character.

CHAPTER 19

Exercises

1. a. Allow competition among telephone companies but subsidize rural service directly. Still better, give the rural people money.
- b. Allow imports, subsidizing the farmers directly, or giving them money.
- c. Allow many different makers to make a certain type of tank, but hold each to its promise to

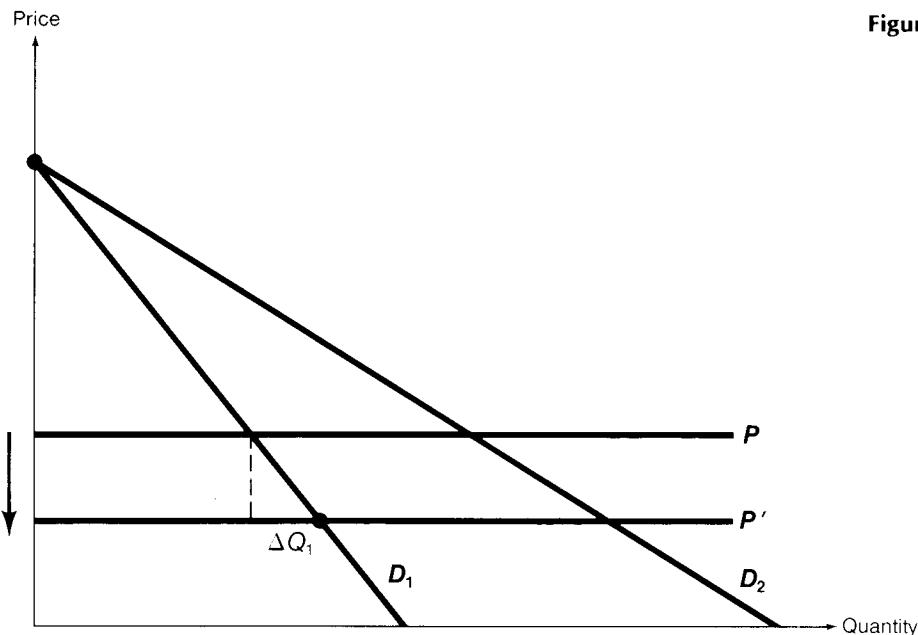


Figure A18.2

make it reliably (compare the same brand of Chevrolet automobiles coming out of many different factories).

3. a. $\frac{1}{2}(\$0.20 \text{ per ton})(-25 \text{ tons}) = \0.025
- b. Price times quantity, or $(\$1.0)(0.80) = \0.80 .
- c. $\$0.025/\$0.80 = 0.02$, or 2% of national income earned.
- d. Marginal net revenue of \$0.20 per ton on an output of 0.80 ton, or $(\$20)(0.80) = \0.16 .
- e. He would waste if necessary the whole of profits (conceived as profits above the normal rate of return, the normal rate being included in marginal cost): \$0.16.
- f. $\$0.16/\$0.80 = 0.20$, or 20% of income earned.
- g. In this case monopoly protection might cost ten times monopoly inefficiency: 20% is 10 times 2%.

Problems

1. True. The point is that this defense of the monopoly profits forgets that if the major league owners did not pay for the training provided players in the minor leagues, the players themselves would pay for it, as do, say, professional golfers or tennis players.

5. If the cartel of doctors were effective the output of medical services would be at Q_0 in Figure A19.1 where profit is maximized (and equal to $E + F$). Compared with the competitive output Q_1 , the lost output is $Q_1 - Q_0$. Its opportunity cost is $C + D$ (the value of resources in alternative employment required to produce the original output) and the willingness of consumers to pay for the additional output, namely, $A + B + C + D$. So the social loss is the difference between these two, namely, $A + B$ (although it might be noted that this argument ignores the resources used in blocking entry to the medical profession, such as the teachers of organic chemistry teaching general practitioners more organic chemistry than they need to know).

The obvious social policy is to break up the cartel, making doctors behave competitively. Doctors would lose the profit E and gain the profit B (profits being the rents to whatever special tastes or skills make for better doctors), so their net loss is $E - B$. Indeed, if the cartel was busted by allowing new entrants, the old doctors would lose E with no offset.

An alternative would retain the cartel, but subsidize doctors by the crosshatched line per unit of

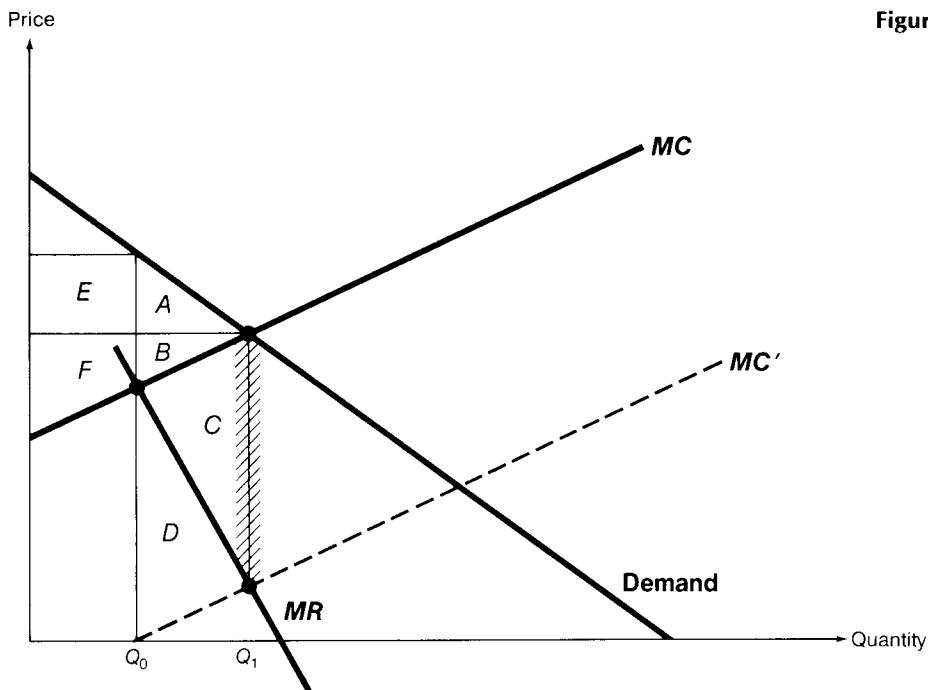


Figure A19.1

medical service produced. This would make the MC facing the cartel MC' , but leave the true social opportunity cost (the subsidy being a mere transfer) at the old MC . It would yield the competitive output (and optimal allocation, if not optimal income distribution). Notice that *any* subsidy would push the cartel *closer* to the competitive output (unless it pushed too far).

One further note: the money to pay the subsidy has to come from somewhere, that is, from a tax. This tax has itself distortion effects elsewhere in the economy (if anything, commodities or leisure, are taxed, and in equilibrium their price will not equal their true MC). We have a second-best piled on a second-best, and strictly speaking it is difficult to say anything at all about the effect of the policy. Note also that the policy does nothing about the resources wasted to blockade entry.

8. a. The inventor clearly has to monitor the output of the licensee after the deal to get his just reward. Otherwise the licensee will use the patent on the sly without paying the fee per unit of use. If he takes his just reward as a lump sum, by contrast, he and the licensee can part company, amicably and cheaply, as soon as the deal is made. No spying, no police, no lawyers need be involved. So the transactions costs are higher in the fee per unit case. Furthermore, charging a fee per unit faces the licensee with a marginal cost for something (namely, knowledge of how to make transistors or stainless blades or Pringle's Newfangled Potato Chips) that does not in fact have a marginal cost. Society sacrifices no output elsewhere in addition to the costs of the ingredients by using once again the knowledge of how to make Pringle's: the knowledge (like television signals or national defense) is not scarce once it is produced. So the fee per unit leads (relative to the lump sum) to an underuse of the invention. On both counts, then, the fee per unit is inefficient.

b. He might want to set the fee per unit to meter (literally, "measure") the demand for the invention. He is a monopolist searching out his demand curve. Monopolists can merely demand a single lump sum; but they might over- or underestimate the position of the demand curve, and thereby reduce their profits.

c. It's then quite clear why novel product inventions would have fee per unit licensing: their novelty and their dependence on the vagaries of consumer taste make their demand curves obscure (how much would *you* have paid in 1970 for world rights to the idea of a pet rock?). The fee per unit

allows the inventory to ride the demand curve out on its unknown path (to use an inexact metaphor). Routine process inventions, on the other hand, reduce costs by known amounts. A slightly better alloy in sheet steel or a method of making a purer chemical have easily determinable values. And that's what they will sell for.

Notice that all this has something to do with the puzzle of why any monopolist (not just the holder of a patent) does not simply extract a profit- and efficiency-maximizing lump sum from consumers and have done with it. Instead (we assert) he often charges a high marginal price.

11. But the triangle is not the only loss. So the answer here is False. This information is enough to construct an estimate of the triangle, but does not give you a clue to the social cost of the excessive education of doctors and so forth that amount to costs of maintaining the monopoly.

CHAPTER 20

Exercises for Section 20.1

1. The marginal customer would be where a dashed line drawn from the price at Harvard Avenue (like the Harvard-Delivered Price lines) intersected the tent side. The quantity corresponding to this (measured horizontally, of course) is the quantity corresponding to the middle price chosen.

Exercises for Section 20.2

1. a. 50 each, because the delivered cost of political soap from the two stores is the same exactly at the midpoint.

b. The Democrats will get all the voters out to 12.5 yards below the 50-yard mark, that is, $50 + 12.5 = 62.5\%$ of the vote as against $100 - 62.5 = 37.5\%$ to the Republicans. You can show this by noting that the new Democratic starting point bisects its old share, which means it bisects the portion of the Republican line between 25 and 50% of the vote. This leaves the Republicans with $(25 + 50)/2 = 37.5\%$. It's easier on the brain to merely draw the figure accurately and measure the new share.

3. a. Look at the equation in the text: namely, willingness to pay for an acre = $Y(P - C - TD)$. With the data for eggs the equation is: $(2143 \text{ pounds per year}) [62 \text{ pence per pound} - 30 \text{ pence per pound} - (0.5 \text{ pence per pound mile}) (10 \text{ miles})] = 57,861 \text{ pence per year}$.

b. The wool grower, using the same equation, is willing to pay only $(120 \text{ pounds per year}) [230 \text{ pence per pound} - 100 \text{ pence per pound} - (0.1 \text{ pence per pound mile}) (10 \text{ miles})] = 14,280 \text{ pence per year}$. The egg farmer clearly gets the land.

c. At 100 miles, however, the wool grower is willing to pay: $120[230 - 100 - (0.1)(100)] = 14,400$, whereas the egg farmer is not willing to pay even a positive rent for the land: $(2143)[62 - 30 - (0.5)(100)] = -38,574$. He would need to be subsidized to raise eggs there. The wool grower gets the acre 100 miles from London.

d. By now the routine is familiar. Take the two equations with D , the distance from London, unknown, and set them equal to each other (to represent the equal bids):

$$2143[62 - 30 - (0.5)(D)] = 120[230 - 100 - (0.1)(D)]$$

Do the high school routine on this, to get $D = 50$ miles. At 50 miles the bid rent curve has a kink, wool growing begins, and egg farming stops.

Problems for Section 20.2

1. A simple case which makes the point is this: assume each consumer will buy at most one unit for at most \$5. Let costs of transport be \$20 per mile along a road of one mile. If both firms locate at the middle, each will reach only one-fourth of the customers on the road. A firm that moves slightly away from the center gains customers; the other firm's response to such a move would also be to move away. The profit-maximizing equilibrium in this case is for each firm to locate one quarter of a length from an endpoint. This is stable: Neither firm gains by moving in either direction. True.

5. See Figure A20.1. The W is the hourly wage. The critical assumption is that consumers are identical—all Seattlers are alike, and a Seattler is like a Buttean, and so forth.

8. The optimal location is determined by the method outlined in problem 6. For the situation discussed in the text, the optimal position is at the $\frac{1}{4}$ and $\frac{3}{4}$ marks on the road. The Hotelling point is of course the $\frac{1}{2}$ mark (with the two stores side by side). True.

CHAPTER 21

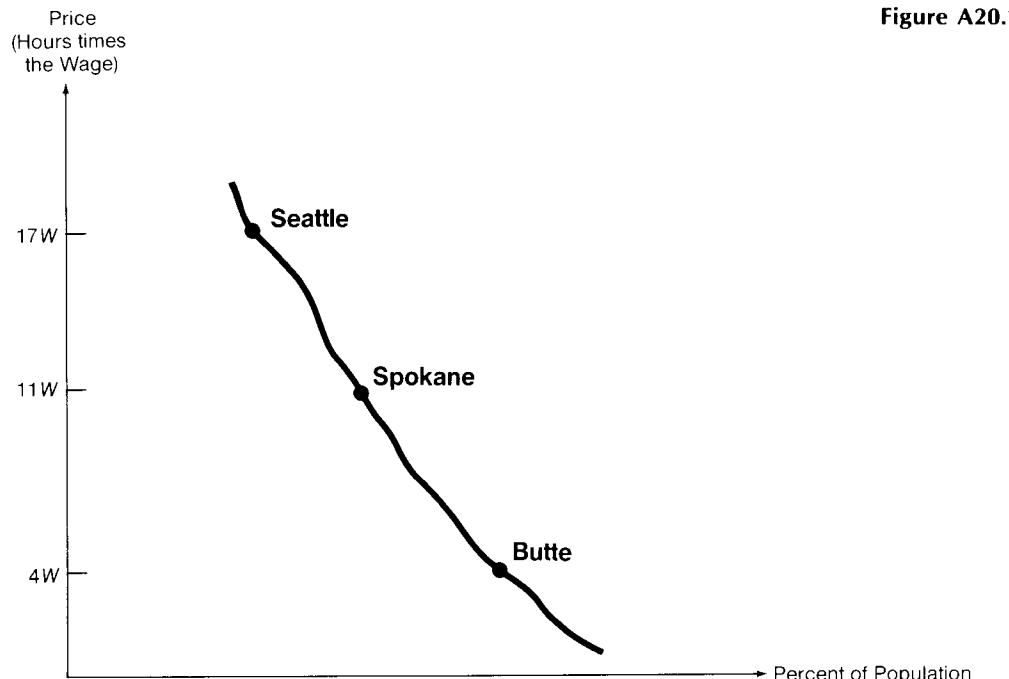
Exercises for Section 21.1

1. a. The monopoly price is \$5 per bottle. Since the monopoly has no costs, the midpoint of its demand curve is the point of maximum revenue—and therefore maximum profit (at the midpoint marginal revenue is zero, which is what marginal cost is by assumption). The intercept on the price axis is \$10 (since $P = \$10$ when Q is set equal to zero). So the price corresponding to the monopoly output is the midpoint between 0 and \$10, or \$5.

b. The profit-maximizing quantity is again the midpoint along the intercept, this time the quantity intercept. The intercept is 1000 bottles, so the monopoly output is 500 bottles. The revenue is therefore $(\$5)(500) = \2500 , which is the best that can be done with Mudavia.

c. $N = 2$ in this case, $\alpha = \$10$, and α/β (which is the quantity intercept) is 1000 bottles. So the industry output under duopoly is $[2/(2 + 1)](1000) = 667$ bottles. The output is larger than under monopoly, as one might expect. It is a third higher. The price must of course be lower, therefore. Using the (price form) of the demand equation, the market-clearing price is $P = \$10 - (0.01)(667) = \3.33 . The price is fully a third less than under monopoly. Total revenue is then $(\$3.33)(667) = \2220 . This is only about 11% less than under monopoly. Each duopolist gets half because the market divides randomly between them, or \$1110 each.

d. Alternatively, one could use the Cournot equation directly to get the monopoly figures, setting $N = 1$: $[1/(1 + 1)](1000) = 500$ bottles, as before.



Exercises for Section 21.2

1. "I don't think much of Poe's argument. Were it so simple there would be more good poker players. Indeed, someone who thought he was the victim of such psyching out could randomize his moves to avoid being psyched out: you can't predict a coin flip, no matter what the intelligence of the coin."

Exercises for Section 21.3

1. a. For instance, Death = $-\$10,000$, Free = $+\$10,000$, Light Sentence = $+\$5000$, Heavy Sentence = $-\$5000$. When presented with the first column (Morris Cooperates), McAlpin prefers Free ($\$10,000$) to Light Sentence ($\5000). When presented with the second column (Morris Defects), McAlpin prefers Heavy Sentence ($-\$5000$) to Death ($-\$10,000$).

b. The general relation is merely that b must be greater than a and d must be greater than c . If it is true for both players (looking at the situation from each player's point of view in turn), a prisoner's dilemma is the outcome. Each always defects.

3. a. Each trawler has an incentive to defect from a collective agreement to fish moderately. If the agreement could be enforced (in fact it is), then the cooperative solution is attained.

b. Each voter has no incentive to go to the polls, since his vote will have no discernible effect on the election, yet if all defect from their duty as citizens the Republic will fall. In Australia all must vote, or face a fine (compare Figure 21.10).

Problems for Section 21.3

3. The four have less to gain by acting in concert but a much lower cost of doing so (finding out who is cheating, taking reprisals against the cheaters) than the 10,000,000. A broad tendency for governmental actions to favor the few at the expense of the many can be inferred from this rudimentary piece of analysis. False.

4. Since Figure 21.7 in the text is merely the industry diagram with the scale changed, true.

CHAPTER 22**Problems for Section 22.1**

1. The subsidy from the local government ("from the rates" as it is put in English, that is, out of taxes) would increase the demand for labor. An employer who formerly would have been willing to pay 5 shillings per week to a worker would be willing to pay 6 shillings if the village gave him a shilling for every person employed. The 5 shillings was his profit from hiring the person before; the extra shilling is icing on the cake, making it still tastier. The demand curve, then, rises, forcing up the wages in the neighborhood as the intersection of the rising demand curve and the supply curve climbs up the supply curve. In other words, the subsidy accrues in part to laborers.

The subsidy results in a lessening of dependence on local farmers in one sense: now only the lower wages paid by the farmers flow directly from the farmers to the workers, the rest coming from the state. In another sense, the result is the opposite: workers now have more to gain from employment, relative to not being employed, and they cannot be employed without being hired by a farmer. The first proposition about the Old Poor Law interpreted as a wage subsidy, then, may or may not be true. But the second is false: the subsidy increases, not decreases, the amount of labor forthcoming. (This assumes, quite reasonably, that the supply curve of labor is upward sloping; see Chapter 24). (It is doubtful that the Old Poor Law was in fact a subsidy to wages in the strict sense. It is more likely to have been a subsidy to income, that is, a lump sum payment unrelated to the amount of work a person did. The effects of such a subsidy are the opposite; see Section 24.1.)

Exercises for Section 22.2

1. The total revenue product is merely \$200 times the total product. At 400,000, for instance, it is $(\$200) \times (20) = \4000 . The marginal product is the change over the next interval in total revenue divided by the number of hours used to produce it (100,000 until the next to last).

Hours per Year	Total Revenue Product per Year	Marginal Revenue Product per Hour (base of initial output)
400,000	\$ 4,000,000	\$14
500,000	5,400,000	26
600,000	8,000,000	42
700,000	12,200,000	36
800,000	15,800,000	30
899,999	18,799,985	15
900,000	18,800,000	12
999,999	19,999,990	10
1,000,000	20,000,000	8
1,100,000	20,800,000	

3. Lumberjacks earn (\$10 per hour) (1,000,000 hours) = \$10,000,000. The firm earns \$20,000,000—namely, the total revenue product. Nonlumberjacks earn \$10,000,000.

Problems for Section 22.2

1. If the marginal product of labor is 0, output would not have fallen at all. If labor's *MP* was less than the wage, output would have fallen by less than the wage (this relies on the fact that the change—5 million out of several hundred million—was small; see Problem 4.) If labor's *MP* equaled the wage, output would have fallen by the amount of the wages paid to workers killed. On this evidence, there was no surplus labor in India.

3. True. With rising marginal product (at any rate for a firm facing a given price for its product, so that larger output does not spoil the price), the demand curve for *X* slopes up. Firms will substitute toward *X* until nothing else is hired; indeed, until output is infinite. The situation is similar to that of a consumer with concave utility contours, with the difference that the firm can choose its budget constraint as well.

CHAPTER 23

Exercises for Section 23.1

1. a. Set the marginal revenue products equal to the corresponding factor price, giving two equations in two unknowns, *L* and *K*:

$$\$500 = 2000 - 25L + 1000K$$

$$\$50,000 = 100,000 - 150,000K + 1000L$$

b. Labor's marginal revenue product depends on *K*, which you must know already. The same holds for capital's marginal product—it depends on *L* as well as *K*. You must solve them, as Shlomowitz implicitly does, simultaneously.

c. Solving for *L* in the first equation, we have

$$L = \frac{500 - 1000K - 2000}{-25}$$

Substitute this into the second equation:

$$50,000 = 100,000 - 150,000K + 1000 \left(\frac{500 - 1000K - 2000}{-25} \right)$$

This results in *K* = 1, which is to say one sugar mill a year. Substituting back into the first equation gives 500 = 2000 - 25*L* + 1000, or 25*L* = 2500, or *L* = 100 man-years.

3. Only (b) and (c): use the method of pretending that L is identical to K .
 5. a. Pencils: $(100\%)(0.005) = \frac{1}{2}$ of 1%. Labor, which is much more important, $(100\%)(0.90) = 90\%$.
 b. Tubas: $(100\%)(0.095) = 9.5\%$ increase. Labor: $(-10.56)(0.90) = -9.5\%$ decrease. The cost is more sensitive to the price of the more important factor.

Problems for Section 23.1

2. $Q^* = s_E E^*$, where Q^* is the percentage difference in output, s_E is the share in costs of entrepreneurship (namely, 10%), and E^* is the percentage difference in entrepreneurship. If German entrepreneurs were a third more effective the consequent difference in output would be only $s_E E^* = (0.10)(0.33) = 3.3\%$.
 6. Labor's share is constant and equal to the elasticity of output with respect to labor. Therefore, true.
 7. There is no reason to believe that universities are profit maximizers or that they have constant returns to scale. If either condition fails, the adding up property does not hold. Therefore, true: the Fundamental Theorem does not hold.

Exercises for Section 23.2

1. a.	Output per Man-Day	Output per Acre
Stultior	10	33.33
Sapientior	9.09	50.00

b. Since $Q = AL^{0.5} T^{0.5}$ for both, for Stultior $1,000,000 = A(100,000)^{0.5}(30,000)^{0.5}$, or $1,000,000 = A(316.23)(173.21)$, which implies that $A = 18.26$. For Sapientior, $2,000,000 = A(220,000)^{0.5}(40,000)^{0.5}$, which implies similarly that $A = 21.32$. Sapientior is about 17% wiser than Stultior.

3. The rate is $2.06\% - (1.5)(1\%) + (0.5)(0) = 1.56$ per year. Multiplying this (namely, 1.0156) by itself ten times, or using some other method, gives 1.16742, or the "about 17%" of Exercise 1. Stultior will catch up.

Problems for Section 23.2

1. The rate of productivity change is the rate of growth of output (from the ratio 2040/1000 over 30 years, 2.376% per year) minus the rate of growth of inputs. The rate of growth of inputs are the share weighted sum of their individual rates of growth:

Input	Share	Rate of Growth (%)	Share-Weighted Rate (%)
Land	0.04	0	0
Capital	0.44	1.443	0.635
Labor	0.52	1.021	0.531
Sum	1.00		1.166

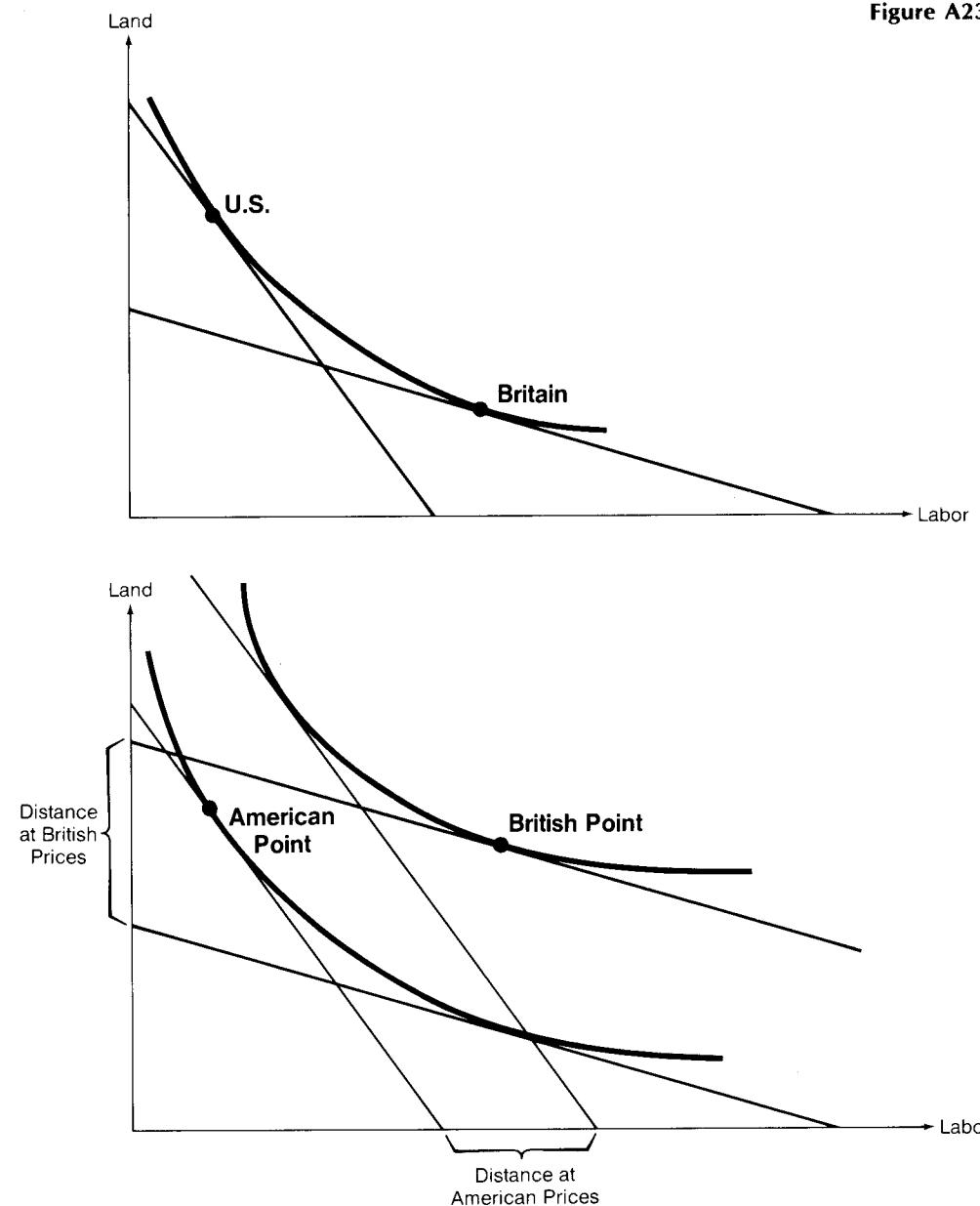
British productivity change was about $2.376\% - 1.166\% = 1.21\%$ per year, as against American of about 1.5% per year over the same period. The difference would cumulate to a difference in levels of productivity of only about 9% by 1900, even if (as seems unlikely) American productivity was as high as British in 1870. See D. N. McCloskey, "Did Victorian Britain Fail?" *Economic History Review*,

2nd ser. 23 (December 1970), reprinted in McCloskey, *Enterprise and Trade in Victorian Britain* (London: Allen & Unwin, 1981).

4. See Figure A23.1, top half. Neither point is revealed preferred to the other; that is why the productivity issue can only be resolved by knowing the unit isoquants.

5. See the bottom half of Figure A23.1. The ambiguity is the same as the ambiguity in measuring changes in income. The Americans are better off, but by how much depends on whether you view matters as an American or an Englishman. True.

Figure A23.1



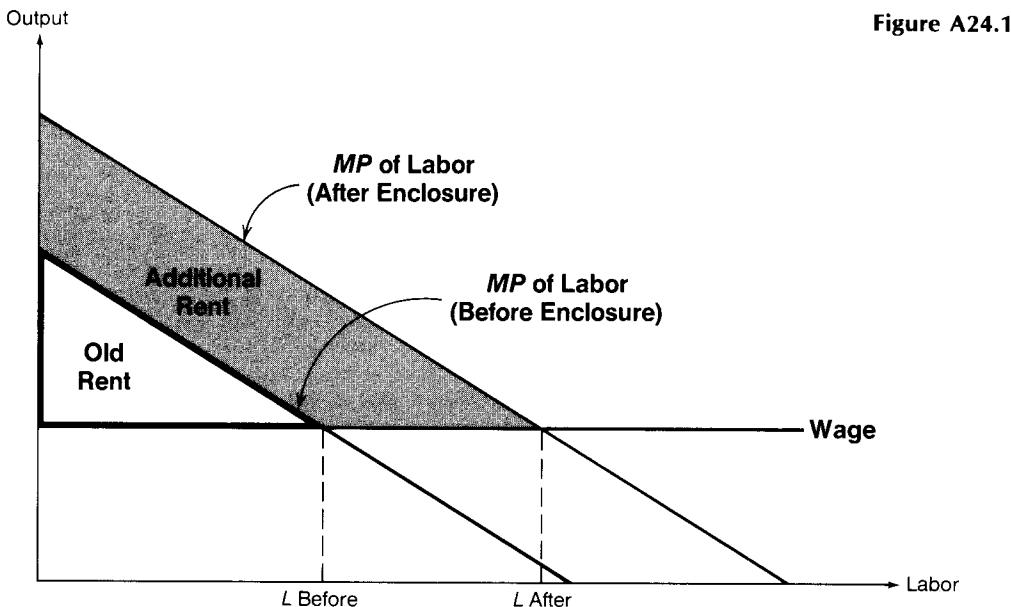
CHAPTER 24

Exercises for Section 24.1

1. a. At 20 cents an hour the value of marginal product falls to the wage in alternative employment at 2400 hours per year. Lebergott would earn $(\$0.20)(2400) = \480 per year.
- b. From 0 to 200 hours Lebergott brings in output worth \$0.56 for each hour, or $(200)(\$0.56) = \112 . From 200 to 400 he brings in $(200)(0.53) = \$106$. And so forth, in a series like this: $\$112 + \$106 + 100 + \$94 + \dots + (200)(\$0.20) = \$958$ per year. Updike pays Lebergott out of this. So she keeps $\$958 - \$480 = \$478$ per year, or half of gross revenues, near enough.
- c. At 1000 hours a year (notice that the value of marginal product is \$0.41, about double its level at 2400 hours) the series of gross revenue is $\$112 + \$106 + \$100 + \$94 + \$88 + \$82 + \$582$. The wage bill is $(1000 \text{ hours})(\$0.20 \text{ an hour}) = \200 . So Updike's net earnings are $\$582 - \$200 = \$382$. She loses $\$478 - \$382 = \$96$ a year, or about one-fifth of her best income.
- d. Lebergott is employed by Updike 2400 hours a year. The other $3200 - 2400 = 800$ hours he supplies to some other farm or business, earning the same \$0.20 per hour in money. Or else he takes the 800 hours in leisure, which he must value at \$0.20 because that is what he gives up to "buy" the leisure. In either case his full income is all his hours available evaluated at the going wage: $(3200 \text{ hours})(\$0.20) = \640 per year.
3. The amount of corn would increase (actually, it did not, but its cost of production fell). Its price would fall, driving down the value of the marginal product (which you recall is the marginal *physical* product times the price of corn). Updike's net earnings would therefore decline. Consumers benefit in the last instance. Lebergott is indifferent as a worker; Updike loses her initial gain. Both benefit only as *consumers*, like the rest of society.

Problems for Section 24.1

2. See Figure A24.1. If the supply curve of labor is perfectly elastic to each town (a reasonable assumption, since only a few areas were enclosed at a time), then true. The line marked Wage is in this case the supply curve, and by reasoning similar to that in Figure 24.1 in the text, all the rise in



output is paid to rent owners. If supply were not perfectly elastic, some of the rise in output would be paid to labor.

5. a. Since the economy is split into equal halves, the optimal allocation must be half to each. The corresponding national income is

$$2K^{0.5} = 2\sqrt{1} = 2$$

This is the maximum. With identical production functions the marginal products of K have identical curves in the two halves. Any deviation of allocation from

$$K_2 = K_1 = 1.0$$

will cause a loss in income (a triangle of welfare loss).

b. Insert the bad choices for the allocation of capital into the marginal product function given in the question:

$$0.5(0.4)^{-0.5} = 0.80 \text{ unit of output per unit of } K$$

$$0.5(1.6)^{-0.5} = 0.40 \text{ unit of output per unit of } K$$

That is, the marginal product of capital in the capital-poor region is twice that in the capital-rich region, as required. Capital earns twice as much per unit in one place as in the other. The national income is:

$$(0.4)^{0.5} = 0.65 \text{ and } (1.6)^{-0.5} = 1.26, \text{ for a sum} = 1.91$$

This is only 4.5% below the maximum income of 2.0! Evidently, misallocations even of an apparently radical sort do not result in large losses of income. Large increases in income will probably be difficult to explain as a consequence of correcting misallocations.

7. The point is that he has fortuitously set the relevant wage (the wage the sharecropper faces) at 50% of the true wage, and this is equal to the share. The halving of the wage exactly offsets the halving of the marginal product. The calculated equilibrium gives therefore the same labor employed as the optimal equilibrium. He therefore gets, by accident, the optimal result even though he makes Marshall's (dubious) assumption that the sharecroppers by themselves decide how much labor is to be applied to the acre. He had better look closely into the assumption that the alternative wage is 50% of the wage paid by landlords.

11. The supply of cab drivers may be presumed to be perfectly elastic in the long run. An increase in fares will raise the drivers' marginal product for a short time; but in the long run the number of drivers will adjust (they will increase, because cab revenues from monopolistically set prices are monopoly rents; see Chapter 19) but the drivers' wage will stay the same. False, if the short run is short enough.

Exercises for Section 24.2

1. The equilibrium condition is $P(MPP_L) = w[1 + (1/E_L^S)]$. The term involving the elasticity of demand for the wool blankets, E_Q^D , drops out: as E_Q^D approaches infinity, $1/E_Q^D$ approaches zero. In the Monopsony Only case with the elasticity of supply of labor equal to 1.0, the equation is then $P(MPP_L) = 2w$. That is, the wage paid, w , is in equilibrium half the full value of the marginal product. If the monopsony power disappears, too, the term with E_L^S drops out as well, leaving just $P(MPP_L) = w$. The wage w is 100% of the value of the marginal product: the mill is at Perfect Competition in Figure 24.4.

3. According to the equation in the text, with the values $E_Q^D = 2$ and $E_L^S = 1$, the equilibrium condition is

$$P(MPP_L) \left(1 - \frac{1}{2}\right) = w \left(1 + \frac{1}{1}\right)$$

This implies that $P(MPP_L) = 4w$. The wage is now only one-fourth of the value of marginal product.

Problems for Section 24.2

3. a. OPEC could extract all the monopoly rents from gasoline without integrating into refining. Because of the fixed production technology, OPEC's price does not affect factor proportions but is simply passed on, plus a competitive markup on refining services, to consumers. It would not care.
- b. If technology is not fixed, OPEC's high price would cause refiners to substitute away from OPEC oil. This means that production would be inefficient. A monopolist maximizes profits by producing efficiently; so OPEC would have an incentive to refine as well, and to become an octopus.
- c. Coming into refining would allow OPEC to price discriminate, giving another reason for wanting to integrate forward into refining. Recall the formula for optimal price discrimination in two markets from Chapter 18; the prices charged are different if the demand elasticities are different. Adding a constraint hurts; OPEC would rather be able to charge different prices in the two markets.
9. The price of labor rises only because the demand for cotton rose, which shifted out labor's VMP schedule. Since the supply schedule of labor did not shift, profits necessarily increased as a result of the upward shift in the demand schedule. False.
11. The incentive in either case is the triangle of welfare loss. It might be thought, mistakenly, that the victims would want to buy out the monopolist so that they would no longer have monopoly rents extracted from them; but these rents will be capitalized into the monopolist's sell-out price. True.

CHAPTER 25

Exercises for Section 25.1

1. a. The economist who is just indifferent between professing and businessing is *the* marginal economist. There may be many like this.
- b. She earns exactly \$10,000, to compensate for the \$10,000 less in money salary.
- c. $8 + 5 + 8 - 4 - 3 + 5 = +16$ thousand dollars which more than compensates Professor X for her \$10,000 lower salary. She stays in professing quite happily.
- d. $2 + 2 - 1 - 4 - 2 = -7$, which drives the value of Professor Y's compensating differentials below \$10,000. He leaves professing.
- e. If many Ys leave, the supply of professors is reduced and the monetary differential shrinks. Professors—even those among them who delight in obnoxious students—have to be compensated more because the marginal professors (who set the salary) will leave without it.

Problems for Section 25.1

2. True. Ability as a lawyer is presumably correlated with a lawyer's salary as a lawyer. The higher-paid lawyer (that is, the one with more ability) sacrifices relatively more to become a judge than does the lower-paid lawyer. The value of the compensation for being a judge, money gain plus prestige gain, is relatively higher for the less able lawyers.
5. Quite the contrary. On this account their wage would be higher, not lower. That it is in fact lower is testimony to other differences between the two jobs: for example, engineers must at great expense learn something about engineering.
6. False. Wages taken in psychic satisfaction are untaxable; so the lower the money wage the lower the effective tax rate on total wages, and this skews the choice of occupations away from money-intensive but unpleasant ones. The executive limousine, the three-martini lunch, and the spacious office, which are untaxed, might merely be a cost-effective way of compensating workers, especially those with high salaries and therefore high income taxes.

Exercises for Section 25.2

1. a. The hours and wages are:

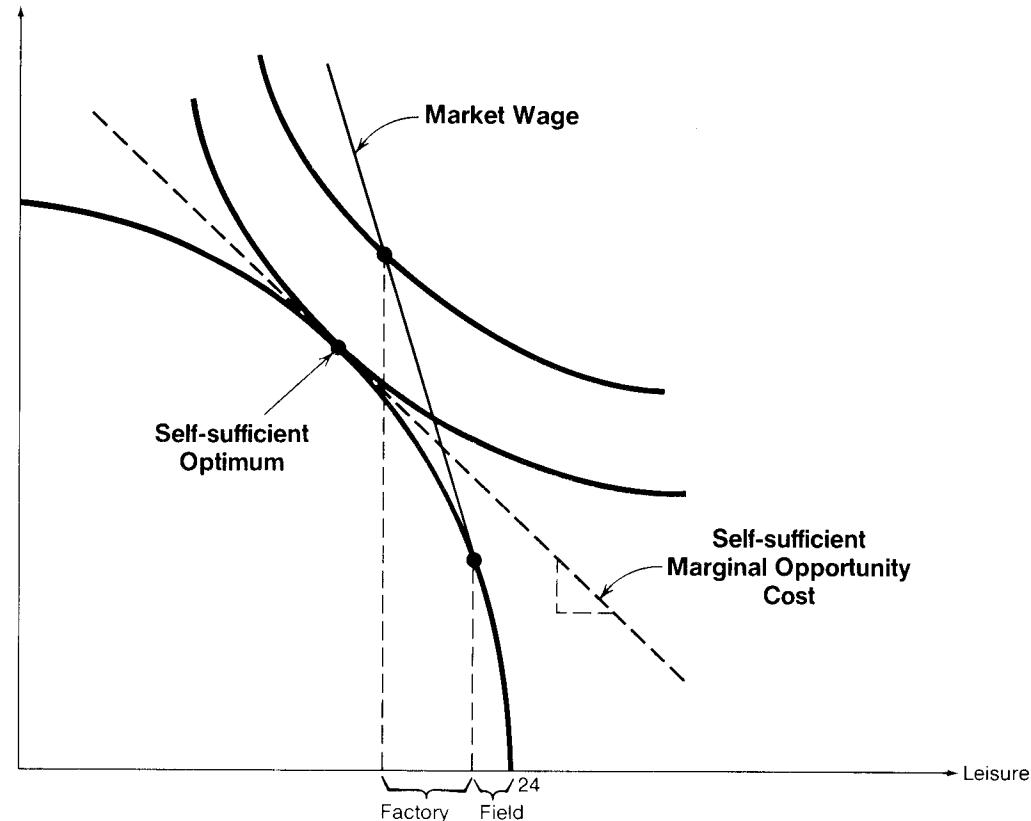
	Hours	Wages (tons per hour)
High wage	9	0.44
Equilibrium	8	0.38
Low wage	4	0.25

The wage is calculated as the commodities gotten per hour worked. For instance, at Equilibrium it is 3 tons/8 hours, or 0.38 ton per hour. These are the three supply points.

b. The income effect is the fall in hours supplied (the *rise* in leisure consumed) as one moves along the indifference curve from the old Equilibrium (an actual point) to Income Effect (a hypothetical point where a budget line with the old wage at Equilibrium is tangent to the new indifference curve through High Wage). In terms of hours it is $6 - 8 = -2$ hours of work. In terms of commodities, looking on it as an increase, it is $3.4 - 3.0 = 0.4$ ton.

Commodities
(Oats)

Figure A25.1



Problems for Section 25.2

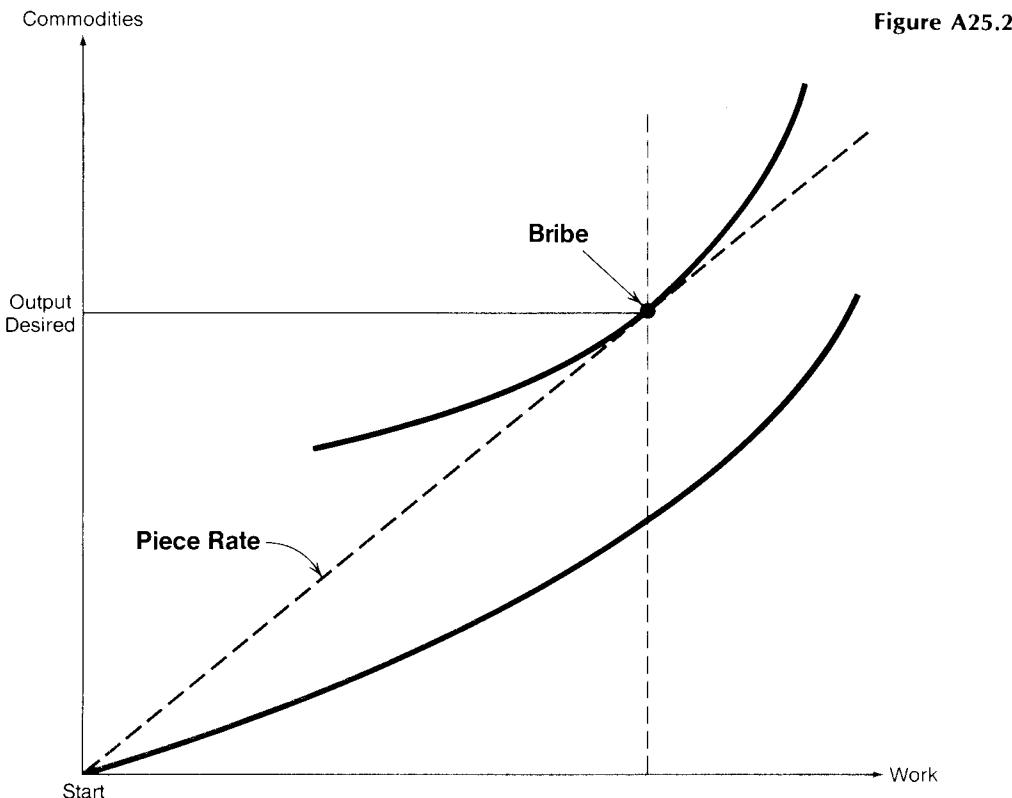
1. The typical American woman reveals that the value she places on housewifery is *less* than \$760; the typical Japanese woman reveals that the value for her is *greater* than \$630. The average shadow wage of housewifery, therefore, may be higher in Japan than America, so true.

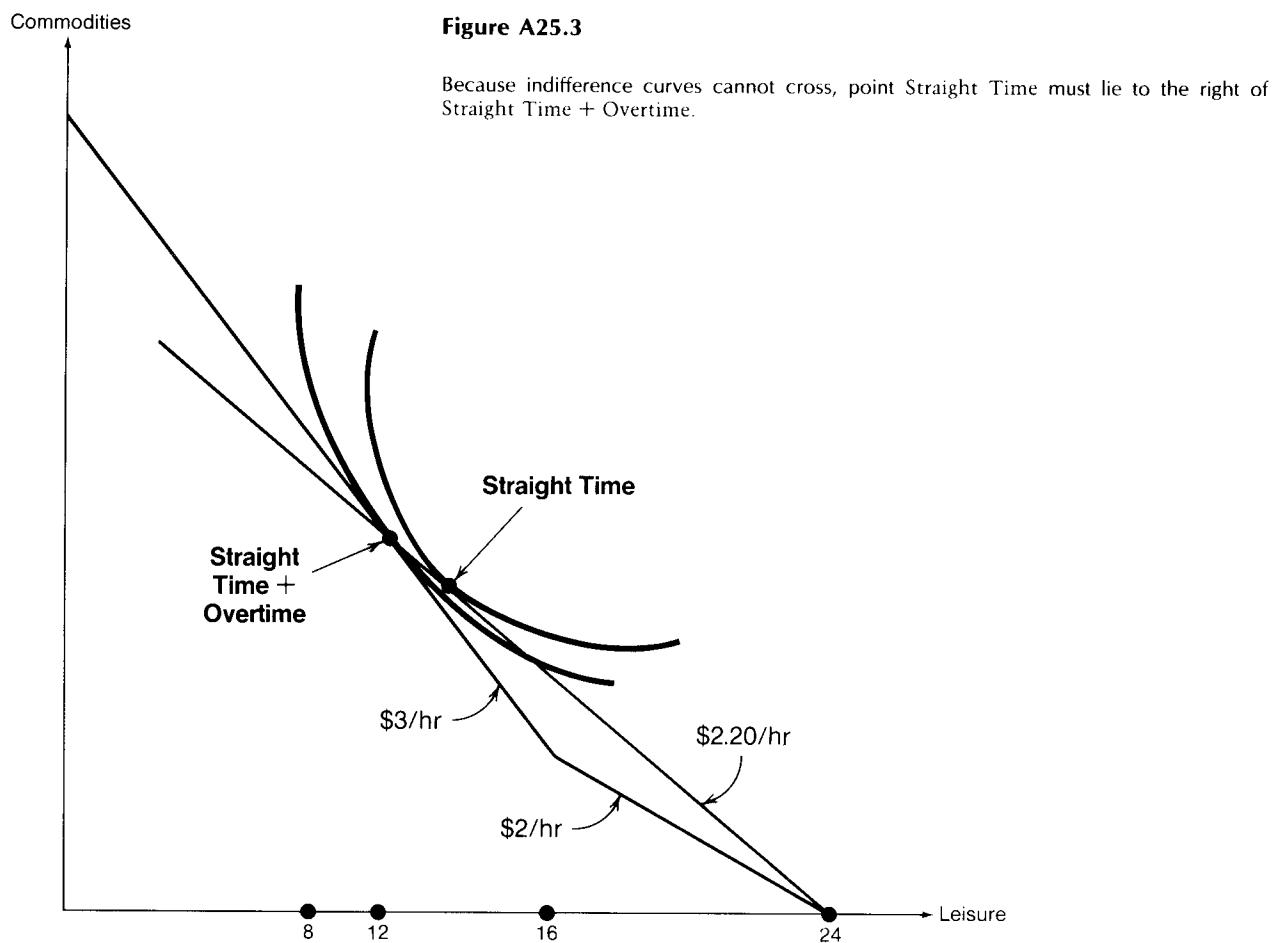
3. a. More educated women have a higher potential market wage and presumably a higher reservation wage as well (since they are probably more efficient homemakers). Thus both demand and supply shift up, and the effective market wage rises, drawing them out of the home.

b. In families with more children the shadow price of the mother's time is higher, assuming that child services are relatively time intensive. Also, the lower the wife's potential market wage, the lower the opportunity cost of having children and the more she will have. Finally, the large number of children affect the mother's human capital if they induce her to stay home more: she will choose to develop her skills less and will receive less on-the-job training.

9. All that is involved is that vacations are a normal good, and this is assumed by the marginal valuation condition in the question (technically speaking, leisure and goods are separable in the utility function). Unemployment insurance in this form increases the endowment of goods associated with maximum leisure; that is, the budget line moves out parallel to its old position. Leisure is a normal good, so more is consumed. True.

11. See Figure A25.1. Any wage greater than the self-sufficient marginal opportunity cost will induce the highlander to work for wages. His productive optimum is at the point where the marginal opportunity





cost on the farm equals the market wage. Because of the income effect of the new wage-earning opportunity, he may work fewer hours than before. That is, true.

Exercises for Section 25.3

1. "You get so many dollars a year and must work so many hours; but [and here's the trick] you get much of the pay in a lump sum *after* the harvest."
3. "I'll give you $(\frac{1}{2} \text{ cent})(20)(12)(6) = \7.20 a week if you produce $(20)(12)(6) = 1440$ collars per week. If you produce less than this I'll fire you." The results are the same.

Problems for Section 25.3

1. See Figure A25.2 here and compare it with Figure 25.6 in the text.
3. See Figure A25.3 and its caption. The diagram is clearer than Figure 25.5 in the text, which illustrates the same idea.
7. True. The argument runs Figure 25.5 in the text into reverse. A tax on inside-the-margin quantities (all income) affects economic decisions less than does a tax on marginal quantities (overtime income).

CHAPTER 26

Exercises for Section 26.1

1. a. $1/1.1 + 1/(1.1)^2 + 1/(1.1)^3 = \2.49 .
b. $1/(1.1)^3 = \$0.75$.
c. $1/(1.1) = \$0.91$.
d. \$3 exactly.
3. The capital value of a perpetuity, namely, $\$50/0.10 = \500 .

Problems for Section 26.1

1. The *real* rate of interest fell from 4% to 3%: from $4 - 0$ to $5\frac{1}{2} - 2\frac{1}{2}$. The value of an infinitely long-lived capital good, in other words, rose 33%. Improving the land by enclosure was a capital good, very long-lived. Therefore landlords had more incentive to enclose. The historians are wrong.
2. As stated in the text, the American interest rate was higher than the British; other things being the same, Americans would prefer to defer expenditures, relative to the British, by building less durable roads. To put it the other way, a project to build railways in a very durable fashion, yielding returns (less maintenance and rebuilding costs) far in the future would be less attractive to Americans: high interest rates make for future returns worth little.

Exercises for Section 26.2

1. $\$100/\$1000 = 10\%$.

Problems for Section 26.2

2. What is happening is that the one-month increase causes a rush to sell off the animals before the month ends; the permanent increase causes withholding of the animals in order to have more animals in the future.
4. On the contrary, he will tend to borrow (dissave), unless he strongly prefers future consumption to present consumption.
7. False. Read the problem carefully. It says "at any time in the next 10 years." You will obviously wish to make the investment on the eve of the return, and not a moment earlier.

Index

- Absolute advantage, 153
Adam Smith's generalization, 330
Adam Smith's Theorem, 101–103
Adding. *See* Summing
Adding constraints hurts, 258–259
Advertising, by monopoly, 374
Agency problem, profit motive and, 231–232
Agreements, inefficiencies of monopoly and, 392–393
Algebra, of constant returns to scale, 470–471
Allocation, 107–110
 good and bad, 487–494
Amenities, value of, 418
American Question, 303
Annuity, 532–533
Arbitrage, 209–210, 211, 212
Arms race
 Cournot solution and, 429–431
 games theory and, 444
Asset prices, expected future return and, 533–534
Autarky, consumers gain from, 215
Average cost, 236, 270–272
 from total cost, 249–250
Average labor cost, 500
Average profit, maximizing, 237
Average propensity to consume, 131–133
Average utility, 54–58
Avoidable fixed costs, 268–270
Backward-bending supply of labor, 515–516
Bargaining, 96–97
 contract curve and, 94–95
 in politics, 189
Baseball, monopsony power of, 500–501, 502–503
Behavior lines, indifference curves and, 52–53
Bertrand solution, oligopoly and, 421–422
Bid rent curves, 416–418
Bigness, optimality and, 499–500
Bliss point, 174, 175
Bonds, 530
Borrowing, at market interest rates, 538
Bribery
 scarcity causing, 109
 voting and, 189–190
 wages and, 525
Budget line, 8
 of businesses, 522 (*see also* Wages)
 for helping the poor, 521–522
 income in, 14–16
 indifference curves and, 42–43
 mathematics of, 16–17
 of nation (*see* Real income)

- Budget line—*Continued*
 opportunity cost as the slope of, 9–10
 prices in, 18–25
 scarcity and, 6–12
 social (*see* Production possibility curve)
- Bulging production possibility curve, 155–156, 162–163
- Calculus
 Cournot solution and, 423–424
 for Euler's theorem on homogeneous functions, 474–475
 for imperfect competition in factor markets, 499–500, 505
 for ordinality of indifference curves, 53–54
 for utility function, 32–34
- Capital
 cost of, 528–529
 rental rate of, 528–529
 supply and demand of, 528–540
- Capital intensive industries, value of the marginal product of labor in, 487–488
- Cardinal indifference curves, 52
- Cartels, 362–363
 as monopsony, 366
 oligopoly problem and, 434–444
 prisoner's dilemma faced by, 436–437, 441
- Chamberlin, Edward, 408
- Change, rates of, 142–143, 146–147
- City, transport costs to and rent, 416
- Coase theorem, 335–338
- Cobb–Douglas production function, 471, 480
- Cocktail Party Problem, prisoner's dilemma and, 440–442
- Common pool problem, 489–491
- Compensated demand, 75–77
- Compensating differentials, 508–512
- Competition. *See also* Perfect competition
 among the few (*see* Oligopoly)
 imperfect in factor markets, 449–500, 505
 by location, 411–418
 monopolistic, 366–367, 402–411
 price, 421–422
 restrictions causing waste in, 324–326
- scarcity causing, 109
 supernormal profits and, 320–327
- Competitive equilibrium, 403
 maximization of sum of consumers' and producers' surplus, 202–203
- monopoly
 barred entry and, 406–407
 free entry and, 407–408
 mutually advantageous exchange and, 203, 204
- Competitive industry, 276–288
- Complement
 effect of price on quantity demanded and, 123
 shape of indifference curves and, 35–36, 123
- Concavity, 248
- Concentration ratio, 375
- Constant elasticity, 140–146
- Constant-returns-to-scale, marginal productivity and, 469–472
- Constant-sum game, 444
- Constraints, on firm, 259–260
- Consumer's choice, 28–49
- Consumers' surplus, 196–221
 change in, 217–221
 marginal productivity and, 461–463
 taxes on, 312
- Consumption
 average and marginal propensities of, 131–133
 diagram of present and future, 536–537
 supply of labor and, 514–515
 theory of, 18
- Contract curve
 bargaining and, 94–95
 points on, 93
 production possibility curve and, 162, 163
- Convexity, of indifference curves, 37–38
- Cooperation
 oligopoly and, 420–421
 prisoner's dilemma and, 435–444
- Corporate income tax, 315–316
- Corruption, restrictions producing, 322–324. *See also* Bribery
- Cost-benefit analysis, consumers' surplus and, 217–218
- Cost of capital, 528–529
- Cost curves. *See* Firm

- Cost function, 257–258
duality and, 254–255
- Cost push, 226
- Counterfactuals, 366, 367
- Cournot solution
irrationality of, 428–433
oligopoly and, 422–427
- Crime, utility-of-income curve and, 66–67
- Crops, theory of location applied to where grown, 416–417
- Cross elasticity, 292
- Deadweight loss, taxation causing, 312–313
- Defection, prisoner’s dilemma and, 442
- Delayed wages, 522
- Delivery costs, 253
- Demand. *See also* Demand curve; Elasticity of demand; Utility function
of capital, 528–540
cross elasticities of, 292
derived, 450–458
inelastic, 136–137
for labor, 450–456
marginal productivity and, 457–465
- Law of, 19–25
need and, 78–80
perfect discrimination and, 395
summing, 117–121
supply and, 84–88, 101–103, 119, 121
using, 107–128
- Demand curve, 23–24, 124
for capital, 535–540
consumers’ surplus related to area under, 198–199
determination of, 107–110
downward sloping, 252–253
economy wide, 165–169
income compensated, 75–77
indifference curves and, 71–81
for local monopolists, 403–406
marginal revenue curve for straight-line, 349–350
monopoly and, 356
monopsony and, 360–361
offer curve and, 75–76
pathologies, 78–81
- Demand pull, 226
- Derived demand, 450–458
- Diminishing marginal returns to scale, 248–249
- Diminishing marginal utility, utility of income curve and, 63
- Diminishing returns
of monopoly in selling, 252–253
rising marginal cost and, 251–252
size of firm and, 253–254
- Discounting, 530–531
- Discrimination, in hiring, 451–452
- Discriminatory monopoly, 380, 382, 383, 384, 393–395
- Discriminatory monopsony, 501–502
- Dissipation of rents, 373–374
- Diversification, risk-aversion and, 68–69
- Division of labor, 153
- Dollar voting, theory of, 188–190, 191–192
- Don’t-do-it-all-yourself principle, 243–245
- Downward-sloping supply curve, 285
- Dual measure, of productivity change, 481–482
- Duality, 254–255
- Duopoly, 346, 420, 421
- Econometrics, 40
- Economic forecasts, 303–304
- Economic good, monopoly as, 434
- Economic growth, fundamental theorem of marginal productivity and, 480
- Economic rent, 201, 294–295. *See also* Producer’s surplus
- Economies of scale, 248
industry, 285–288
intervention and, 286–288
- Economywide diagrams, 167–170
- Edgeworth box, 91–93
monopoly’s behavior in, 391–392
for production, 162
- Efficiency, 93–94
allocation and, 278–279
of equilibrium, 101–103, 167
isoquants showing, 160
of market, 103–105, 167
of monopoly
price discrimination and, 393–395
transaction costs and, 392–393
- Morality and legality, 336–338

- Efficiency—*Continued*
 - profit maximization and, 277
 - restrictions on, 324
- Elasticity
 - constant, 140–146
 - with respect to income, 134–136, 139
 - with respect to price, 136–139
- Elasticity of demand
 - Cournot solution and, 424–425
 - excess, 145–146
 - facing one seller, 144–145
 - monopoly and, 356, 503
 - in foreign competitive market, 377–380
 - Fundamental Equation of Discriminatory Monopoly and, 380–382
 - as measure of power of, 375–376
 - taxation affecting, 309–310
 - Elasticity of supply, 131–139
 - excess, 145–146
 - Employment, discriminatory monopsony and, 502
 - End-to-end monopolies, Cournot assumption and, 425–427
 - Engel curve, 135
 - Entrepreneur, 293
 - Entry
 - monopolistically competitive equilibrium and
 - with barred entry, 406–407
 - with free entry, 407–408
 - to monopoly, 374
 - monopoly blockading, 345–346
 - principle of, 298–304
 - restrictions on, 320–327
 - taxation and, 316
 - Entry price, 293–294
 - Envelope of short-run curves, long-run curve as, 262–263
 - Equalizing differentials, theory of, 511–512
 - Equilibrium. *See also* Competitive equilibrium
 - as best society can, 101–103
 - efficiency of, 167
 - of market, 97–100
 - maximization of sum of consumers' and producers' surplus and, 202–203
 - uses of, 107–110
 - Equilibrium price, 111–114
 - Equimarginality, 241–245
 - violation of, 278–279
 - Ethics, economics of, 171–178
 - Euler's theorem on homogeneous functions, 472, 474
 - Excess demand, 145–146
 - Excess supply, 145–146
 - Exchange
 - between two people or nations, 89–95
 - beyond monopoly point for monopoly, 392
 - Edgeworth box and, 91–93
 - game theory and, 444
 - indifference curves and, 91
 - as mutually advantageous, 89–90
 - competitive equilibrium and, 203, 204
 - taxation as constraint on, 307–308
 - Exhaustion of the product, problem of the, 468–469
 - constant returns-to-scale and, 469–472
 - Euler's theorem on homogeneous functions and, 472, 474–475
 - marginal productivity and, 472
 - Exit, 301–302, 303–304
 - Expected returns, 533–534
 - Exponents, elasticities and, 140–142
 - Exports
 - monopoly power and, 379–380
 - restrictions on, 215
 - External economies, 334
 - External economy of industry scale, 285–288
 - Externalities
 - industry supply and, 283–288
 - nonpecuniary, 331
 - road congestion as, 332–334
 - Extortion, unanimity and, 191. *See also* Bribery
 - Face value, of bonds, 530
 - Factor markets
 - of imperfect competition, 499–500, 505
 - monopoly in, 498–504
 - Factors of production, 161. *See also* Capital; Labor; Land
 - as fungible, 449–450
 - households owning, 448–449
 - Few, competition among the. *See* Oligopoly

- Firm
as agent for other people, 224–225
alternatives to, 227
constrained, 258–260
cost curves of, 249
fixed and variable costs and, 265–272
long and short-run, 257–263
marginal cost, 247–255
equimarginality and, 241–245
existence of, 224–229
marginal cost and, 234–239
number of as measure of monopoly, 375
profit motive and, 229–234
rule of rational life for, 498
size of, 253–254
optimality and, 498–500
- Fisheries problem, 489–491
- Fixed costs, 235
avoidable, 268–270
unavoidable, 268
- Foreign market, monopolist's price in, 377–380
- Foreign trade. *See* Trade
- Free immigration, marginal productivity and, 464–465
- Free markets, 103–105
maximization of sum of consumers' and producers' surplus and, 202–203
- Free riders, prisoner's dilemma and, 435–444
- Free trade
consumers' surplus and, 204
producers' surplus and, 201
- Fundamental Equation of Discriminatory Monopoly, 380–382
- Fundamental Equation of Monopoly, 376
- Fundamental theorem of marginal productivity, 472–473
- Fungibility, 7–8
factors of production and, 449–450
- Future costs, 267–268
- Future income, price of an asset estimating, 533–534
- Gambling, utility and, 54–60
utility of income and, 64–65
- Game theory, oligopoly problem and, 444–445
- General discounting formula, 532–533
- General equilibrium model, 169–170
- General will, 188, 189
- Giffen goods, 78
- Giffen paradox, 78
- Government, prisoner's dilemma and, 438–440
- Hicks neutral, 480
- Hill of utility, 28–29
- Historical costs, 267–268
- Homogeneous functions, Euler's theorem and, 472, 474–475
- Horizontal consumers' surplus, 210–211, 214–215
- Horizontal summation, 124
- Households, factors of production owned by, 448–449
- Human capital, investment in, 540
- Immigration, marginal productivity and, 464–465
- Imperfect competition, in factor markets, 499–500, 505
- Imports
monopoly power and, 379–380
quotas on, 320–321, 324
- Incidence, 301
- Income. *See also* National income; Utility of income curve
in budget line, 14–16
real, 179–180
measurement of, 180–183
segregation by, 301
- Income compensated demand curve, 75–77
- Income distribution
desirability of, 192–193
marginal productivity and, 461, 462
national income and, 463–465
- Income effect, 15–16, 74, 75, 77
inheritances, 516
subsidies, 516–518
- Income elasticity, 134–136, 139
- Income subsidies, 516–518
indifference curve and, 40–41
- Income tax. *See also* Taxation
negative, 104–105, 518–519
progressive, 60–61, 104

- Index number problem, 183–184
 Indifference curve, 29–32. *See also* Marginal utility
 as behavior line, 52–53
 budget line and, 42–43
 calculus for ordinality of, 53–54
 cardinal, 52
 complements and, 35–36
 convexity of, 37–38
 demand curve and, 71–81
 equalizing differentials and, 511–512
 exchange exhibited by, 91
 gifts that foreclose other opportunities and, 41–42
 marginal utility and, 48–49
 order of, 52
 ordinal, 52
 revealed preference and, 52–53
 shape of, 34–39
 subsidies of income and price of, 40–41
 substitutes and, 36–37
 Industry
 competitive, 276–288
 definition of, 290–292
 suppliers to, 291–292
 Industry supply curve
 downward-sloping, 285
 externalities and, 283–288
 fixed number of firms and interdependent costs and, 283–288
 flat, 294
 long-run, 290–296
 marginal cost curve and, 276–281
 upward-sloping, 279–281
 inelastic supplies and, 283–285
 Inelastic demand, 136–137
 Inelastic goods, taxation and, 309
 Infant industry tariff, 389
 Inferior good, 16, 17
 Infinite elasticity, 136
 Inflation
 interest rates and, 529–530
 monopolies and, 365–366
 Inheritances, 516
 Innovation. *See also* Technological change
 fundamental theorem of marginal productivity and, 480
 monopoly and, 389–390
 Input
 many at once as, 467–469
 production function and, 247–248
 for profit-maximizing output, 457–460
 size of firm and, 254
 Interest, real versus nominal, 529–530
 Interest rate, 528–534
 borrowing and lending and, 538
 ubiquity of, 530
 Internal rate of return, 539–540
 Invention. *See* Innovation
 Investment, present value maximization and, 539
 Isolated city, 415
 Isoquants, 160. *See also* Unit isoquant
 Issues, monopoly, 395–398

 Joint supply, 124–128
 Joys, utility measured by, 54–58

 Keynes, John Maynard, 118–119
 Knowledge, technological change and economics of, 484–485

 Labor
 as a commodity, 448–456
 demand for, 450–456
 marginal productivity as, 457–465
 division of, 153
 Law of One Price to, 449–450
 marginal cost of, 499–500, 505
 supply of, 508–526
 backward-bending, 515–516
 consumption and, 514–515
 leisure-commodity choice determining, 514–515
 theory of supply of labor and, 514–515
 Labor monopsony, 500–501, 502–503
 Laissez-faire capitalism, 172–174, 175, 177
 failures of, 285
 Large groups, prisoner's dilemma and, 440–442
 Laspeyres measure, 184–186
 of national income, 219–220
 Law of Demand, 19–25
 Law of One Price, 117–118
 to labor markets, 449–450

- Leisure
choice between work and, 513–519
supply curve of labor and, 514–515
- Lending, at market interest rates, 538
- Leontief production function, 159
- Lexicographic preferences, 80–81
- Liability, efficiency and, 336–338
- Licenses, 321, 322–323
- Local monopolist
demand curve for, 403–406
maximum of profit for, 406–407
- Location, theory of, 411–418
- Long-run cost curve, 261–262
as envelope of short-run curves, 262–263
- Long-run costs, 260
- Long-run industry supply curve, 290–296
- Lump-sum tax, 314
- Lying, as to value of public goods, 191–192
- Major league sports, monopsony power of, 500–501, 502–503
- Margin of Cultivation, 296
transport costs determining, 415–416
- Marginal benefit, revenue product as, 458
- Marginal cost, 200, 235–236
equimarginality and, 241–245
firm and, 234–239
monopoly and, 364–366
Fundamental Equation of Monopoly and, 376
ratio of as measure of monopoly power, 376
without marginal cost, 351–353
- total cost and, 249–250
wage as, 458
- Marginal cost curve
average cost curve and, 249–250
of firm, 247–255
slope of, 251–252
upward curvature, 253–254
- Marginal curves, 46
- Marginal labor cost, 499–500, 505
- Marginal productivity
constant returns to scale and, 469–472
consumers' surplus and, 461–463
demand for labor and, 448–465
- to derive a measure
of invention, 480
of war, 480–481
- distribution of income and, 461, 462
national income, 463–465
- equation of, 460–461
- fundamental theorem of, 472–473
many inputs and, 467–469
- Marginal products, equalization of values of, 487–490
- Marginal propensity to consume, 131–133
- Marginal rate of indifferent substitution, 38
- Marginal revenue, 235–236
monopolist with no costs and, 351–353
monopoly and, 364–366
price and spoilage, 347–349
- Marginal revenue curve, for straight-line demand curve, 349–350
- Marginal revenue product, 499–500, 505
- Marginal revenue product of labor, 458, 459, 460
- Marginal utility
equations of, 47–48
as not measurable absolutely, 59–61
Rule of Rational Life and, 46–47
total utility and, 44–45
utility of income curve and, 63
- Marginal utility of income, 46, 62–63
- Marginal valuation, 38, 124, 167
vertical addition of, 190–191
- Marginal value product, 500
- Marginal willingness to pay, 198
- Market
clearing by, 114–116
Law of One Price and, 117–118
consumers' surplus for examining goodness or badness of, 207–210
demand curve and, 111–114
equilibrium and, 97–100
free, 103–105
noneconomist and, 103
related, 123–124
supply curve and, 111–114
taxes constraining, 307–308
- Market for factors of production, 449
- Market power, elasticity as measure of, 144
- Market sharing, cooperative solution to oligopoly and, 421, 422
- Maturity, of bonds, 530

- Microeconomics, 1
 Middlemen, 207–210, 213
 Minimum wages, 322, 453–455
 Misallocation, 487–494
 Moneymaking. *See* Profit motive
 Monopolistic competition, 402–411
 theory of, 408–410
 Monopoly, 420. *See also* Marginal cost; Oligopoly
 behavior, 344–369
 as common, 371–372
 diminishing returns in selling and, 252–253
 elasticity of demand and, 503
 factor markets and, 498–504
 location and, 411–418
 marginal revenue products and, 499–500, 505
 measuring, 371–385
 monopolistic competition, 402–411
 octopus effect and, 503–504
 output and, 344–345, 353–356, 390–392
 welfare economics of, 388–399
 Monopoly rent, 373, 374
 Monopsony, 346
 applications of, 500–504
 cartel as, 366
 demand curve and, 360–361
 discriminatory, 501–502
 labor, 500–501, 502–503
 marginal revenue product and, 499–500, 505
 straight-line supply curve for, 350–351
 Morgenstern, Oskar, 444
 Multipart pricing, 383–384
 Mutually advantageous (beneficial) exchange, 89
 competitive equilibrium and, 203, 204

 National income, 179–186
 changes in, 217–221
 marginal product for maximizing, 488–489
 value of housework and, 514
 National monopoly, 367, 368
 Need, economics of, 78–80
 Negative income tax, 104–105, 518–519
 New tax, 310–311
 Nominal interest rate, 530

 Noncooperation
 Bertrand solution as, 421–422
 Cournot solution as, 422–427
 Noneconomist, market interference by, 103
 Nonpecuniary externalities, 331
 Normal good, 16, 17
 Normal profits, 298–300
 Normal return, 293
 Normative economics, 171

 Octopus effect, 503–504
 Offer curves, 75–76
 supply equaling demand and, 101–103
 Oligopoly, 346, 420–446
 Opportunity cost, 152
 as slope of budget line, 9–10
 specialization theorem and, 153–156
 uses of, 529
 Optimality *One-man-one Task*
 bigness and, 499–500
 smallness yields and, 498–499
 Ordinal indifference curves, 52
 Organization of Petroleum Exporting Countries (OPEC), 362, 391–393
 Output
 monopoly and, 344–345, 353–356, 390–392
 production function and, 247–248
 input for profit maximization and, 457–460
 taxation and, 309–310
 Output per man, higher production function and, 477–479
 Outsmarting principle, 445
 Cournot solution and, 431–433
 Overfishing, 489–491
 Overtime pay, 522–523
 Ownership, misallocation due to lack of, 489–491

 Paasche measure, 184–186
 of national income, 219–220
 Par value, of bonds, 530
 Pareto efficient, 189n
 Pareto optimal, 93
 Partial equilibrium model, 169–170
 Past costs, 267–268

- Patent, 367–369
secondhand buyer of, 371
- Pathologies, of demand, 78–81
- Payback, 539
- Payoff matrix
prisoner's dilemma as, 435
punishment and, 437–438
- Perfect competition, 403
marginal product of labor and, 500
- Perpetuity, 531–532
- Piece rate of pay, 523–525
- Planned obsolescence, monopolies and, 390
- Political parties, similarity of, 412–413
- Politics
economics of, 187–194
principle of entry and, 300–301
theory of location and, 411–414
- Poll tax, 314
- Pollution, tax on, 324–338
- Polypoly, 346
- Positive economics, 171
- Preferences, of consumer, 28
- Present consumption, diagram of future and, 536–537
- Present value, 531
maximizing, 539
of perpetuity, 531–532
- Prestige value, price and, 128
- Price
in budget line, 19–25
marginal revenue and, 347–349
monopolies raising, 344–345
productivity measured by, 481–482
relative, 180
slope as, 10–12
- Price competition, Bertrand solution as, 421–422
- Price controls, 321–322
eliminating waste from, 330–331
for monopoly, 361–363
- Price discrimination, monopoly and, 382, 383, 393–395
- Price effect, 19
- Price elasticity, 136–139
- Price gouging, 99
- Price movements, supply and demand inferred from, 114–115
- Price searching, by monopolists, 347–349
- Price subsidies, indifference curves and, 40–41
- Price taking, 97, 98, 111, 347
as consequence of summing, 121
exchange and, 101–103
firm as, 252
Law of One Price and, 118
- Price war, oligopoly as, 422
- Priorities, 80–81
- Prisoner's dilemma, oligopoly and, 435–444
- Producers' surplus, 201–203
change in, 217–221
taxes on, 312–313
- Product, 468. *See also* Exhaustion of the product
- Product differentiation, 409
- Production, substitutability in, 159–163
- Production function, 247, 257–258
changes in, 476–485
Cobb–Douglas, 471
duality and, 254–255
input for profit maximizing output determined by, 457–460
overexhaustion and, 468–469
production possibility curve and, 158–163
total cost curve from, 249–250
unit isoquant and, 477
- Production possibilities, 153–170
- Production possibility curve, 8
production function and, 158–163
specialization theorem and, 153–156
supply and demand curve and, 166
- Productivity, price measuring, 481–482
- Productivity change, new knowledge and, 484–485
- Profit, 201. *See also* Producers' surplus
average, 237
competition for, 320–327
marginal productivity as theory of, 461–463
as measure of monopoly, 372–373
normal, 298–300
restrictions on entry resulting in, 322–323
short-run versus long-run, 237
supernormal, 294–295, 298–300
total, 237
zero, 300, 301

- Profit margins, fallacy of, 236
- Profit maximization, 237–239, 266–267
 - efficiency and, 277
- Profit motive, firm and, 229–234
- Progressive taxation, 60–61
- Property rights, 320–340
 - reasons for, 338–339
- Public goods
 - truthful valuations of, 191–192
 - voting about, 301
- Public projects, consumers' surplus for, 198–281
- Public utilities, 367
- Punishment
 - prisoner's dilemma and, 437–438, 441, 442
 - utility-of-income curve and, 66–67
 - wages and, 525
- Pure bargaining, Cournot solution and, 428–429

- Quantity, supply or demand inferred from, 114, 115
- Quantity competition, Cournot solution as, 422–427
- Quasi rent, 294–295
- Queuing, 324, 325–326
- Quotas, import, 320–321, 324

- Railway, as industry, 290–291
- Rates of change, 142–143, 146–147
- Ratio of prices, Law of Demand and, 23
- Ration coupons, 326–327
- Rational expectations, 303–304
- Rational Life. *See* Rule of Rational Life
- Rationality, of Cournot solution, 428–433
- Real income, 179–180
 - measurement of, 180–183
- Real interest rate, 529–530
- Real wage, 463
- Realpolitik, 189
- Recontracting, 99n
- Regulation, 302
- Related markets, 123–124
- Relative price, 180
 - slope as, 10–12

- Rent, 294–295
 - bid rent curve and, 416–417
 - dissipation of, 373–374
 - monopoly and, 373, 374
 - rise in due to inelastically supplied factors, 488–489
 - transport costs to central city determining, 416
- Rent control, 205, 321–322
- Rental rate of capital, 529
- Residual productivity change, 479
- Return, internal rate of, 539–540
- Revealed preference, indifference curves and, 52–53
- Ricardo, David, 415
- Rising production functions, 477
- Risk aversion, unfair insurance and, 65–66
- Roads, ownership of, 332–334
- Rule of Rational Life, 46–47, 234–236
 - for a firm, 498
 - for monopoly, 347, 364–366
 - in two markets, 380–382
 - with no costs, 351–353
- Run profits, maximizing, 237

- Salary. *See* Wages
- Say's law, 86
- Scalping, scarcity causing, 101–110
- Scarcity
 - budget line and, 6–12
 - competition due to, 109
- Schumpeter, Joseph, 389
- Second best theory, 313–314
- Seeds of growth theory, 537–538
- Segregation, by taste and income, 301
- Sensitivity. *See* Elasticity
- Separation theorem, 538
- Shadow wage, 513–514
- Sharecropping, as taxation, 491–494
- Short-run cost curves, 257–264
 - long-run as envelope of, 262–263
- Slope
 - of budget line, 75–76
 - as relative price, 10–12
- Small groups, prisoner's dilemma and, 441–444
- Smallness of firms, optimality and, 498–499

- Smith, Adam, 277, 344
Social budget line. *See* Production possibility curve
Social cost, 300
Social good, monopoly for, 388–389
Social indifference curves, 174–178
 egalitarian, 174–175, 176
Social loss, from tax, 312–313
Specialization by comparative advantage, 153
Specialization theorem, 153–156
Spoilage, marginal revenue and, 347–349
Spoiling the market of monopolist, higher output and, 447–448
State, prisoner's dilemma and, 438–440
Steady state, 538
Straight-line demand curve
 elasticity of, 138
 marginal revenue curve for, 349–350
Straight-line production possibility curve, 153
Straight-line supply curve, monopsonist and, 350–351
Subsidies, 516–518. *See also* Income subsidies
Subsistence theory of wages, 300
Substitutability
 for definition of an industry, 291–292
 octopus effect and, 504
 in production, 159–163
Substitutes
 effect of price on quantity demanded and, 123
 shape of indifference curves and, 36–37, 123
Substitution effect, 74–75, 77
Summing
 horizontal, 124
 price taking as consequence of, 121
 of supply and demand, 117–121
 vertical
 joint supply requiring, 124–126
 of marginal valuation, 190–191
Supernormal profit. *See* Profit
Supply. *See also* Labor; Opportunity cost; Production functions
 applied to entire economies, 165–170
 of capital, 528–540
demand and, 84–88, 101–103, 119, 121
 elasticities of, 131–139, 145–146
 excess, 145–146
 joint, 124–128
 perfect discrimination and, 395
 price movements and, 114–115
 summing of demand and, 117–121
 taxation affect on, 308
Supply curve, 85–86. *See also* Industry supply curve
 for capital, 535–540
 economywide, 165–169
 of market, 111–114
 monopoly, 360–361
 upward-sloping, 279–281
Surplus of producers' revenues over costs, 201

Tariff, 321
 infant industry and, 389
Taste, segregation by, 301
Taxation, 307–317. *See also* Income tax
 of pollution, 334–338
 progressive, 60–61, 104
 sharecropping as, 491–494
Technological change. *See also* Innovation biased, 483
economics of knowledge and, 484–485
marginal product for beneficiaries of, 488–489
spillovers and, 483
Temperature, measurability of, 56
Tiebout effect, 301
Total cost curve, 266–267
 from production function, 249–250
Total factor productivity change, 479. *See also* Marginal productivity
Total production of labor, 458, 459
Total productivity, 479
Total revenue
 for marginal revenue for straight-line demand curves, 349–350
 monopolist with no costs maximizing, 351–353
Total revenue curve, 266–267
Total revenue product of labor, 458, 459
Total supply, total demand equal to, 119, 121

- Total utility, marginal utility and, 44–45
- Total utility curve, 44–45
- Trade, 84–105. *See also* Exchange
 - among many people, 96–100, 101–105
 - quotas, 320–321, 324
 - tariffs, 389
- Trade-off, 8
- Transaction costs, inefficiencies of monopoly and, 392–393
- Transformation curve, 8
- Transport cost, 252–253
 - margin of cultivation determined by, 415–416
- Two-good diagram, 111

- Unanimity
 - extortion and, 191
 - value of, 189
- Unassigned property rights, 330–339
- Unavoidable fixed costs, 268
- Uncompensated demand, 75–77
- Unemployment, minimum wage causing, 453–455
- Unfair insurance, 65–66
- Unit elasticity, 136
- Unit isoquant, 161n
 - production function and, 477–478
- Upward-sloping supply, 279–281
- Utilitarianism, utility possibility curve and, 174, 175
- Utility. *See also* Marginal utility
 - average, 54–58
 - of different prospects, 54–58
 - gambling theory and, 54–60
 - from goods, 28
 - measurable, 44–49, 54–58, 62–69
 - nonmeasurable, 51–54
 - as not comparable between people, 60–61
 - as not measurable absolutely, 59–61
 - total, 44–45
- Utility function, 32–34
 - production function different from, 159
- Utility of income, 62–63
- Utility of income curve, 62–63
 - crime and punishment and, 66–68
 - gambling and, 64–65
 - risk aversion and, 65–66

- Utility possibility curve
 - laissez-faire* and, 172–174
 - utilitarianism and, 174, 175

- Value in exchange, 196–198
- Value of marginal product, 499, 500
- Value-in-use, 196–198, 199
- Variable costs, 268–269
- Variable sum game, 444, 445
- Varying costs, of monopoly, 354–355
- Vertical consumers' surplus, 210, 214–215
- Vertical summation. *See* Summing
- von Neumann, John, 444
- von Thünen, Johann H., 415
- Vote trading, 189
- Voting
 - about public goods, 301
 - bribery, 189–190, 191
 - dollar, 188–190, 191–192
 - general will and, 188

- Wages, 458–460
 - bribery and, 525
 - definition, 513
 - delayed, 522
 - minimum, 322
 - overtime, 522–523
 - punishment and, 525
 - shadow, 513–514
 - subsistence theory of, 300
 - theory of compensating differentials and, 508–512
 - by time versus by piece, 523–525
 - whole pay and, 509–511
- Waiting time, as wasteful, 324, 325–326
- Walras's law, 86n
- War, fundamental theorem of marginal productivity and, 480–481
- WARP (Weak Axiom of Revealed Preference), 21–23
- Waste, from monopoly seeking, 397–398
- Weak Axiom of Revealed Preference (WARP), 21–23
- Welfare economics
 - laissez-faire* and, 172–174
 - monopolistic competition and, 409
 - of monopoly, 388–399

- Wheel of wealth, 225–226, 448–
449
- Whole pay, 509–511
- Work
- choice between leisure and, 513–
519
 - payment (*see* Wages)
- Working conditions, whole pay and, 509–
511
- Zero elasticity, 136
- Zero-profit, 300, 301
- Zero-sum game, 444