Second Edition

Incentives

Motivation and the Economics of Information

Donald E. Campbell

CAMERINGS

CAMBRIDGE www.cambridge.org/9780521832045

This page intentionally left blank

INCENTIVES

Second Edition

This book examines the incentives at work in a wide range of institutions to see how and how well coordination is achieved by informing and motivating individual decision makers. Incentives work well when they result in a high level of individual welfare generally. This is problematic because each individual acts to maximize his or her individual payoff, regardless of its implications for the welfare of others. The book examines the performance of agents hired to carry out specific tasks, from taxi drivers to CEOs. It investigates the performance of institutions, from voting schemes to kidney transplants, to see if they enhance general well-being. The book examines a broad range of market transactions, from auctions to labor markets to the entire economy. The analysis is conducted using specific worked examples, lucid general theory, and illustrations drawn from news stories. The theory and examples are presented rigorously but not in an overly "high tech" way. Of the seventy different topics and sections, only twelve require a knowledge of calculus. The second edition offers new chapters on auctions, matching and assignment problems, and corporate governance. Boxed examples are used to highlight points of theory and are separated from the main text.

Donald E. Campbell is CSX Professor of Economics and Public Policy at The College of William and Mary, Williamsburg, Virginia, where he has taught since 1990. He previously served as professor of economics at the University of Toronto from 1970 to 1990. He is the author of *Resource Allocation Mechanisms* (Cambridge University Press, 1987) and *Equity, Efficiency, and Social Choice* (1992). His published research has appeared in leading journals such as *Econometrica, Journal of Political Economy, American Economic Review, Journal of Economic Theory, Review of Economics Studies*, and the *Journal of Mathematical Economics*.

SECOND EDITION

Incentives

MOTIVATION AND THE ECONOMICS OF INFORMATION

Donald E. Campbell

The College of William and Mary



CAMBRIDGE UNIVERSITY PRESS Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press The Edinburgh Building, Cambridge CB2 2RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org Information on this title: www.cambridge.org/9780521832045

© Donald E. Campbell 1995, 2006

This publication is in copyright. Subject to statutory exception and to the provision of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published in print format 2006

 ISBN-13
 978-0-511-21966-5
 eBook (EBL)

 ISBN-10
 0-511-21966-0
 eBook (EBL)

 ISBN-13
 978-0-521-83204-5
 hardback

 ISBN-10
 0-521-83204-7
 hardback

 ISBN-13
 978-0-521-53974-6
 paperback

 ISBN-13
 978-0-521-53974-6
 paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLS for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

For Soren, Rosie, and Edie, and their parents, Samantha and Tyler

Contents

Pre	efac	e to the Second Edition	page xi						
1	Eq	quilibrium, Efficiency, and Asymmetric Information							
	1.	Asymmetric Information	10						
	2.	Taxi!	16						
	3.	Acid Rain	18						
	4.	Efficiency	23						
	5.	Equilibrium	30						
	6.	The Prisoner's Dilemma Game	45						
	7.	Repetition and Equilibrium	53						
2	Ba	sic Models and Tools	71						
	1.	Maximizing a Quadratic	72						
ė	ð2.	Overview of Calculus	76						
	3.	Lagrangian Multipliers	86						
	4.	The Composite Commodity Model	98						
	5.	Quasi-Linear Preferences	102						
	6.	Decision Making Under Uncertainty	112						
	7.	Insurance	124						
3	Hie	dden Action	136						
	1.	Resource Allocation	139						
	2.	Marketable Pollution Rights	143						
	3.	Incentive Regulation of the Telecommunications Industry	152						
	4.	The Savings and Loan Debacle	155						
	5.	Personal Bankruptcy	164						
	6.	Mandatory Retirement	165						
	7.	Tenure and the Performance of Professors	174						
	8.	Pay and Performance in U.S. Public Schools	177						
	9.	Moral Hazard and Insurance	179						

4 C	orporate Governance	196
1	. A Brief Tour of Several Countries	197
2	. Partnerships	198
3	. The Owner-Employee Relationship	207
4	. The Owner-Manager Relationship in Practice	212
5	. Agency Theory	231
5 H	lidden Characteristics	256
1	. Price Discrimination	257
2	. Two-Person Exchange	259
$\partial 3$. The Used-Car Market	269
4	. Credit Rationing	272
$\partial 5$. Bundling and Product Quality	280
6	. Job-Market Signaling	290
7	. Competitive Insurance Markets	303
6 A	uctions	325
1	. Introduction	326
2	. The Vickrey Auction	334
3	. Four Basic Auction Mechanisms	349
4	. Revenue Equivalence	358
5	. Applications of the Revenue Equivalence Theorem	374
6	. Interdependent Values	377
7 V	oting and Preference Revelation	384
1	. Voting Schemes	385
2	. Preference Revelation in General	402
3	. General Proof of the Gibbard-Satterthwaite Theorem	411
4	. The Revelation Principle	418
8 P	ublic Goods and Preference Revelation	420
1	. The Economic Model	422
2	. The Pivotal Mechanism	440
3	. Groves Mechanisms	453
4	. Efficiency and Incentive Compatibility	457
9 N	Natching	467
1	. Students and Advisors	469
2	. College Admissions	480
3	. Hospitals and Doctors	496
4	. Allocating Dormitory Rooms	499
5	. Kidney Transplants	510

Contents

10	General Competitive Equilibrium				
	1.	Competition, Property Rights, and Prosperity	514		
	2.	The Arrow-Debreu Economy	523		
	3.	Nonconvex Economies	538		
	4.	Efficiency and Incentive Compatibility	543		
	5.	Common Property Resources	556		
Ref	erer	nces	561		
Au	thor	r Index	579		
Sul	ojec	t Index	583		

Preface to the Second Edition

I am pleased to have this opportunity to express my appreciation to the following students and colleagues who assisted me at various stages: My former students Hanley Chiang, Ryan Mutter, and Sita Slavov discovered some glitches in the first edition and brought them to my attention. My current students David Hansen, Jonathan Kuzma, and Emma Murray helped me fill in many of the boxes that connect the theory to contemporary events, and Matthew Draper did some preliminary spadework for Chapter 9. I am grateful for the superb diagrams produced by Carrie Clingan, a student in the Masters in Public Policy program at William and Mary. Jerry Kelly of the Syracuse University Economics Department made copious comments on early drafts of Chapter 7. I have benefited from the insightful comments of Ed Nelson of the Tulane Economics Department, John Weymark of the Vanderbilt Economics Department, and David Ellerman of the World Bank. My colleague and coauthor for two undergraduate texts, Alfredo Pereira, taught me how to write textbooks. I express my deep gratitude to these people. I assume responsibility for any errors in the book.

It is a pleasure to acknowledge the support and encouragement of Scott Parris, the economics and finance editor at Cambridge University Press, and the diligence of his assistant, Brianne Millett. My readers will benefit significantly from the finishing touches of my copy editor, Nancy Hulan. Renee Redding, of TechBooks, did a first-class job of guiding me through the production process. The superb index is the work of Jake Kawatski. I tip my hat to these five, and I express my gratitude.

The new edition is an improvement over the first in many ways. The material is much better organized, with examples, definitions, and theorems properly identified and displayed.

There are dozens of one-paragraph stories from current—and occasionally ancient—events to illustrate or reinforce the theory. These are displayed in boxes and separated from the main text.

The first edition claimed to be grounded in calculus, but in preparing this edition I discovered that calculus isn't really used that much. Where it was used in the first edition it was often employed to maximize a quadratic function, and this can be done perfectly rigorously using high school algebra. (See Section 1

xii Preface to the Second Edition

of Chapter 2 on maximizing a quadratic function.) For instance, to work out the symmetric equilibrium bidding strategy in a two-person, first-price auction with values distributed uniformly, one simply has to maximize a quadratic function. Where calculus *is* used I have identified the relevant section with the ∂ symbol (even though there is only one variable).

There is lots of new material, including an entire chapter on auctions, which includes a noncalculus proof of the revenue equivalence theorem. (There is also a simple integral calculus version.) There is a new chapter on matching, with sections on the assignment of advisors to advisees, students to colleges, doctors to hospitals, and students to dormitories. There is now a separate chapter on corporate governance, about half of which is new.

Chapter 7 presents a proof of the Gibbard-Satterthwaite Theorem that is different from the one in the first edition. I now begin by proving the result for two people and three alternatives and then generalize in stages. The hidden action chapter has new sections on resource allocation, marketable pollution rights, incentive regulation of the telecommunications industry, personal bankruptcy, and pay and performance in U.S. public schools. Also, the moral hazard and insurance section contains a new subsection on the binary choice model of moral hazard. The discussion of the savings and loan crisis has been expanded. The hidden characteristics chapter has a new section on two-person exchange (including subsections on dominant strategy equilibrium and Nash equilibrium) and a new section on credit rationing. The bundling and product quality section now includes the simple binary model. Chapter 8 on preference revelation with public goods has new sections on dominant strategy equilibrium and efficiency and incentive compatibility (with subsections on dominant strategy equilibrium and Nash equilibrium).

Chapter 1 has a new illustration of hidden characteristic problems, based on the problem of reducing acid rain at low cost, as well as brief subsections on harboring terrorists and on the invisible hand. Chapter 2 includes new sections on decision making under uncertainty (asset preferences, etc.) and on competitive insurance markets under full information (to establish a benchmark, of course). The discussion of efficiency with quasi-linear preferences (in Chapter 2) is much improved. It includes a very easy—but perfectly rigorous—proof that efficiency is characterized by maximization of total utility if there is no nonnegativity constraint on consumption. Chapter 10 briefly considers why the Industrial Revolution did not first take root in China during one of its periods of great inventiveness.

I dedicate this book to my exemplary grandchildren Rosie, Soren, and Edie, ages five, seven, and two. They live seven hundred miles away but the bond with my wife and me couldn't be stronger. I salute their mom and dad, Samantha and Tyler, whose "attachment parenting" has produced extraordinarily happy, healthy, creative children who are a joy to be with.

1.	Asymmetric Information							
2.	Taxi!							
3.	Acid Rain							
4.	Efficiency							
5.	Equ	ilibrium	30					
	5.1	Dominant strategy equilibrium	31					
	5.2	Nash equilibrium	32					
	5.3	The invisible hand	33					
	5.4	The incentive to harbor terrorists	34					
	5.5	Dissolving a partnership	35					
	5.6	The centipede game	37					
	5.7	Subgame-perfect Nash equilibrium Problem set	40 44					
6.	The	Prisoner's Dilemma Game	45					
	6.1	Economic sanctions	47					
	6.2	Public opinion	48					
	6.3	Pollution	48					
	6.4	Beggar-thy-neighbor policies	49					
	6.5	Disarmament	49					
	6.6	Cartels	49					
	6.7	Hostile takeovers Problem set	50 51					
7.	Rep	etition and Equilibrium	53					
	7.1	Repeated prisoner's dilemma with terminal date	54					

2

7.2	Infinitely repeated prisoner's dilemma	55
7.3	Equilibrium theorem for infinitely repeated games	60
7.4	Terminal date and unknown type	64
	Problem set	70

A successful institution, whether large or small, must coordinate the activities of its individual members. In this book, I examine the incentives at work in a wide range of institutions, to see how—and how well—coordination is achieved by informing and motivating individual decision makers. Incentives work well when they result in a high level of individual welfare generally. This is problematic because each individual acts to maximize his or her individual payoff, regardless of the implications for the welfare of others. In other words, we examine incentives to determine the extent to which they prevent the pursuit of self-interest from being self-defeating. We look at an entire economy, as well as a single firm in that economy. Even two-person institutions receive attention: a car owner and a mechanic hired to repair the car, for instance. In all cases, a satisfactory outcome requires coordination among the participants, and coordination requires *information transmission* and *motivation*, as shown in Table 1.1.

The individual members of the institution cannot do their part unless they receive information telling them what their roles are. In the case of a market economy, much of the vital information is transmitted by prices. In a wide range of situations, the consumer's budget constraint and the firm's profit motive give the respective decision maker the incentive to use the information embodied in prices in a way that enhances the welfare of all households. However, in many significant political and economic interactions, the relevant information has been received by individuals but they have no incentive to use that information in a way that enhances the welfare of others. If everyone chooses a strategy that benefits himself or herself a little and harms others a lot, the outcome will leave everyone with a lower payoff than the system is capable of delivering. For instance, each individual in a town knows that everyone can benefit from an Independence Day fireworks display. But there is no incentive for anyone to use this information about the spillover benefit in deciding whether to finance the display. In most towns, no individual would gain by watching fireworks if that person also had to pay the entire cost. If the decision were left to the market system there would be no fireworks. This is typically not a good outcome. If the display would cost \$100,000 and there are 50,000 townspeople, then the fireworks spectacular could be produced by having each person contribute \$2. In most towns, everyone would be better off if he or she gave up \$2 to watch a fireworks display. Although everyone knows that there would be a high level of total benefit from the display, no individual has an incentive to act on that information. The economic theory of incentives is devoted in part to the design of mechanisms that give the decision maker an incentive to use information about spillover benefits.

In rare cases there is a natural alignment of the incentives of the decision maker and the rest of the community. For instance, the pilot of an aircraft is just as determined as the passengers to arrive safely at the destination.





However, the welfare of an airport security guard or mechanic on the ground is not *directly* linked with that of the passengers. The passengers need to be reassured that the mechanic, say, has a strong incentive to act *as though* his or her chief concern is the passengers' well-being. With inappropriate

In 1979 all DC-10 airplanes were temporarily grounded after one of them crashed upon takeoff. The crash was caused by a crack in one of the engine attachment assemblies. The crack resulted from the way that the engine was replaced after servicing. It had been reinstalled in a way that was not recommended or even anticipated by the plane's designer. Reattachment was henceforth done with special care (Petrosky, 1992, pp. 95–6). incentives, a mechanic may succumb to the temptation to avoid hard work by doing a superficial job of inspection and repair.

Incentives are obviously of vital concern to air travelers and are worth studying for that reason alone. But they are also vital to society as a whole. Given the decisions made by others, a worker—whether a mechanic or professor or company president—may find it in his or her interest to expend little effort on the job while drawing a full salary. If a large fraction of the labor force can get away with shirking, the economy's output of goods and services will be greatly diminished and per capita consump-

tion will be very low. In that case, each worker will wish that *everyone* had been prevented from shirking, to enable each to consume less leisure but more produced goods and services. *The pursuit of self-interest is self-defeating in this case*. A more appropriate system of incentives could have prevented this—making everyone better off, even though each individual is maximizing his or her own welfare *given the decisions of others* when everyone shirks as a result of poor incentives.

Appropriate incentives are crucial to the success of any institution, whether large or small. This book examines incentive environments and evaluates each in terms of its ability to promote individual welfare generally. In most cases, the pursuit of self-interest can lead to a high level of individual welfare generally only if the individual taking action incurs a cost equal to the cost that his or

her action imposes on the rest of society. We refer to this as *social cost pricing*. Here is an informal explanation of why social cost pricing works: Let U_i be the payoff (or utility) to individual *i*, who will act to maximize U_i . This will

Four hundred people died in January 1996 when the Indonesian ferry *Gurita* sank. The boat sailed even though the captain knew that the cement that had been used to patch holes in the hull had not dried. A government official had ordered the captain to sail or lose his job. typically affect the payoffs of others, and we let C_i be the total decline in the payoffs of everyone but individual *i*, resulting from *i*'s decision. Then C_i is the cost that *i* imposes on the rest of society. We modify the rules of the game so that the payoff to *i* is now $U_i - C_i$, which is what individual *i* will now maximize. But the change in $U_i - C_i$ is obviously equal to the change in the sum of the payoffs of everyone in society,

including individual *i*. By imposing a cost on individual *i* equal to the cost that *i*'s actions impose on the rest of society, we induce individual *i* to act to maximize the total social payoff, even though *i* is only directly interested in maximizing his or her own payoff.

DEFINITION: Social cost pricing

An institution uses social cost pricing if each decision imposes a cost on the decision maker that is equal to the total cost incurred by the rest of the group as a result of that decision. If there is in fact a net benefit realized by everyone else then the decision maker receives a reward equal to that net benefit.

In many situations individuals must be sheltered from uncertainty if high levels of individual welfare are to be achieved. Full social cost pricing then would leave maximum exposure to risk or uncertainty. In other words, in the presence of uncertainty, incentives have to be less than fully efficient, to allow for insurance.

We look at incentive schemes currently in use, and we also consider the prospects for designing superior schemes in particular situations. The starting point is the realization that, although the decision maker's actions affect the welfare of a wider group, the decision maker has private information that is not available to members of that wider group-nor to a representative of the group, such as a government agency-and that the decision maker will act to maximize his or her payoff, without taking into consideration any resulting side effects on the other members of the group. For example, the manager of a factory has much better information about the production process and product quality than the firm's consumers or the residents of the neighborhood in which the factory is located. If the government attempts to regulate the firm-to affect product quality or the emission of toxic waste-it can do a much better job if it taps the manager's private information instead of issuing direct commands. If the government orders each factory to modify its production process in the same specific way, it may achieve the desired level of pollution abatement. However, it will usually be possible to achieve the same pollution reduction at a lower total cost in resources that have to be diverted from other uses by having the individual factories adjust in quite different ways, depending on their specific input

requirements and production technologies. Doing so requires the provision of incentives to harness the factory manager's self-interest and inside information. We refer to this as *incentive regulation*, and it is coming into increasing use, replacing the old *command and control* approach.

Transmission of information goes hand in hand with incentives. Market prices have their limitations as conduits of information, but they do a superb job in a wide range of situations. For example, wages are important determinants of individual career choices, and wages contain information about the value of various skills to all consumers. An occupation will command a high wage if it contributes significantly to the production of highly valued goods and services. That's because the high demand for a consumer good translates into high prices and profit for the producer. There will be great demand for workers who are crucial to the production process because they generate substantial revenues for their employers. The high demand for these workers leads to high wages. Competitive bidding in the labor market raises the wage of the most productive workers above that of other workers. A particular wage signals information to the economy as a whole concerning the value of the associated skill. We not only acquire the information that a particular occupation is valuable to consumers as a whole; at the same time, an individual has a strong incentive to take this information into consideration in choosing a career, because higher wages provide more income and thus more consumption opportunities.

In general, the way prices enter our budget constraints gives us the incentive to use the information embodied in those prices. All individuals maximize their own payoffs, but because the prices embody information about the welfare of others, the pursuit of self-interest induces individuals to take the welfare of others into consideration, without realizing that they are doing so.

Information transmission and motivation do not always go hand in hand. Commuters know that traffic is congested during rush hour. If individual driver A joined a car pool, other drivers would benefit from the reduction in the number of cars on the road. But the benefit to A is slight, and A's own welfare would decrease because of the inconvenience of not having his or her own car. Selfinterest leads all motorists—well, almost all motorists—to drive their own cars to work. It's plausible that if everyone joined a car pool the improved traffic flow would leave everyone better off, *net* of the inconvenience of carpooling. As it is, everyone knows about the social value of carpooling but no one has an incentive to act on that information. However, information technology now allows municipalities to charge for the use of designated high-speed lanes. Such lanes remain uncongested because their user fee gives motorists for whom time is relatively less valuable the incentive to use the lanes that are free but more crowded.

Information transmission can be more or less costly. Low-cost information transmission is problematic. If the institution is the entire economy, the delivery of information throughout the economy can be exceedingly costly. For one thing, contracts must be enforced, and legal costs can be very high. Prices transmit information at low cost but, as Table 1.1 indicates, other devices such as warranties are important. An extensive warranty on a manufactured "More than 2000 television sets a year exploded in Moscow alone" before the collapse of the Soviet Union (Milgrom and Roberts, 1992, p. 13). good is a signal that the manufacturer believes that the likelihood of a defect is small. If an entrepreneur set out to deceive customers by manufacturing low-quality television sets and passing them off as high-quality sets, he

could not offer a good warranty without losing the profits that his deception was designed to yield. He would know that a very high number of sets would be returned for replacement or repair under the warranty. Competition between manufacturers in a private ownership market economy induces each producer to make a high-quality appliance and offer an extensive warranty.

Even when the information transmission problem is solved, the motivation problem remains. As with the highway congestion example, there must also be an incentive for the individual to use the information in a way that promotes the goals of the institution—a high level of welfare by commuters generally, in the case of the traffic example. Incentives are essential because individuals' paramount concern is their own welfare, not the welfare of others. This book is devoted to the study of material incentives—incentives that have their impact on the decision maker's welfare through their impact on his or her consumption opportunities. How can they be designed to harness self-interest and prevent the pursuit of self-interest from being self-defeating?

An automobile repair shop illustrates nicely how incentives will come into play in this book. A car owner who brings his car to the shop for repair wants a reliable job done at low cost. He has neither the expertise nor the time required to monitor the mechanic. If the car owner suspects that the mechanic has cut corners he is likely to broadcast his suspicions to acquaintances. This implicit threat, along with the existence of other repair shops competing for business, gives the owner of a garage some incentive to ensure that the repairs are well done and that customers are not overcharged. But how does the garage owner motivate the mechanic that she employs? Competition and reputation effects may give the right incentives to the owners of firms, but they are just part of the solution. The owner—in general, the *principal*—now has the problem of providing appropriate incentives to the *agents* (mechanics) that she hires. We attempt to solve this problem—with considerable success. The private ownership market economy is very sophisticated when it comes to generating devices for solving

In World War II the United States won the race with Germany to develop the atomic bomb. Computers were not available, of course, and the United States depended on a team of high school graduates to do a staggering amount of calculating. The productivity of the calculators increased almost tenfold when they were told what they were working on (Gribbin and Gribbin, 1997, p. 97).

these principal-agent problems. But there are serious limits to the ability of *any* institution to overcome incentive difficulties in many situations. The difficulties are compounded by the presence of random effects. If the car breaks down a week after it was repaired, should that be attributed to shirking on the part of the mechanic or to bad luck?

Although this book is almost exclusively concerned with material incentives, we acknowledge that nonmaterial incentives play

a role in any institution. In one of the first influential articles on the modern economics of information, Kenneth J. Arrow (1963b) noted that the information advantage possessed by physicians in treating their patients has led to the emergence of institutions based on trust and delegation to supplement market incentives. Hence, the code of medical ethics.

Each of us does things that benefit others, at some personal sacrifice. Nevertheless, we employ a model that assumes that each individual always pursues

As early as 1931 the Soviet ruler Joseph Stalin deviated from the egalitarian wage ethic, realizing that a high level of economic performance could not be achieved without material incentives. The opportunity to work for the common good was not sufficient motivation (Laffont and Martimort, 2002, p. 23). narrow self-interest. One reason for doing so is that we are alerted to potential difficulties if our model results in low levels of individual welfare generally. Moreover, we are much less likely to recommend policies that are naively utopian when we work within this framework.

The importance of incentives has been documented in many ways and in many contexts, although the specific contractual form derived from economic theory is not always reflected

in contracts as actually written (Chiappori and Salanié, 2003). For the specific case of the relationship between a tenant farmer and the landowner, Allen and Lueck (2002) show convincingly that incentives are central to understanding the nature of the contracts that are employed.

In 1896 South Carolina enacted a law levying a fine on any county in which a lynching took place. No county that had been fined for this abuse ever had a second lynching (Dray, 2002). Lynching of African Americans by white mobs was common from the late nineteenth century until the middle of the twentieth and was one of the many devices by which African Americans were terrorized. One measure of the importance for public policy of a formal study of incentives is provided by McAfee and McMillan (1988). They estimate that switching to appropriate contract design could reduce government costs by at least 8%, and sometimes by as much as 30% (p. 149). The switch to the *responsibility system* in Chinese agriculture in the 1980s resulted in a remarkable increase in productivity over a short period of time (McMillan, 1992, pp. 96–8). The responsibility system requires each farm to deliver a fixed amount of output to the state, but

the farm keeps the proceeds of all output above this quota. This is an example of social cost pricing: The social cost of the farmer's leisure consumption is the output that society loses when the farmer consumes an hour of leisure. But that is also equal to the cost imposed on the farmer under the new system because the farmer would have been allowed to keep the harvest from that hour of labor. Under the old system, the cost to the farmer of an additional hour of leisure consumption was zero because all of the output from an additional hour of labor goes to the state. It was the farmer whose return was fixed.

Another reason why we assume selfish behavior at every turn is that, although it abstracts from important features of the real world, it gives us a simple model with a lot of explanatory power. We have come to accept abstract models in everyday life and should not be reluctant to employ them in economics. A road map, for instance, is a representation of a particular region. It abstracts from

almost everything that is important—scenery, the location of shops, and so on. Because it *is* so abstract it is very easy to use to work out a route from one location to another; it can even be used to compute a short route. Similarly, an economic model can be exceedingly abstract and still allow us to determine the effect of an excise tax on a commodity's price or the nature of a salary contract that will be offered when the employer can observe the quality of the employee's work but cannot validate that observation with evidence that would be credible to a third party, such as a judge. Conclusions are drawn from abstract, formal economic models via theorems.

Many people are impatient with economists for abstracting—and worse, employing assumptions that are at odds with reality. It may comfort you to know that this is standard practice in physics. It can even be useful for a physicist to assume that a cow is a sphere! (See Krauss, 1993, pp. 1–7.) "The set of tools physicists have to describe nature is limited. Most of the modern theories you read about began life as simple models by physicists who didn't know how else to start to solve a problem....*Before doing anything else, abstract out all irrelevant details*!...Overcoming the natural desire *not* to throw out unnecessary information is probably the hardest and most important part of learning physics" (Krauss, 1993, p. 4). The classical model of the motion of the planets around the sun assumes that the mass of each planet is concentrated at a single point of zero breadth. That's absurd. Nevertheless, the model is extremely useful. It was used to predict the existence of the planet Pluto, for example, which was discovered in 1930.

We begin then by assuming that all individuals evaluate outcomes exclusively in terms their effect on their own well-being. This allows us to work

Public drunkenness is not uncommon in Japan, but drunk driving is very rare because of the severe penalties. A professional person can even be disqualified from practicing if convicted of driving while intoxicated. out an individual's response to a change in the incentive environment. The assumption of selfish utility maximization implies that there *will* be a response. Not everyone is able to grasp this point. For example, a lot of people argue against long prison sentences for drunk drivers who kill or maim others: "It could happen to anyone." Well, wouldn't you make sure that it

couldn't happen to you if a long prison sentence were the penalty for drunk driving? To adapt a phrase of Dr. Johnson's, the prospect of a long jail sentence focuses the mind wonderfully.

We examine incentives at work to see whether we can expect outcomes that maximize individual welfare generally when individuals are motivated by selfish considerations. In each case study we assume that an individual takes whatever available course of action leads to the highest possible personal benefit for himself or herself. Of course, in real life there are situations in which some or all individuals behave altruistically, at least up to a point. But self-seeking behavior is pervasive enough to warrant independent study, particularly when the economy as a whole is our concern. Therefore, our goal is to work out the implications of self-motivated behavior, by means of examples and theorems, and we try to learn from them without being distracted by the many real-world features that

are left out of the models. We discover that the need to provide individuals with socially beneficial incentives imposes constraints on the economic system as a whole, forcing us to make trade-offs. For instance, giving individuals an incentive to truthfully reveal their preferences for public goods leads to government budget imbalance. By identifying such trade-offs we can design better public policies. In particular, we won't waste resources trying to accomplish goals that are mutually exclusive.

Links

McMillan (2002) is a superb but non-technical account of how, and to what extent, markets can provide the incentives that lead to a high standard of living. The role of the CIA in supplying the Bush administration with evidence of Iraq's weapons of mass destruction, prior to the invasion of March 2003, is a reminder that the performance of an organization is a function of worker and management incentives. See *The Economist*, July 15, 2004 ("The weapons that weren't"). Baumol (1993) contains many examples of entrepreneurial responses to incentives, some of which reach back to ancient Greece and Rome. See Sappington (1993), Laffont (1994), and Sappington and Weisman (1996) for further discussion of incentive regulation. Stiglitz (1993) has a good discussion of the limits of prices in transmitting information. See Chapter 4 of Baumol (1993) for examples of the costs of contract enforcement.

Problem set

- 1. The services of garbage collectors have far more total value to the community than the services of heart surgeons: Compare a world without garbage collection—plagues, low life expectancy, only 50% of children surviving to the age of five—to a world without heart surgeons—no appreciable difference in life expectancy. But heart surgeons are paid far more per hour than garbage collectors. What information is being signaled by this wage rate differential?
- 2. I drive a car made in 1990, before air bags became mandatory in all cars sold in the United States. I could buy a safer car—a new Mercedes Benz, for example—but I prefer a basket of goods and services that includes my present car and an annual vacation on the ocean to a basket with a safer car but an annual vacation consisting of croquet in the backyard. Is it in society's interest for firms to devote enough resources to the production of consumer goods to ensure that there is absolutely no chance of a defective product injuring someone?
- **3.** A barber will not stay in business long if he gives bad haircuts. Competition among barbers ensures that each attempts to build a reputation for high-quality service. What about an industry in which problems do not show up until long after the commodity has been purchased—housing construction, for instance? Is there a role for some form of government regulation in these cases?

ASYMMETRIC INFORMATION

When you hire a taxi you are employing an *agent* to carry out an assignment. You, the *principal*, want to get to your destination quickly and at low cost, but the taxi driver wants to maximize his revenue. The driver appears to have an incentive to overcharge, and your ability to monitor this is very limited because you know very little about traffic patterns and expedient routes, especially if you are a visitor to the city. This is an instance of a *hidden action* problem. The passenger cannot directly determine if the driver has acted in a way that minimizes the travel time.

DEFINITION: Hidden action problem

A principal hires an agent to carry out a task, but it is impossible or extremely costly for the principal to monitor the agent.

In Section 2, we demonstrate that the conventional taxi fare schedule induces the driver (the agent) to choose the route that the principal (the passenger) would select if the principal had as much information about routes and traffic patterns as the agent-even though the principal in fact has very little information, and the agent knows it. In general, we investigate the possibility of providing appropriate incentives to agents to induce them to behave in the way the principals would instruct them to act if the principals themselves possessed the relevant information-even though the principal is in fact unable to monitor the agent, and the agent knows this. There are three reasons why the principal may want to employ an agent: The agent may possess a skill that is particularly appropriate to the task at hand. (I hired a specialist to remove a tree that had fallen over my driveway during a storm.) The principal may not have the time to carry out the task herself. (I sometimes eat in restaurants where the chef is not as good a cook as I am.) Finally, even if the principal and the agent are "twins," economies of scale can justify the delegation of some tasks by one to the other.

Providing the agent with an incentive that is optimal from the standpoint of the principal requires us to choose the incentive scheme that maximizes the principal's payoff subject to constraints. These constraints embody the notion that agents will act to maximize their payoffs subject to the incentive scheme governing their behavior, and the notion that agents have alternative job opportunities and hence must do at least as well working for principals as they would in the next best alternative. These constraints on principals' choices of incentive schemes result in principals achieving lower payoffs than if principals had all of the information available to agents and could simply instruct agents to carry out the actions that maximized the principals' payoffs. In some cases, the principal's loss is the agent's gain, but in other cases the constraints result in a net loss to the principal-agent duo. If there were no asymmetry of information, the

1. Asymmetric Information

principals would offer the agents a contract that specified precisely what task was to be performed and how and when it was to be carried out. A court could easily determine if the conditions had been met.

A second family of hidden information problems concerns the attempts by a principal to elicit a specific piece of information that is known only by the agent, but which affects the principal's welfare. For example, the principal is the owner of an asset that is up for sale. If the owner knew the maximum amount that each potential buyer is willing to pay, the current owner could offer the asset to the individual with the highest willingness to pay at a price that is just below that value. That would clearly maximize the owner's return from the sale. For that very reason, all potential buyers have incentive to conceal their maximum willingness to pay, which is the hidden information in this case. This is called a *hidden characteristic* problem. In Chapter 6, we show that it is possible to design an auction that motivates the bidders to reveal their true willingness to pay.

DEFINITION: Hidden characteristic problem

Information possessed by one individual (or firm or institution) is concealed from everyone else, but the welfare of others depends on that information.

In many cases we employ a fictitious principal, usually a surrogate for society as a whole. Maximizing the principal's payoff is then a metaphor for maximizing consumer welfare generally, subject to the limitations imposed by resources and technology. The goal is to provide individuals and firms with an incentive to disclose their private information-specifically, individual preferences and firms' production recipes. In Section 3, we show that a firm can be given an incentive to disclose how much it would have to pay to reduce the amount of pollution it generated. In this case, the firm's true pollution abatement cost is the hidden characteristic. The incentive scheme induces the firm to reveal its true cost, even though the firm with the lowest cost will actually have to make the adjustment. The fact that information is hidden from the principal reduces the principal's payoff. In the case of the pollution example, the principal is a metaphor for general consumer welfare, and the payoff reduction is the money that has to be paid to firms to induce them to reveal their costs truthfully. (Pollution is reduced in the end, so there is an overall net gain for consumers. But the gain would be even larger if the government knew what the individual firm costs were.)

There are many types of hidden characteristics. Here are some examples:

An individual's preference scheme, or some statistic based on that preference scheme; the marginal rate of substitution at a point (elicited by some resource allocation mechanisms); the elasticity of demand (elicited by a price-discriminating producer).

- The probability that an automobile driver will have an accident. This information is sought by an insurance company. The probability affects the driver's preference scheme via the expected utility function.
- A voter's most-preferred candidate (required by the plurality rule voting mechanism).
- The cost that would be incurred by a firm if it were to reduce its pollution output by 15%.
- A firm's production function.

Here is an interesting example with a global perspective. Worldwide reduction of carbon dioxide (CO_2) emission is advocated by many as a way of slowing global warming. One widely supported policy would require each country to pay a tax on CO_2 emissions greater than a specified quota. A country's quota would be a fraction of its current CO_2 emission rate, with the fraction determined by an international committee. The current emission rate is the country's hidden characteristic. It is naive to assume that each country would report its true emission rate if it could be assigned a higher quota by reporting a higher emission rate.

The owners of baseball franchises have become adept at hiding the team's profit to make it easier to deal with the players' union (Zimbalist, 2004, Chapter 4). All of the models and examples discussed in this book can be placed in either the hidden action or the hidden characteristic category, and some have elements of both. These two categories constitute the family of *principalagent* models. The principal is the individual whose welfare is to be served, and this welfare

is affected by an agent who makes decisions on behalf of the principal. The principal knows that the agent will choose a course of action that maximizes the agent's own welfare. But the principal may be able to provide the agent with incentives that cause the agent's welfare to reach its maximum when the agent takes the action that leads to the maximization of the principal's welfare. This is problematic because the principal cannot observe the agent's action or cannot determine if the agent has acted appropriately.

In other situations the principal's welfare depends on the agent's characteristic, which cannot be observed or even verified by the principal. There are two ways in which the principal's utility can depend on agent characteristics: In general equilibrium resource-allocation models the principal is an abstract planner whose utility is identified with social welfare. Social welfare in turn is a function of the characteristics (preferences and technology) of the economy's agents (consumers and firms). We examine a special case of this in Section 1 of Chapter 3. In Chapters 8 and 10, we investigate this subject thoroughly. In more narrowly focused models the principal may be an insurance company, for instance, and the company's profit depends on the number of claims submitted, and that in turn is a function of the probability that a policyholder has an accident. The potential policyholders are the agents in this case. The agent cannot be relied on to act in the principal's best interest—either to take the appropriate action or disclose the agent's characteristic—because the agent wants to maximize his or her own utility. We see whether and to what extent the agent's self-interest can be harnessed by a judicious deployment of incentives that induce the agent to act in a way that promotes the principal's welfare. In hidden characteristic models, the incentive structure will be deemed a success when the agent's action reveals the agent's characteristic.

The hidden action and hidden characteristic phenomena are often called *moral hazard* and *adverse selection* problems, respectively, echoing insurance terminology. I prefer the terms hidden action and hidden characteristic, which are more appropriate for economic applications. In this book, we apply the term *moral hazard* to situations in which there is a hidden action problem that is not handled successfully. Similarly, we use *adverse selection* to refer to welfare losses due to a hidden characteristic problem.

Hope Scholarships, proposed during the 1992 U.S. presidential campaign, would allow students to borrow money for college and then pay back a specified fraction of their *incomes* after graduation. This plan would be plagued by adverse selection. Students expecting to go into high-income occupations after graduation would opt for conventional loans, leaving only those heading for low paying—but perhaps socially valuable careers to apply for Hope Scholarships (Wheelan, 2002, p. 81). Hidden information problems are everywhere. What guarantee do you have that your instructor devotes a reasonable amount of time to designing the course, preparing lectures, and grading tests? Surely there is a temptation to increase leisure time and reduce preparation time or to substitute consulting activity—or research activity, in general—for lecture preparation. In this case the student is the principal and the instructor is the agent. This is clearly a hidden action problem. The committee that *hires* a new faculty member has a hidden characteristic problem, and the characteristic in this case is the prospective employee's quality. Here are two more examples: Business travelers

sometimes choose unnecessarily expensive flights (paid for by their employers) to get higher "frequent flyer" bonus points, which are then applied to personal travel. (Can you identify the *social* waste in this case?) The United States federal student loans program involves billions of dollars. Private contractors are hired to collect student debts, and some of the collecting companies are financially tied to firms in the profitable secondary loan market, giving the collectors an incentive to allow students to default on the original loans (*Washington Post*, June 19, 1993, p. 2).

Moral hazard can *create* adverse selection! A retail store that did not monitor incoming cash would give employees insufficient incentive to be careful with that cash. This would also invite unscrupulous people to apply for work at this store, in the expectation that they could embezzle easily.

An adverse selection problem can be so severe that the market can disappear completely. Consider the viability of unemployment insurance if it were to be provided by the private ownership market economy. It would be costly to purchase, so individuals who know that the likelihood of their becoming unemployed is low would not buy it. This would result in a higher number of claims "Your pizza's free if you don't get a receipt." Sales receipts make it much easier for a store owner to monitor incoming cash. But they only work if the employee issues a receipt. By giving free pizza if there is no receipt, the owner gives the customer an incentive to monitor the employee working the cash register. per insured worker, leading to an increase in premiums to enable the insurance companies to offer unemployment insurance without taking a loss. This would lead to more individuals opting out—those who were willing to buy when the premium was lower but who feel that the probability of their being unemployed is not high enough to justify paying the slightly higher premium. As the premium increases, it

is always the low-probability individuals within the group of previously insured workers who discover that it is now rational for them to cancel their insurance coverage because of the increase in the premium. This means that the number of claims per insured person will rise after an increase in premiums, resulting in another round of premium increases. The whole market can unravel in this way. And if that's the case, and protection against the risk is socially desirable—that is, provides net benefit to workers generally—there is a case for provision by the public sector.

A democratic political and legal system also exposes its participants to risk. I might be formally charged with a crime that someone else committed. Part of the benefit of a democracy has to do with competition for political office, and the consequent realization of incumbents that they will be punished by defeat at the polls (or worse) if too many constituents are falsely accused of crimes. However, it is in our interest as law-abiding citizens to have arrests made *before* the authorities are perfectly certain that they have identified the culprit. If they waited until they were certain there would be too few arrests and too much crime,

Cardozo Law School's Innocence Project uses DNA evidence to determine the culpability of U.S. defendants convicted (primarily of rape and murder) before accurate DNA testing became available in the 1980s. By April 2002, 104 inmates had been exonerated-almost two-thirds of the cases examined. (The project is the brainchild of lawyers Barry Scheck and Peter Neufeld. Cardozo Law School is part of Yeshiva University in Manhattan.) Factors leading to the faulty convictions include mistaken eyewitness reports, coerced confessions, police corruption, poor legal representation, prosecutorial misconduct, and inaccurate laboratory work (Weinberg, 2003, pp. 200-1). Do you think that police and prosecutors have too strong an incentive to obtain convictions?

and the arrests that were made would be obtained at too high a cost in resources. (What does "too high a cost" mean? How do we know the cost would be too high?) So, there remains some risk that a law-abiding citizen will be arrested and forced to defend himself in court. Why don't private markets insure against that risk by offering policies that pay legal costs? Legal services don't come cheap. (And legal defense insurance would cause legal fees to soar. Why? Why have physicians' incomes soared over the past few decades in all countries that provide national health insurance, even if it's only to those over sixty-five?) The premium would not be trivial and hence not everyone would purchase insurance. But why isn't some legal defense insurance provided by the market? The adverse selection problem is quite evident here. The individuals who are most willing to buy this policy would be those who know themselves most likely to be in hot

1. Asymmetric Information

water. This means that the premium would be higher than if everyone in the community purchased a legal defense policy. But, the higher the premium the higher the percentage of lawbreakers among the policyholders. There is no premium at which the claims paid out could be covered by the premiums paid in, and the market breaks down. (Private legal defense insurance is available in the United States—mostly in group form—but it does not provide significant coverage for criminal cases.)

Can a case be made for public provision of legal defense insurance as with unemployment insurance? Probably not. Whether the insurance is provided by the public or private sector, there is a severe moral hazard problem. This doesn't apply to you or me, but a lot of people would increase the scope of their criminal activities if they knew that any necessary legal defense would be funded by taxpayers or holders of private insurance policies. There would be such an increase in the demand for the top spellbinding courtroom orators that their fees would increase and then so would the flow of students into law schools. This waste of resources is perhaps the least of the antisocial effects of the provision of legal defense insurance, a commodity that would substantially increase individual utility were it not for the moral hazard and adverse selection problems.

Sources

The theory of principal and agent is now central to economic theory. K. J. Arrow (1984) proposed the terms hidden action and hidden information as substitutes for the terms *moral hazard* and *adverse selection* in widespread use. (We refer to hidden information as hidden characteristics.) Arrow (1963b, 1971) was the first to draw attention to the economic significance of moral hazard, called hidden action throughout this book. The modern theory of principal and agent was introduced in Ross (1973) and Stiglitz (1974) and given its modern expression in Mirlees (1999), which debuted in mimeograph form in 1975. The optimal income tax problem, a special case of the principal-agent model with the tax authority as the principal and taxpayers as the agents, was proposed by Vickrey (1945) and examined by Mirlees (1971). The pioneering articles by Akerlof (1970) on the used-car market and Spence (1973) on education as a signal of worker quality are credited with turning the attention of the profession to hidden characteristic problems. Mirlees and Vickrey shared the Nobel Prize in 1996, and Akerlof, Spence, and Stiglitz shared the prize in 2001. K. J. Arrow, considered by many to be the most significant economist of the twentieth century, received the Nobel Prize in 1972.

Links

Stiglitz (2000) and Chapter 1 of Laffont and Martimort (2002) outline the history of the treatment of asymmetric information in economic theory. The former emphasizes the ways in which earlier theory was misleading because of failures to acknowledge problems caused by asymmetric information, and the latter highlights the ways in which modern information theory was anticipated.

TAXI!

2

You have just landed at the airport in a city that you are visiting for the first time. You hail a cab to take you to your hotel. How can you be sure that the driver chooses the quickest and cheapest route to your destination? You can't, unless you make an investment beforehand; an investment of money to purchase a map and of time to compute the shortest route between your departure point and your destination. Even then, you will not know which streets are normally congested, so it would be very costly to discover the cheapest route. Assuming that you are not prepared to incur that cost, is there any way of ensuring that the taxi driver will not take you out of your way to enhance his or her income at your expense? We need to find a way of providing the driver with an incentive to choose the least-cost route, so that even though you don't know what that route is you will be sure that the driver has chosen it because that choice maximizes the *driver's* return from operating the cab. This is the purpose of the fixed part of the nonlinear pricing schedule for taxi rides. The fare is F + cDwhere D is the distance to your destination in miles, c is the charge per mile, and F is the fixed initial fee which is independent of the length of the ride. (In fact, you will be charged for time spent idling in traffic, but let's keep things simple.)

If *F* is zero, and hence the fare is *cD*, then the driver has a strong incentive to make each trip as long as possible. That's a consequence of the fact that when passengers are dropped off at their destinations, it takes the taxi driver time to find a new passenger. On one hand, from the driver's standpoint, it would be better to keep the meter running by keeping the original passenger in the cab, and that requires taking a much longer route than necessary. On the other hand, if the fixed fee is relatively large—say \$3.00 when the average variable cost per ride is \$6.00—then the driver has a strong incentive to maximize the number of trips per day. But maximizing the number of trips per day can be accomplished only by making each trip as short as possible.

Example 2.1: The linear fare induces shirking

F = 0 and c = 1. Hence the fare is equal to *D*, the distance of the trip. To simplify, each trip is 5 miles long when the taxi driver takes the short route, and the long route is 10 miles long. The driver can make 30 trips a day of 10 miles each or 55 trips a day of 5 miles each. (Remember, time is lost between trips.) When the driver works efficiently her revenue is $55 \times \$1 \times 5 = \275 . But when the driver shirks, and takes the long route, her daily revenue is $30 \times \$1 \times 10 = \300 : She makes more money by shirking.

The linear fare schedule (with F = 0) gives the agent (the taxi driver) incentive to behave in a way that makes a single ride unnecessarily expensive. It also wastes a valuable resource—time. Both the driver's and the passenger's labor are wasted,

2. Taxi!

and if this fare schedule were used throughout the economy the accumulated waste would be enormous.

When *F* is positive and greater than the value of the time consumed searching for a new passenger, the income maximizing strategy is for the driver to get the passengers to their destinations as quickly as possible: The driver will lose *L* dollars, if we assume that *L* dollars of income is lost while waiting for another passenger, but the next passenger will pay *F* dollars in addition to the variable fee of *c* dollars per minute. This yields a net gain of F - L, compared to the strategy of driving twice as far as necessary on each trip. Of course, any particular trip of distance *D* would cost less if the charge were merely *cD* instead of F + cD, but the fee schedule F + cD results in a lower actual cost to the passenger because it induces the driver to choose a route with the smallest value of *D*.

Example 2.2: The nonlinear fare motivates the agent to perform well

F = 3 and c = 1. Hence the fare is 3 + D. As in Example 2.1, each trip is 5 miles long by the short route and 10 miles by the long route. The driver can make 30 trips a day of 10 miles each or 55 trips a day of 5 miles each. When the driver works efficiently her revenue is $55 \times \$3 + 55 \times \$1 \times 5 = \$440$, but she if shirks her revenue is only $30 \times \$3 + 30 \times \$1 \times 10 = \$390$. The driver's revenue is *lower* when she shirks.

If the driver can make only 50 trips a day when she takes the short route, the revenue would only be $400 = 50 \times 3 + 50 \times 1 \times 5$. That's only slightly more than the 390 a day that she collects when she shirks. But the point is that the nonlinear fare schedule 3 + cD undermines the strong incentive to shirk that is built into the linear fare. The nonlinear fare is an effective solution to the principal-agent problem. The passenger is unable to monitor the driver, and *the driver knows that she cannot be monitored*. Nevertheless, in her own self-interest the driver chooses the action that the passenger would mandate *if* the passenger had the necessary information about the best route.

We said that shirking by taxi drivers wastes both the driver's and the passenger's labor. It is worth noting that the nonlinear fare results in a cheaper ride for the passenger. *Given the length D of the trip*, the fare *cD* is obviously lower than the fare F + cD. However, the latter changes the driver's incentive and, because it eliminates shirking, the passenger actually pays less. For Examples 2.1 and 2.2 the passenger is charged \$10 for a trip under the linear fare $$1 \times D$ but pays only \$8 for the same trip when the nonlinear fare $$3 + $1 \times D$ is used.

Links

It would be interesting to look into the emergence of the nonlinear taxi fare schedule as a response to market forces. It might have been a device introduced

by owners (or managers) of taxi fleets to enhance the performance of their drivers and hence the market share of the company. If it was introduced as a crude device for extracting more money from customers—with the effect on driver performance unanticipated—the company that introduced the nonlinear fare would acquire a reputation for speedy service. This would result in a larger market share and the other companies would likely imitate in an effort to catch up.

Chou (2000); Camerer, Babcock, Lowenstein, and Thaler (2004); Farber (2003); and Sutton (2000, pp. 1–2 and 87–9) provide different perspectives on the taxi industry.

Problem set

- 1. The fare is 3 + D, and each trip is 5 miles long by the short route and 10 miles by the long route. The driver can make 30 trips a day of 10 miles each or k trips a day of 5 miles each. Calculate the value of k for which shirking and minimizing the length of a trip are equally profitable. Now show that shirking is unprofitable for any higher value of k.
- **2.** Each trip lasts *m* miles when there is no shirking and 2m miles when the driver shirks. A taxi does *s* trips per day when the driver shirks and *n* trips per day otherwise. Of course, s < n. The fixed fee is *F* and the charge per mile is *c*. Characterize the values of *F*, *c*, *s*, and *n* for which shirking will not take place.

3 ACID RAIN

Consider an economy that produces electricity primarily by burning coal, discharging sulphur dioxide into the air in the process. Sulphur dioxide (SO₂) dissolves in water to produce sulphuric acid, the principal form of acid rain. Suppose that the government wants one of the firms to reduce its output of SO₂ by 25%, and it's going to choose the firm that can do so at lowest cost. (We are simplifying the story by requiring only one firm to adjust.) The chosen firm will have to make some costly adjustments—purchasing more expensive coal with a lower sulphur content, for example. Other firms might have to install very expensive new equipment to achieve the same reduction in SO₂. General consumer welfare will be best served by selecting the firm that can make the adjustment at the lowest cost. That will minimize the value of resources that have to be diverted from the production of other goods and services to reduce SO₂ emissions.

A firm's adjustment cost is a hidden characteristic. Suppose that the government simply asks each firm to disclose that cost. We'll call this the *naive mechanism*. It will not motivate firms to provide truthful information. Each firm will overstate its adjustment cost by a wide margin, hoping that some other firm will report a lower cost and thus be the one forced to assume the burden of adjustment. We'll not get anything close to truthful revelation of costs, and thus the designated firm could be one of the relatively high-cost companies.

3. Acid Rain

Table 1.2

Firm	1	2	3	4	5	6	7	8	9	10	
Cost	300	120	200	100	150	180	130	300	160	140	

The following *Vickrey mechanism* gives the firm an incentive to report truth-fully.

DEFINITION: Vickrey mechanism

Each firm is asked to report its adjustment cost, and the firm reporting the lowest cost is the one that is forced to reduce its SO_2 by 25%. Then the government compensates that firm by paying it an amount equal to the second-lowest reported cost.

Let's look at this scheme from the standpoint of the individual firm. We see that whatever a firm's true adjustment cost, it can never do better than reporting that true cost, even though no one outside of the firm knows what that cost is.

Example 3.1: Ten utilities with different production processes

The true adjustment cost of each firm is given by Table 1.2.

Firm 4 is the low-cost firm, so it is the one required to reduce its SO₂ output by 25% if everyone reports truthfully. Firm 4's true adjustment cost is 100, and the second-lowest cost is 120, so firm 4 will be paid 120 according to the mechanism's rules. This will give the firm a profit of 20. If firm 4 claims its cost is higher than 120 then the lowest reported cost will be 120, and firm 4 will forego the profit of 20. If firm 4 reports any cost figure below 120 then the outcome will be the same as under truthful revelation: firm 4 will be selected and will be paid 120, which will still be the second-lowest reported cost. Firm 4 cannot benefit by misrepresenting its true cost, but it can harm itself by doing so. *Note: A firm's profit is its revenue minus its true cost*, not revenue minus reported cost.

Let's look at firm 7, with a true cost of 130. If firm 7 reported a higher cost, it would have no effect on its profit—firm 7 would not be selected with 130 or with anything higher. Firm 7 would only be required to make the adjustment if it reported a cost below 100. Suppose, for example, that firm 7 reported an adjustment cost of 90. It would be the low-cost firm, so it would be required to adjust, and it would be paid the second lowest cost—100, reported by firm 4. In that case, firm 7 would actually incur a cost of 130 but would only be paid 100. It would suffer a loss of 30. A firm that does not have the lowest cost cannot benefit from reporting falsely, but it can hurt itself by doing so.



Figure 1.1

Let's apply the Vickrey mechanism to the general case: There are *n* firms and C_j is the true cost of firm j (j = 1, 2, ..., n). No one outside of firm j knows this true cost. We'll show that j cannot gain by misrepresenting that true cost. It is assumed that each firm cares only about its own profit and is indifferent between two outcomes in which it receives the same profit, however much the payoffs of other firms may differ. This is the classic starting point for economic analysis.

Let *D* represent the lowest reported cost among all firms *except j*. Firm *j* can submit its true adjustment cost C_j or not. Suppose that $D > C_j$. Under truthful revelation (i.e., reporting C_j) firm *j* would be required to adjust and would be paid *D*, so its profit would increase by $D - C_j$. What is the effect of a *false* report on *j*'s profit?

There are three subcases to consider, L, M, and H, as suggested by Figure 1.1. On one hand, if firm *j* reports an adjustment cost in region *L*, below C_j , or in region *M* above C_j but below *D*, firm *j*'s profit will increase by $D - C_j$ because either report would be lower than the next lowest cost *D*. In either subcase, firm *j* would be required to adjust, incurring a cost of C_j but receiving a payment of *D* at the same time. That is also what happens with truthful revelation—that is, with a report of C_j by firm *j*. On the other hand, suppose that firm *j* reported a cost in region *H* above *D*. Then it will be not required to make the adjustment. The firm reporting *D* will be selected. Consequently, *j* would forgo the profit of $D - C_j$ that can be realized with truthful revelation. Either a deviation from truthful revelation has the same effect on *j*'s profit as revelation of the true cost C_j , or misrepresentation by overstating cost results in a lower profit than does truthful revelation.

Now, suppose that D, representing the lowest reported cost among all firms other than j, is less than C_j . Under truthful revelation firm j would not be required to adjust, and there would be no change in its profit. Could a false report by j ever be profitable? Again, there are three subcases to consider (Figure 1.2).

Suppose that *j* reports an adjustment cost above *D*, in region *M* between *D* and C_j or in region *H* above C_j . Either way, the firm reporting *D* would be the low-cost firm, and *j* would not be required to adjust. The effect on *j*'s profit would be the same as under truthful revelation. If, however, *j* reports an adjustment cost in region *L* below *D*, then *j* would be the low-cost firm and would have to adjust its SO₂ output. This will actually cost C_j but the firm will only be paid *D*, the next lowest reported cost, resulting in a loss of $C_j - D$. (Profit is always revenue minus *actual* cost. The firm can't reduce its costs by telling a third party that they are lower than they actually are!) If firm *j* had truthfully reported C_j then it would not have been the low-cost firm, and it would not have been required



Figure 1.2

3. Acid Rain

to adjust. Truthful revelation does not result in any change in profit, but an understatement of the true cost results in a loss.

We have examined all the possibilities. Reporting a cost that is different from the true cost can never benefit a firm, but it can be harmful. The Vickrey mechanism provides appropriate incentives for truthful revelation. In contrast, the naive mechanism—ask each firm its cost, select the low-cost firm, which is then required to finance the necessary changes with its own funds—will induce vast overstatement of the costs by the firms. Only by chance would the firm reporting the lowest cost be the one with the lowest true cost. The naive mechanism does not serve the interest of consumers. The Vickrey mechanism works well even when the government authorities have no clue about the true adjustment costs of individual firms. The individual firm has absolutely no incentive to deviate from truthful revelation.

What if firm *j* suspects that other firms will not calculate their profitmaximizing strategies correctly and report costs that are not *their* true costs? Would *j* then have an incentive to deviate from truthful revelation? No! The previous argument, showing that misrepresentation can only harm firm *j* and can never be beneficial, does not depend on an assumption that other firms are reporting truthfully. We demonstrated that whatever these reports are, and *for whatever reason they are submitted*, firm *j* can never profit by misrepresenting its cost.

Source

The Vickrey mechanism is a special case of the Vickrey auction, which is discussed at length in Chapter 6 and which was introduced into economic theory by Vickrey (1961).

Links

To provide clear insight, this section has examined an extreme case: Only one firm will reduce its pollution output by modifying its production process. Section 2 of Chapter 3 presents an incentive scheme that harnesses the individual firm's profit motive in a way that induces the firms to cooperate with each other to determine the assignment of pollution abatement targets to each firm in a way that minimizes the cost to consumers of the total reduction in pollution.

Problem set

- 1. The discussion in this section does not acknowledge the possibility of a tie. If two firms report the same cost, and all other firms report a higher cost, then the tie is broken by flipping a coin. The firm that is selected will still be paid an amount equal to the second-lowest reported cost. When there is a tie, that will be the same as the cost reported by the winner of the coin toss. Prove that it is still not possible for a firm to gain by deviating from truthful revelation.
- **2.** This question concerns the government's attempt to determine which of five firms can reduce its emission of sulphur dioxide by 1000 tons per year at

the lowest cost. For each of the following four schemes the government will require each firm to report its adjustment cost, and it will impose the burden of adjustment on the firm reporting the lowest cost. In each case, determine whether a firm can ever profit by deviating from truthful revelation. If not, explain why. If it is possible, demonstrate that fact with a numerical example.

- **A.** The firm reporting the lowest cost is paid that cost plus 5%.
- **B.** The firm reporting the lowest cost is paid the second-lowest reported cost plus 5%.
- **C.** The firm reporting the lowest cost is paid the average of the lowest and the second-lowest reported cost.
- **D.** The firm reporting the lowest cost is paid an amount equal to 50% of the highest reported cost.
- **E.** The firm reporting the lowest cost is paid the second-lowest cost, and all other firms are paid 10% of the lowest reported cost.
- **3.** This question also concerns the attempt to reduce acid rain by identifying the firm that can reduce sulphur dioxide emissions at lowest cost. Five formulas are given below for determining the payment to be given to the firm reporting the lowest cost. In each case there will be a situation in which one of the firms has an incentive to misrepresent its true adjustment cost. (A "situation" is a specification of the true cost for each firm.) For *Part 1* of your answer you have to present a table of true costs, identify the firm that has an incentive to misrepresent its cost, and give an example of a misrepresentation that will give the firm more profit than truthful revelation. (You won't need more than three firms.) *Part 2* requires you to explain why the specific misrepresentation strategy that you propose in part 1 would *not* be more profitable than truthful revelation *if* the low-cost firm received a payment equal to the second-lowest reported cost—that is, if the Vickrey mechanism were employed.
 - **A.** The firm reporting the lowest cost is paid an amount equal to 150% of the second-lowest reported cost.
 - **B.** The firm reporting the lowest cost is paid an amount equal to the average of the lowest-reported cost and the second lowest reported cost.
 - **C.** The firm reporting the lowest cost is paid an amount equal to the third-lowest reported cost.
 - **D.** The firm reporting the lowest cost is paid \$10 less than the second-lowest reported cost.
 - **E.** The firm reporting the lowest cost is paid \$100, regardless of what any other firm reports.
4. Efficiency

4 EFFICIENCY

We sometimes advocate taxing household *A* to benefit household *B*—"because it results in a more equitable distribution of welfare." Whether it's a good idea or not, it's always possible to change the configuration of production and consumption activities to benefit one individual at the expense of another. Clearly, it is not possible to maximize the utility of everyone simultaneously. How, then, can we formalize the notion that an economic system should maximize consumer welfare generally? By requiring that it exploit every opportunity to benefit some individuals that can be achieved without penalizing anyone else. This is the *efficiency* criterion.

DEFINITION: Efficient and inefficient outcomes

Outcome *A* is efficient if it is feasible and there is no other feasible outcome *B* that gives everyone at least as high a payoff as *A* and gives at least one individual a strictly higher payoff than outcome *A*. An outcome is inefficient if it is not efficient.

An economic system is efficient if it coordinates individual production and consumption activities so well that it uses all opportunities to increase welfare *without* reducing anyone's well-being. If a system is not efficient, then there are equilibria that could be improved to the extent of making some people better off without adversely affecting anyone else. This would be a serious waste because it is extremely costly to identify the individuals in question and to bring about the necessary changes in economic activity. The economic system should not burden public policy makers with this kind of adjustment. The adjustments should be made by the economy itself, before equilibrium is reached.

The efficiency test can be applied to any institution, not only to an economic system. For instance, suppose that there are three candidates, *A*, *B*, and *C*, for a political office, and the community's election rules result in *A* winning even though half of the voters would have been just as happy with *C* and the rest actually prefer *C* to *A*. Then we can say that the result of the election is inefficient, and that the election rules fail the efficiency test. A minimal test of the ability of any system, or institution, to promote individual well-being generally is its ability to eliminate every conceivable kind of waste. There is waste somewhere in the system if it is possible to make at least one person better off without making anyone else worse off. If it is *not* possible to do this we say that the outcome is efficient.

To apply the definition of efficiency we must be able to identify the group of individuals under study and also the set of feasible outcomes. Because the various outcomes are evaluated in terms of the preferences of the members of the group, it is vital to know these preferences. The set of efficiency outcomes can change if we change the group whose welfare is being evaluated, if we change the feasible set, or if we change the preferences of the individual group members.

Example 4.1: The movie or the restaurant

Suppose that the group in question is a family of five who must decide whether to spend a total of \$50 attending a movie (M) or going to a restaurant (R). For this specific group decision problem there are only two feasible outcomes, M and R. If individuals 1, 2, and 3 prefer M to R but 4 and 5 prefer R to M then both outcomes are efficient: A move from M to R would make 1, 2, and 3 worse off, and a move from R to M would make individuals 4 and 5 worse off. Therefore, whatever the feasible outcome, it is not possible to make someone better off without making someone else worse off. Suppose also, that preferences are such that persons 1, 2, and 3 each prefer outcome S to R, where S is obtained from M by taking \$4 from each of the first three individuals and giving \$6 each to persons 4 and 5. Suppose that persons 4 and 5 also prefer S to R. If we change the decision problem so that *S* is now feasible, then *R* is not efficient in this new context: Everyone prefers S to R. Finally, return to the case for which M and R are the only feasible outcomes. Individuals 1, 2, and 3 are the children, and the parents (4 and 5) want the choice of activity to be a function of the children's preferences alone. In that case the relevant group is the set consisting of 1, 2, and 3 only, and hence outcome R is not efficient: Each of the three individuals in the group prefers M to R. But suppose that on a different weekend individuals 1 and 2 prefer M to R but person 3 prefers R to M. If the group is $\{1,2,3\}$ and the feasible set is M, R then both M and R are efficient. We see that a change in preferences can change the set of efficient outcomes.

If there are 1000 individuals in the group, the feasible set is $\{A,B\}$, and 999 people prefer *A* to *B* but the remaining person prefers *B* to *A*, then outcome *B* is efficient, however unfair it might be. Efficiency has to do with the elimination of waste and does not address fairness at all. Consideration of efficiency does not prevent fairness from playing a role: If only 1 of 1000 individuals prefers *B* to *A* then both *A* and *B* are efficient. We can choose *A* from the set of efficient outcomes on the basis of *fairness* or *equity*.

There are two sufficient conditions for efficiency that are easy to apply: First, if an alternative maximizes the total payoff then it must be efficient. We prove this by demonstrating that if an alternative is not efficient then it can't maximize the total payoff. Here's the proof: Suppose there are *n* individuals. We will let $U_i(X)$ denote the payoff (or utility) to individual *i* from generic outcome *X*. Suppose that alternative *Y* is not efficient. Then there is a feasible alternative *Z* such that some individual prefers *Z* to *Y* and no one prefers *Y* to *Z*. In symbols, $U_i(Z) > U_i(Y)$ for at least one *i* and $U_i(Z) \ge U_i(Y)$ for all i = 1, 2, ..., n. The latter implies

$$U_1(Z) + U_2(Z) + \dots + U_n(Z) \ge U_1(Y) + U_2(Y) + \dots + U_n(Y).$$

Because $U_i(Z)$ is strictly greater than $U_i(Y)$ for at least one person we actually have

 $U_1(Z) + U_2(Z) + \dots + U_n(Z) > U_1(Y) + U_2(Y) + \dots + U_n(Y).$

Therefore, Y does not maximize total utility.

4. Efficiency

The second sufficient condition is even easier to apply and defend: If alternative X gives some individual—say, j—a strictly higher payoff than any other alternative then X is efficient. There can't be another feasible outcome that makes one person better off without harming anyone because every other outcome would lower j's payoff. There typically exist efficient alternatives that do not satisfy either of the sufficient conditions, but any alternative that does satisfy one of them is guaranteed to be efficient.

Two sufficient conditions for efficiency:

- 1. A feasible outcome is efficient if it maximizes the total payoff (over the set of feasible outcomes).
- 2. A feasible outcome is efficient if some individual strictly prefers it to every other feasible outcome.

Example 4.2: Three individuals and five feasible alternatives

Table 1.3 gives the utility of each of the individuals 1, 2, and 3 for each of the feasible alternatives *A*, *B*, *C*, *D*, and *E*. *U*₁ is the utility of person 1. *U*₁(*A*) > *U*₁(*B*) indicates that individual 1 strictly prefers alternative *A* to *B*. *U*₂ and *U*₃, the utility functions of 2 and 3, respectively, are interpreted similarly. We know immediately that *D* is efficient because individual 3 strictly prefers *D* to every other outcome. Alternative *C* is efficient because it maximizes total utility: $U_1(C) + U_2(C) + U_3(C) = 125$, which is higher than the total utility from any other feasible alternative. Alternative *B* is efficient, although we can't use either of our

Table 1.3

Alternative	U_1	U_2	U_3
А	25	50	25
В	20	25	60
С	25	50	50
D	10	15	70
E	5	10	60

sufficient conditions to prove it. Moving from *B* to *A* or *C* would harm person 3, and moving from *B* to *D* or *E* would harm person 2. Therefore, starting from *B*, we can't make anyone better off without making someone worse off. Outcomes *A* and *E* are inefficient. *C* gives persons 1 and 2 the same (utility) payoff as *A*, but *C* gives person 3 a higher payoff than *A*, demonstrating that *A* is not efficient. *D* gives each person a higher payoff than *E*, so *E* is inefficient. Note that *A* generates more total utility than the efficient alternative *D*, but *A* is not efficient.

In *economic* models we typically assume that each individual cares only about the direct impact of an outcome on his own welfare, and that more is better.

DEFINITION: Self-regarding and monotonic preferences

Individual *i*'s preference scheme is self-regarding if *i* cares only about the amount of goods and services that he or she consumes. Monotonicity means that *i*'s utility increases if his or her consumption of each good increases.

The simplest economic model requires a cake to be divided among a fixed group of individuals.

Example 4.3: Dividing a cake

There are *n* individuals who are to share one cake. Assume that each person's preference scheme is independent of the amount of cake received by anyone else and that each person always prefers more to less. The feasible set consists of the different ways—*allocations*—of dividing a single cake among the *n* persons. Allocation *x* assigns the fraction x_i of the cake to individual *i*. Of course, $x_i \ge 0$ for all *i* and $\sum x_i = x_1 + x_2 + \cdots + x_n \le 1$. These are the feasibility conditions. Our assumption of self-regarding and monotonic preferences implies that individual *i* will prefer allocation *x* to allocation *y* if and only if $x_i > y_i$.

If $\Sigma x_i < 1$ then *x* is not efficient because we can set $y_j = x_j + (1 - \Sigma x_i)/n$ for each *j*, resulting in an allocation *y* such that $y_j > x_j$ for all *j*. On the other hand, if $\Sigma x_i = 1$ then *x* is efficient because

$$y_i \ge x_i$$
 for all j and $y_h > x_h$ for some h

implies $y_1 + y_2 + \cdots + y_n > x_1 + x_2 + \cdots + x_n = 1$ and thus *y* is not feasible. In short, an allocation $x \ge 0$ is efficient for the division of a cake problem if and only if $\sum x_i = 1$. This means that there are many efficient allocations, and that is typical of almost all economic models. For the division of a cake problem, if all waste has been eliminated, it is still possible to increase someone's utility, but only by transferring some commodity or benefit—"cake"—from someone else. If the transfer is made in a nonwasteful fashion we will have a new efficient outcome. This can be done in many ways, accounting for the large number of efficient outcomes.

Assume that n = 3 for Example 4.3. Then $x = (x_1, x_2, x_3)$ assigns the fraction x_1 of the cake to person 1, x_2 to person 2, and x_3 to person 3. The allocation $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ is efficient and (0.4, 0.4, 0.1) is not. However, both persons 1 and 2 prefer (0.4, 0.4, 0.1) to $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. *Therefore, it is false to say that everyone prefers any efficient allocation to any inefficient allocation*. Of course, there is

4. Efficiency

some allocation that everyone prefers to (0.4, 0.4, 0.1). For example, everyone prefers the feasible allocation (0.42, 0.42, 0.16) to (0.4, 0.4, 0.1).

Note that (0.2, 0.2, 0.2) is not efficient; the allocation $({}^{1}/_{4}, {}^{1}/_{4}, {}^{1}/_{4})$ gives everyone more utility. But $({}^{1}/_{4}, {}^{1}/_{4}, {}^{1}/_{4})$ is not efficient either. So, it's false to say that if y gives everyone more utility than x then y is efficient.

Now, compare allocations $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ and (1, 0, 0). Both are efficient. Therefore, a move from (1, 0, 0) to $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ will make one person worse off. *But it is false to say that the efficiency criterion stands in the way of such a change*. All that we are entitled to say is that there is no efficiency argument justifying a move from allocation (1, 0, 0) to allocation $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. There may be a strong fairness or equity argument for the change, however.

There is a weak version of efficiency that is often easier to work with. Its value lies in the fact that it is an easier definition to apply and that for most economic models the two definitions yield the same set of efficient outcomes.

DEFINITION: Weakly efficient outcome

An outcome is *weakly* efficient if it is feasible and there is no feasible outcome that would make *everyone* strictly better off.

Obviously, an efficient allocation is weakly efficient in general. If everyone can be made strictly better off then it is certainly possible to make one person better off without harming anyone. Consequently, an outcome cannot be efficient if it is not weakly efficient. However, in noneconomic contexts it is possible to have weakly efficient allocations that are not efficient. Consider, for example, a house party with n guests. One may dress casually or formally. Consequently, there are then 2^n outcomes. Assume that no one cares how anyone else dresses so each person is one of two types: C (someone who prefers to dress casually) or F (someone who prefers to dress formally). There is only one efficient outcome, the one that assigns to each person his or her most-preferred mode of dress. Any other outcome has at least one person in his least-preferred attire. This person can be made strictly better off without affecting anyone else and thus the original outcome is not efficient. However, every outcome but one is weakly efficient. Unless each guest is assigned his or her least-preferred mode of dress the outcome is weakly efficient. If at least one person is in his or her mostpreferred attire then that person cannot be made better off so it is impossible to make everyone better off.

In the division of a cake model (Example 4.3) an allocation is efficient if and only if it is weakly efficient. If, say, *y* is feasible, $y_1 > x_1$, and $y_i \ge x_i$ for all *i* set $\epsilon = y_1 - x_1$. Define *z* by setting $z_i = y_i + \epsilon/2n$ for i > 1 and $z_1 = y_1 - \epsilon/2$. Then everyone is better off under *z* than under *x*, and *z* is feasible because $\Sigma z_i < \Sigma y_i$. Therefore, an allocation is not weakly efficient if it is not efficient. In other words, a weakly efficient allocation is efficient in the division of the cake model. In any model, an efficient outcome is weakly efficient. Hence, efficiency and weak efficiency are equivalent for the division of a cake problem.

Table	1	.4
-------	---	----

	Charli	
	Opera	Hockey
Nan Opera Hockey	10,4 0,2	2,1 3,9

We conclude this section by proving that a weakly efficient allocation is efficient in any standard *economic* model. The general proof requires the assumption of a commodity such as money (or cake) that can be divided into arbitrarily small amounts, and which everyone wants more of, and such that each person cares only about his own assignment of that good. We take it as axiomatic that this is possible in any *economic* context.

To show that weak efficiency implies efficiency in an economic model, suppose that feasible outcome *y* makes person 1 strictly better off than *x* and leaves no one else worse off. Construct outcome *z* from *y* by having person 1 give up a small amount of some commodity, money perhaps. Make this amount small enough so that person 1 prefers *z* to *x*. Now divide this amount evenly among the remaining individuals to complete the specification of *z*. Each person likes *y* at least as well as *x*. Thus, with the extra money each person $i \neq 1$ will be strictly better off (at *z*) than under *x*. We already know that person 1 prefers *z* to *x*. Therefore, everyone strictly prefers *z* to *x*. Hence, if it is possible to make one person better off without leaving anyone worse off then it is possible to make everyone strictly better off.

Problem set

- 1. This question concerns two roommates, Nan and Charli, who must decide how to spend their evening. Each prefers being with the other to any outcome in which they attend different events, but Nan likes opera better than hockey and Charli likes hockey better than opera. Table 1.4 displays the payoffs, which can be used to recover their preferences. The first number in each cell is Nan's payoff, and the second is Charli's payoff. List the efficient outcomes.
- **2.** This question concerns a simple economic problem of distribution involving three people, 1, 2, and 3. Specifically, there is a six-pound cake to be divided among the three. Assume that *only* the following five assignments are feasible:

(6, 0, 0), (2, 2, 2), (2, 1, 2), (1, 2, 3), (2, 0, 4).

(The first number is the amount of cake assigned to person 1, the second is the amount of cake assigned to person 2, and the third number is the amount of cake assigned to person 3.) Each individual cares only about his

4. Efficiency

Table	1	.5
-------	---	----

	Kyle's utility	Jackson's utility	Mia's utility
X	1	2	3
Y	2	4	1
Z	3	1	0

or her own consumption of cake and prefers more to less. Of the five specified assignments list the ones that are efficient.

- **3.** Three siblings, Jeremy, Kerri, and Tom, have jointly inherited three assets: *X*, a large house; *Y*, a yacht; and *Z*, a very valuable painting. Each individual must receive one of the assets, so there are six possible assignments of assets to individuals. For each of the following three cases specify the preferences of each of the individuals so that no individual is indifferent between any two assets:
 - **A.** There is only one efficient outcome.
 - **B.** Every outcome is efficient.
 - **C.** There are at least two efficient outcomes and at least one that is not efficient.
- **4.** Christine, Jay, and Christy-Ann have jointly inherited five assets (call them *A*, *B*, *C*, *D*, and *E*). The assets are indivisible—an antique car, a sailboat, and so forth. It is left to the heirs to allocate the assets among them. Therefore, the feasible outcomes are the set of all possible ways of assigning the assets to the three individuals. The individual preferences are as follows:

Christine strictly prefers A to B, B to C, C to D, and D to E.

Jay strictly prefers *E* to *D*, *D* to *C*, *C* to *B*, and *B* to *A*.

Christy-Ann is indifferent between each pair of assets. (If you took one asset away from her and gave her a different one in its place she would be no better off and no worse off.)

Each person gets positive utility from each asset. (If you gave an individual an additional asset, he or she would be better off, whatever the asset.)

- A. List five efficient outcomes that leave Christy-Ann with nothing.
- **B.** List five efficient outcomes that leave Jay with nothing.
- C. List five efficient outcomes that give each person at least one asset.
- **5.** This question concerns a situation in which three roommates, Kyle, Jackson, and Mia, have to choose between the following three alternatives:

X: studying together,

- Y: going to the basketball game together, and
- *Z*: going their independent ways.

	F	G	Н	J	K	М
Cathy's payoff	0	60	200	100	40	205
Vince's payoff	170	60	65	40	110	95

Table 1.6

These three alternatives, and only these alternatives, are feasible. The utility derived by each individual from each alternative is revealed Table 1.5. "Alternative X is efficient yet it does not maximize the sum of individual utilities." Is this statement correct ? Explain.

- **6.** This question asks you to identify the efficient outcomes in a simple model with two individuals, Cathy and Vincent, and six outcomes, *F*, *G*, *H*, *J*, *K*, and *M*. Table 1.6 gives the level of utility obtained by each individual under each outcome. All six outcomes are feasible, and there are no other feasible outcomes. Which of the outcomes are efficient? Suppose that a seventh option becomes available, and it provides utility levels of 206 for Cathy and 172 for Vincent. How would the set of efficient outcomes be affected?
- **7.** There are three individuals (1, 2, and 3) and five feasible outcomes (*A*, *B*, *X*, *Y*, *Z*). Table 1.7 specifies the utility function for each person. List the efficient outcomes. Now, list the weakly efficient outcomes.
- **8.** Return to Example 4.2. Multiply each of the utility numbers for person 3 by 10, leaving the utility numbers of 1 and 2 unchanged. Show that the set of efficient outcomes is unchanged, even though a different outcome now maximizes total utility. This demonstrates that efficiency depends only on individual preference rankings and not on the utility numbers that we use to represent those rankings. To drive this point home, for each individual list the alternatives in order of preference. Now work out the efficient outcomes, using only those rankings. (If two alternatives have the same utility number put them in the same row of your list for the individual in question.)

5 EQUILIBRIUM

Each person has a given set of actions from which he or she is allowed to choose. When each person employs a strategy that maximizes his or her payoff, given the choices made by others, we will be at equilibrium. In most situations the

Table	1.7
-------	-----

A	В	X	Y	Ζ
U_1 1	2	5	4	3
U_2 50	0	100	1	1
U_3 1	1	1	1	1

5. Equilibrium

strategy that is best for individual *A* depends on what individual *B* is expected to do. For instance, in a game of soccer—known as football outside of North America—if *A* has a clear shot on goal, whether *A* decides to kick to the right or the left depends on whether *A* expects the goalie to move left or right.

We begin by examining a special family of games in which each person's best strategy *can* be determined independently of what the opponent is expected to do.

5.1 Dominant strategy equilibrium

This section considers a small but important family of games in which the individual's payoff-maximizing strategy is independent of the strategies that others

Table 1.8

	Play L	ver 2 R
Player 1 H	5.5	0 10
D	10, 0	1,1

pursue. Consider Table 1.8: Player 1 has to choose between two strategies U and D, and Player 2 has to choose between the two strategies L and R. The first number in a cell is player 1's payoff, and the second number is player 2's payoff. On one hand, if person 1 thinks that her opponent will choose L then she'll do better playing D than playing U. When person 2 plays L, player 1 gets 5 by playing U but 10 from D. On the other hand, if player 1 expects her opponent to play R then she'll also

do better playing *D* than playing *U*. The former yields 1 but the latter yields 0 when person 2 plays *R*. Therefore, player 1 should play *D*, whatever she thinks her opponent will do. We say that *D* is a *dominant strategy*.

DEFINITION: Dominant strategy

We say that S^* is a dominant strategy for player A if, for *any* strategy T available to A's opponent, none of A's strategies yields a higher payoff than S^* when A's opponent plays T. (We use an asterisk to distinguish a salient strategy or outcome.)

Notice that a dominant strategy does not necessarily give a player the highest possible payoff. It is not even the case that a dominant strategy gives a player the same payoff for each of the opponent's strategies. *D* is clearly a dominant strategy for player 1 in the game of Table 1.8, but when she plays *D* she will get 10 if player 2 chooses *L* but only 1 if player 2 chooses *R*. The payoffs are quite different. But *D* is a dominant strategy because when player 2 plays *L*, player 1's payoff is higher from *D* than from *U*, and when player 2 plays *R*, player 1's payoff is also higher from *D* than from *U*.

Both players have a dominant strategy in the above game. Person 2 will do better playing *R*, whichever strategy person 1 has chosen. When player 1 plays *U*, player 2 gets 5 from *L* and 10 from *R*. If player 1 were to play *D*, player 2 would get 0 from *L* and 1 from *R*. Therefore, *R* is a dominant strategy for player 2.

If each individual has a dominant strategy then we can say with confidence that the outcome that has each individual playing his or her dominant strategy is an equilibrium.

DEFINITION: Dominant strategy equilibrium

If each individual has a dominant strategy then there is a dominant strategy equilibrium, and it results when each person chooses his or her dominant strategy.

The game that we have been analyzing is an example of a *prisoner's dilemma* game, which demonstrates that the pursuit of self-interest does not always lead to an outcome that benefits individuals in the end. In the game of Table 1.8 when each individual is guided by self-interest, person 1 will play *D*, person 2 will play *R*, and each will get a payoff of 1. However, if each had chosen the alternative strategy, then the payoff for each would have been 5. The pursuit of self-interest is self-defeating in this game. (If person 1 thinks that person 2 has studied this game and will play *L*, doing her part so that they can each get 5, then person 1 has a strong incentive to play *D* because she gets 10 that way.) The prisoner's dilemma is more fully examined in Section 6. We remind you that dominant strategies do not usually exist. For example, the game resulting from repeated play of the prisoner's dilemma does not have dominant strategies.

5.2 Nash equilibrium

A dominant strategy equilibrium is a special case of a *Nash equilibrium*, in which each person's strategy is a *best response* to the strategies chosen by the

Table 1.9

	Rob	
	L	R
Pat		
U	12, 10	15, 5
D	10, 20	5, 25

other players. We say that S^* is player *A*'s best response to player *B*'s strategy *T* if there is no other strategy available to player *A* that gives her a higher payoff than S^* , given that the opponent has selected *T*. Consider the game described by Table 1.9: (*U*, *L*) is a Nash equilibrium because Pat's best response to *L* is *U*, and Rob's best response to *U* is *L*. It is the only Nash equilibrium, because if Pat were to play *D* then Rob would respond with *R*. But *D* is not Pat's best response to *R*.

The unique Nash equilibrium for this game is *not* a dominant strategy equilibrium. Although Pat has a dominant strategy (her best response to *L* is *U*, and her best response to *R* is also *U*), Rob does not. Rob's best response to *U* is *L*, but if Pat were to select *D* then Rob's best response would be *R*, not *L*.

In any two-person game, each player has a set of available strategies, and if player 1 chooses strategy S_1 and 2 chooses S_2 , we let $U_1(S_1, S_2)$ represent the resulting payoff to player 1, with $U_2(S_1, S_2)$ denoting player 2's payoff. We say that (S_1^*, S_2^*) is a Nash equilibrium if, given that player 2 plays S_2^* , there is no strategy available to player 1 that gives him a higher payoff than S_1^* and given

5. Equilibrium

that player 1 plays S_1^* there is no strategy available to player 2 that gives him a higher payoff than S_2^* .

DEFINITION: Nash equilibrium

 (S_1^*, S_1^*) is a Nash equilibrium if $U_1(S_1^*, S_2^*) \ge U_2(S_1, S_2^*)$ for every strategy S_1 available to person 1 and $U_2(S_1^*, S_2^*) \ge U_2(S_1^*, S_2)$ for every strategy S_2 available to person 2.

Because the payoff $U_1(S_1^*, S_2^*)$ to player 1 is the highest payoff available to him given that his opponent plays S_2^* we say that S_1^* is a *best response* to S_2^* . Note that a dominant strategy equilibrium is a special case of a Nash equilibrium: A dominant strategy is a best response to *anything* that the opponent might do.

In general, there are *n* players and each has a set of available strategies. We say that the strategy list $(S_1^*, S_2^*, S_3^*, \ldots, S_n^*)$ is a Nash equilibrium if for each player *i* the strategy S_i^* is a best response by *i* to the choice of S_j^* by each $j \neq i$. We may want an equilibrium to have additional properties but it should at least be self-enforcing in the sense that each person's strategy is a best response to what the others are doing.

5.3 The invisible hand

The prisoner's dilemma shows that without appropriate incentives the pursuit of self-interest can be self-defeating. Adam Smith identified a range of situations in

	Play L	ver 2 R
Player 1		
U	5, 5	7,2
D	2, 7	1,1

Table 1.10

which the pursuit of self-interest promotes the well-being of everyone, *without the need for regulation by any central authority*—except that there must be some agency to enforce the rules of the game. Consider the game represented by Table 1.10: Person 1 has to choose between two actions *U* and *D*, and person 2 has to choose between *L* and *R*. If person 2 plays *R* then person 1 does better playing *U* than playing *D*. But when person 1 plays *U*, person 2 does better switching to

L. And when person 2 plays *L*, the best response for person 1 is to play *U*. Then we have a Nash equilibrium with person 1 playing *U* and person 2 playing *L*. Although an individual would rather have 7 than the 5 that he or she gets at equilibrium, the temptation to get the big payoff doesn't ruin things. The incentives still take this tiny society to the (*U*, *L*) outcome and would still do so if we changed 7 to 700 in the description of the rules of the game: Person 1 has to play *U* to get 700, but person 2's best response to *U* is to play *L*. We don't get the bad outcome in this situation, even though each player is pursuing narrow self-interest, as in the prisoner's dilemma game of Section 5.1.

It's all a matter of incentives. If the appropriate incentives are in place, then the pursuit of individual self-interest leads to an outcome that benefits society as a whole—without the need for a government to guide the participants. We

Table 1.11

Country B	
Retaliate	Don't retaliate
120, 150	100, 125
95, 130	105, 95
	Ca Retaliate 120, 150 95, 130

say that we have a decentralized system—that is, the individuals are on their own to follow their self-interest. The games of this section and Section 5.1 are both decentralized; in one of them we get the bad outcome and in the other we do not.

5.4 The incentive to harbor terrorists

There is a spillover benefit from any effort to eliminate terrorism undertaken by an individual country. The elimination of any terrorist cell by any country reduces the threat to other countries. With many cases of spillover benefits, the

In the 1980s the United States became the main target of Middle East terrorist attacks, in part because of Israel's effective protection of El Al Airline flights and its efforts in other aspects of security (Hill, 1986). resulting game in which the players are involved is a prisoner's dilemma. But not in the case of countries seeking to protect themselves from terrorism. When country *X* puts more resources into shielding itself from terrorism, country *Y* becomes more vulnerable, to the extent that terrorists shift their activities away from *X* and toward *Y*. The cost of attack-

ing a country increases when that country increases its level of protection, and hence the probability of an attack on other countries increases.

Because of the shift in terrorist activity, the specific benefits to country *X* when it retaliates for acts of terrorism increases with the level of retaliation by country *Y*. Even if *X* would not have an incentive to retaliate if *Y* did not retaliate, if *Y* does retaliate then *X* is better off retaliating than being passive. Let's examine the resulting two-agent game of Table 1.11. The first number in a cell is country *A*'s payoff, and the second number is country *B*'s payoff. (When a country retaliates, it attacks the terorists, not the other country.)

Table 1.11 shows that it is to country *B*'s advantage to retaliate, whatever country *A* does. If *A* does *not* retaliate, then *B*'s payoff is 130 for retaliation but only 95 for passivity. However, if *A does* retaliate, then *B*'s payoff is 150 for retaliation and only 125 for passivity. We say that for *B*, retaliation is a *dominant strategy*. Therefore, we can be sure that country *B* will choose to retaliate. How will country *A* respond? Its payoff is 120 if it retaliates, and only 95 if it doesn't. Therefore *A* will retaliate. We have a unique equilibrium in which both countries retaliate. (Note that *A* would have an incentive not to retaliate if *B* did not retaliate. But, of course, we know that *B* will retaliate.)

Table 1.12 portrays the situation in which country *A* has a third strategy—to harbor terrorists within its borders, in hopes of winning the terrorists' favor, and

5. Equilibrium

Table 1.12

	Co Retaliate	ountry B Don't retaliate
Country A		
Retaliate	120, 150	100, 125
Don't retaliate	95, 130	105, 95
Harbor	140, 75	115, 80

in that way shield itself from attack. Often the host country obtains a promise from the terrorists that it will not be attacked. A country that would benefit considerably from a strategy of retaliation—because a terrorist attack could be particularly devastating—might be a country that would also benefit considerably

France, Italy, Greece, and Cyprus are among the many countries that have allowed foreign terrorists to establish a base within their own borders. Cuba has accepted a dozen United Nations' counterterrorist conventions, but it hosts a number of Latin America's most wanted terrorists, in addition to Basque terrorists and Irish Republican Army nationalists (*The Economist*, May 25, 2002, p. 30). from buying protection—that is, by harboring terrorists. Note that harboring the terrorist group is now a dominant strategy for *A*. On one hand, if country *B* retaliates then *A* gets a payoff of 140 from harboring, and that is higher than the payoff from either retaliating or being passive, given that *B* retaliates. On the other hand, if *B* does not retaliate then *A*'s payoff is 115 from harboring, but *A* gets only 100 from retaliating and 105 from passivity when *B* does not retaliate. Therefore, we can expect *A* to harbor the terrorists. *B*'s best response to that is to refrain

from retaliating because it gets a slightly higher payoff from doing so when *A* hosts the terrorists. Why does *B* retaliate in Table 1.11 but not in Table 1.12? Because part of the benefit to any country *X* from retaliating in the first scenario comes from offsetting the shift in terrorism activities against *Y* to *X* that results from *Y*'s retaliation. In the second scenario, country *A* does not retaliate because it pays *A* to harbor the terrorists.

The relative magnitude of the numbers that we have used is not the only plausible choice. In Table 1.11, we can think of *B* as a country like Israel that is plagued by local terrorists whose objective is to destroy *B*, whereas country *A* is victimized by foreign terrorists whose grievances are primarily against *B*. For that story we would probably want to increase the 75 in the bottom left-hand cell (excuse the pun) of Table 1.12 to 95, in which case *B* would be better off retaliating than being passive when *A* harbors the terrorists.

5.5 Dissolving a partnership

Two companies, located in different countries, embark on a joint project in a third country. If one of the parties wants to be released from its commitment at some stage, how should the breakup of the partners be adjudicated?

Before addressing this question, we look at the two-person division of a cake problem. (Example 4.3, with n = 2.) An allocation *x* assigns the fraction x_1 of

the cake to person 1 and x_2 to person 2. Allocation x is efficient if and only if $x_1 + x_2 = 1$. In particular, the allocation that assigns all of the cake to one person is efficient. (Either $x_1 = 1$ and $x_2 = 0$, or $x_1 = 0$ and $x_2 = 1$.) Such an outcome is far from fair, of course. Let's agree that $x_1 = \frac{1}{2} = x_2$ is the only *fair and efficient* allocation.

Suppose that our aim is to implement the fair and efficient allocation $x_1 =$ $\frac{1}{2} = x_2$ in a decentralized way so that the cake is distributed evenly as the result of selfish utility-maximizing behavior by each individual. We want to design a game for which the unique Nash equilibrium gives exactly half of the cake to each person. This cannot be done without specifying the rules of the game. These rules will detail the strategies available to each player and will also specify the allocation as a function of the pair of individual strategies. An appropriate mechanism is not hard to find. Let person 1 cut the cake into two pieces. With the sizes of the two pieces determined by individual 1, let person 2 choose one of the two pieces for his own consumption. Person 1 then consumes the remaining piece. The only equilibrium allocation generated by this game is the fair and efficient allocation $x_1 = \frac{1}{2} = x_2$ because person 1 knows that person 2 will choose the larger piece if 1 cuts the pieces unequally. Therefore, person 1 is sure to receive less than half the cake if she cuts the pieces unequally at stage 1. She can prevent this by cutting the cake precisely in half and this, therefore, is the strategy that ensures her the largest payoff. Consequently, person 2's choice becomes irrelevant.

This simple game has an important application in the business world. It often happens that two companies from different countries find themselves involved in a joint business venture in a third country. If at some point one of the parties is presented with a more profitable opportunity elsewhere and wants to abandon the project, there will be considerable uncertainty about the legal resolution that would be handed down by the courts. This prospect could inhibit the firms from undertaking the project in the first place, and thus some socially desirable investments may not be adopted.

Some joint ventures have been undertaken after the two participants agree to settle disputes over withdrawal by the following straightforward variant of the division of the cake mechanism: The partner that wishes to withdraw from the project names a price P at which he is willing to sell his share of the venture to the second partner. If that were all there were to it, the withdrawing partner would have a strong incentive to name an exorbitantly high price (and there would be a strong incentive to withdraw) just as person 1 would have a strong incentive to cut a cake unequally if he were the one to decide who gets the larger piece. But there is a second stage to the game: The second partner now chooses whether to buy the other out at the named price P or to sell out to the partner that set the price P. This forces the withdrawing partner to set a price equal to one half of the present value of the project. Proof: Suppose that the present value of the project is V and the contracting parties have equal ownership shares. If the project is completed then each gets a payoff of $\frac{1}{2}V$. If partner 1 (the company wishing to withdraw) names a price $P > \frac{1}{2}V$ then partner 2 is better off selling to 1 at that price than insisting on completion. If

5. Equilibrium

$$(4, 1) (2, 8) (16, 4) (8, 32) (64, 16) (32, 128) (256, 64)$$

Figure 1.3

partner 1 sets *P* below $\frac{1}{2}V$ then partner 2 will want to buy out partner 1 at that price and complete the project at her own expense for a net gain of $V - P > \frac{1}{2}V$. In either case, by choosing a value of *P* different from $\frac{1}{2}V$, partner 1 will wind up with less that $\frac{1}{2}V$ and he can always ensure a payoff of exactly $\frac{1}{2}V$ by setting $P = \frac{1}{2}V$.

Because the only equilibrium solution has the withdrawing partner setting $P = \frac{1}{2}V$, why don't they simply agree in advance that $\frac{1}{2}V$ will be the price at which a partner can be bought out should he or she decide to withdraw before the project is completed? Because there will be disagreement about V. The remaining partner will claim that the project has little likelihood of generating substantial profit and will offer to buy out the other at a very low price, claiming that she is offering $\frac{1}{2}V$ but that V is very small. The partner selling his share in the enterprise will have a strong incentive to claim that V is very large, whatever he really believes, and hence $\frac{1}{2}V$ is large. Suppose, however, that partner 1 can name a price P and then partner 2 has the right to buy out the other at P or to sell her share to company 1 for P. Then partner 1 could lose heavily by naming a price that was much greater than $\frac{1}{2}V_1$, where V_1 is partner 1's estimate of the present value of the project. If partner 2's estimate of the present value were no higher than partner 1's, then partner 2 would opt to sell to partner 1 at any price P greater than $\frac{1}{2}V_1$ and the net value to partner 1 of the project would then be $V_1 - P$, which is less than $\frac{1}{2}V_1$, the payoff that partner 1 could get just by setting $P = \frac{1}{2}V_1$.

5.6 The centipede game

In spite of the plausibility of Nash equilibrium, there are games that have a single Nash equilibrium that is not a reasonable forecast of the game's outcome. One of the niftiest examples is the so-called centipede game characterized by the game tree in Figure 1.3. The two players are Samantha and Tyler. As time passes we move from left to right along Figure 1.3.

The players take turns moving, and when it's a player's turn to move he or she has to choose between grabbing the money (g) and passing (p). If he or she passes then the total amount of money available doubles. When one of the players grabs then the game is over and Samantha's payoff is the first number in parentheses and Tyler's is the second number. If each player passes at each turn then the game ends at the extreme right of the diagram with Samantha receiving 38



\$256 and Tyler receiving \$64. (Note that the only efficient outcomes are this one and the second-last outcome at which Samantha receives \$32 and Tyler receives \$128.)

What makes analyzing the game tricky is that, although the total payoff doubles every time a player passes, the amount that he or she will receive if the other person responds by choosing *g* is cut in half. Suppose that each player passes every time it is his or her turn to move. Then Tyler will receive \$64. But he can get twice as much money by grabbing on his last move instead of passing. Passing at the last stage is not a best response by Tyler to Samantha's strategy of passing at every opportunity. Therefore, the outcome that results when each player passes at each opportunity is not self-enforcing and hence not part of a Nash equilibrium.

Both players can predict that the game will not end with a player passing. Suppose both players anticipate that the game will end after move t with Tyler grabbing at that stage. Then Samantha will not let the game survive to that stage because she can get twice as much money by grabbing on the previous move, instead of passing and letting the game continue. Similarly, if Samantha were expected to end the game by grabbing at stage t > 1, Tyler's best response would be to grab at the previous stage because he would double his payoff by doing so. Therefore, the only self-enforcing outcome has Samantha and \$1 for Tyler. This is obviously far from efficient. (The game is called the centipede game because the associated diagram looks like a centipede. Moreover, one could extend the game by adding 94 more moves, with the pot continuing to double each time. The starting point would remain the only equilibrium, but it offers minuscule payoffs relative to those available later on.)

Our intuition tells us that the two players would not end up at the Nash equilibrium. In fact, McKelvey and Palfrey (1992) conducted experiments and found that the players typically finish somewhere near the middle of the centipede, not at either extreme of grabbing at the first opportunity or passing until the last move or two. Therefore, Nash equilibrium is an inappropriate solution concept in this case. Why? To identify the difficulty, we will truncate the game so that each player potentially has only two moves, as illustrated in Figure 1.4. We refer to the later move as the player's second move.

5. Equilibrium

Here is a difficulty: We have implicitly assumed that both players are "rational." Rationality means that agents care only about the effect of an outcome on their own welfare, and they always act to enhance their welfare in any situation where that has an unambiguous meaning. We have also assumed that each player believes that the other is rational. Here is the argument that establishes that the unique Nash equilibrium has Samantha grabbing on the first move and receiving \$4, with \$1 going to Tyler: If Tyler is rational and he has the opportunity to make the last move—his second move—he will grab rather than pass because he gets \$32 by grabbing and only \$16 by passing. Nothing remarkable about the background assumptions so far.

Now, suppose that Samantha is rational *and* that Samantha knows that Tyler is rational. Then Samantha will anticipate that Tyler will grab if he has a second move. This means that Samantha deduces that she will get \$8 if Tyler is given an opportunity to make a second move. Therefore, if Samantha has the chance to make a second move, she knows that she is really choosing between \$8—if she passes—and \$16—if she grabs. She is rational, so she will grab if she has a second move. Now, suppose that Tyler is rational, Tyler knows that Samantha is rational, *and* Tyler knows that Samantha knows that Tyler is rational. Then Tyler can anticipate that Samantha will grab if Samantha has an opportunity for a second move. Then Tyler will wind up with \$4 if Samantha has a second move. Therefore, on Tyler's first move—if he has one—he can obtain \$8 by grabbing or \$4 by passing. He is rational, so he will grab on the first move if Samantha hasn't grabbed first. And so on. The conclusion that Samantha will grab at the first opportunity is based on the following suppositions:

- 1. Samantha and Tyler are each rational.
- 2. Samantha knows that Tyler is rational.
- 3. Tyler knows that Samantha knows that Tyler is rational.
- 4. Samantha knows that Tyler knows that Samantha knows that Tyler is rational.

Statement 1 implies that Tyler will grab if he is given a second move. Statements 1 and 2 imply that Samantha will grab if she is given a second move. Statements 1–3 imply that Tyler will grab on his first move if Samantha passes on her first move. Statements 1–4 imply that Samantha will grab on the first move. Therefore, assumptions 1–4 collectively imply that the unique Nash equilibrium has each person grabbing whenever he or she is given an opportunity to move. But these assumptions are extremely unstable. If Samantha actually passes on the first move then Tyler knows that one of the four statements is false—perhaps Samantha is not rational, or perhaps she is unsure that Tyler knows that she knows that Tyler is rational—and the logical chain directing Tyler to grab at the first opportunity is broken. *Anything* can happen now.

The longer the game, the larger is the spread between the payoff a player gets by grabbing early and the payoff that awaits both players if the game ends much later. Moreover, the longer the game, the longer is the chain "I know that he knows that I know that he knows ..." that is required to support the backward induction derivation that the game will end on the first move. For long games

of this nature—or short ones, for that matter—we don't have a good model for predicting behavior, but at least we can see why results in which the game ends after seven or eight rounds of passing are not inconsistent with our basic rational choice model. It's just that the results are inconsistent with the implicit assumption about what individuals know about what others know.

5.7 Subgame-perfect Nash equilibrium

The centipede game of the previous section has a single Nash equilibrium, but we don't have much confidence that it would emerge as the outcome when the game is actually played, and that is confirmed by experiments. Now we examine a game with *two* Nash equilibria, one of which is not a sensible forecast of the game's outcome. In this case the equilibrium is implausible because it is based on a threat that is not credible. Simply put, a *subgame-perfect equilibrium* is a Nash equilibrium that is not sustained by a threat that is not believable. Before we can present a formal definition we need to prepare the ground.

A *strategy* is much more comprehensive than an *action*. "Steal second base now" is a simple instruction by a coach in a baseball game, and the attempted theft is the action. But a strategy specifies an act as a function of every act made by every participant up to the present stage of the game. "Attempt a theft of second base if we haven't reached the fifth inning, or if it is late in the game and we are behind by two or more runs, provided that the batter has fewer than two strikes and the probability of a pitch-out is less than 0.25" is a strategy. We could specify a single strategy for the manager of a baseball game for the entire game. It would specify a decision for every situation that could arise, as well as the decisions made at the beginning of the game before the opponent has taken any action.

Consider a deterministic two-person game—that is, a game between two individuals that is not affected by any random variables. Let S₁ and S₂ denote, respectively, the strategies chosen by players 1 and 2. Then the pair (S_1, S_2) uniquely determines the outcome of the game. If we display the payoffs awarded to player 1 as function of the strategies chosen by players 1 and 2, and similarly for player 2, we have what is called the normal form representation of the game. The normal form payoffs are simply expressed as functions $U_1(S_1, S_2)$ and $U_2(S_1, S_2)$ of the chosen strategies. Recall that a Nash equilibrium is a pair of strategies (S_1, S_2) such that $U_1(S_1, S_2) \ge U_1(T_1, S_2)$ for every strategy T_1 available to person 1, and $U_2(S_1, S_2) \ge U_2(S_1, T_2)$ for every strategy T_2 available to person 2. It is helpful to think of the respective strategies S_1 and S_2 as chosen simultaneously and submitted to a referee who then computes the outcome and assigns payoffs according to the rules of the game. In a nondeterministic game there are points in the game at which a strategy calls for an act to be selected randomly by means of a given probability distribution over a given set of acts. Uncertainty may even be imposed on the players-the arrival of rain during a baseball game, for example. The outcome will be random, but the payoff to an individual associated with any configuration of strategies-one for each player-can still be expressed as a single number by using the probabilities as weights on the different payoffs that could arise. (See Section 6.1 of Chapter 2.)

5. Equilibrium

An *extensive form* representation of the game has much more structure than the normal form. The extensive form provides information about the sequences of moves—whose turn it is to move at each stage and what choices that person has. A strategy for an individual prescribes an action for that person for every situation that could arise.

DEFINITION: Individual strategy

Player **B**

 $R \rightarrow D$

1,5

0,2

 $R \rightarrow U$

1,5

3, 3

At any point in the game at which the player is allowed to move, the strategy specifies an action for that player for each potential history of the game to that point—and a single action for the game's opening move.

Example 5.1: Extensive form two-person game

The game is represented as Figure 1.5. At the first stage player *A* has a choice of moving left or right. If *A* moves left the game is over, and *A*'s payoff is 1 and *B*'s payoff is 5. If player *A* moves right at the first stage then player *B* has the next move and can go up or down. If *B* chooses up then each gets a payoff of 3, but if *B* moves down then *A*'s payoff is 0 and *B*'s payoff is 2. Consider the normal form representation of the same game displayed as Table 1.13. $R \rightarrow U$ represents the strategy "*B* moves Up if *A* has opened by moving Right," and $R \rightarrow D$ represents the strategy "*B* moves Down if *A* opened by moving Right." There are two Nash equilibria here: (Right, $R \rightarrow U$) and (Left, $R \rightarrow D$). Confirm that Right is a best response by *A* to $R \rightarrow U$, and that $R \rightarrow D$ are by *B* to Right. Note also that Left is a best response by *A* to $R \rightarrow D$, and that $R \rightarrow D$ is a best response by *B* to Left.

The equilibrium (Left, $R \rightarrow D$) of Example 5.1 is not a plausible one. It depends on B's threat to move Down if *A* moves Right. In plain words, *A*

announces her intention to move Left, whatever *B* proposes to do should *B* get a chance to move, and *B* announces her intention to move Down if *A* moves Right. If *A* believes that *B* is really committed to Down if *B* gets a chance to move, then Left is the rational choice for *A*: Left gives *A* a payoff of 1, but *A* gets 0 if he moves Right and *B carries out her threat* to move Down. However, *B*'s threat is not credible. If *B does* get a chance to move it will come after *A*'s move and

thus it can have no impact on *A*'s choice. Therefore, the payoff-maximizing move for *B* is Up, yielding a payoff of 3 instead of 2. A Nash equilibrium that does not depend on an incredible threat is termed a *subgame-perfect Nash equilibrium*. *Subgame* refers to the game that would be defined if we were to begin play at

Table 1.13

Player A

Left

Right





some advanced stage of the original game. The players are assumed to play best responses in the subgame. For Figure 1.5, moving Up is *the* best response for *B* at the second stage, so a threat to move Down is not credible. The only subgame-perfect Nash equilibrium is (Right, $R \rightarrow U$).

DEFINITION: Subgame

A subgame of an extensive form game is a game obtained by separating the tree at one node and retaining only that node and all parts of the tree that can be reached from that node by going forward and not backward (in time). Any node may serve as the origin of a subgame, *provided* that the person who moves at that stage knows the entire history of the game up to that point.

The prisoner's dilemma (Table 1.8 in Section 5.1) can be represented in extensive form: Player 1 moves first and chooses between *U* and *D*. At the second node, player 2 chooses between *L* and *R*. However, at this point player 2 will not know what choice player 1 made at the first node. Therefore, player 2 will not know the prior history of the game when it is his turn to move. This game has no subgames (except for the entire game itself).

The game of Figure 1.5 has five subgames, including the original game itself: There are three trivial subgames corresponding to the three terminal nodes with respective payoff vectors (1, 5), (3, 3), (0, 2). The trivial subgames do not allow anyone to move, of course. There is only one proper and nontrivial subgame, obtained by eliminating the branches Left and Right.

If the original game includes moves in which the player taking action is not perfectly certain of what has gone before, then a subgame must have an additional property: at the node *N* where the separation identifying the subgame occurs, any act *A* by the mover *M* (the player who moves at *N*) must be included in the subgame if there is *some* prior history of the game that would make *A* available to *M* if *A* is not ruled out by the information available to *M* at *N*. A

5. Equilibrium

subgame-perfect equilibrium is one that remains an equilibrium for all subgames—with the equilibrium strategies amputated to fit the subgame.

DEFINITION: Subgame-perfect Nash equilibrium

A Nash equilibrium β is subgame perfect if the strategies specified by β constitute a Nash equilibrium in every subgame.

For the game of Figure 1.5, if we begin at the point where *B* moves we have a subgame in which *B* chooses between Up and Down. Clearly, Up is the only Nash equilibrium in this one-player game. Therefore, the equilibrium (Left, $R \rightarrow D$) of the original game is not subgame perfect.

For *finite* games we locate subgame-perfect equilibria by backward induction: Begin with the proper subgames that are closest to a terminal node. In Figure 1.5, that would be the subgame beginning with *B*'s move. Replace those subgames with their Nash equilibrium payoffs. For the subgame of Figure 1.5 that begins with *B*'s move, player *B* has a simple choice between Up, with a payoff of 3 to herself, and Down, which gives her a payoff of 2. She would choose Up,



resulting in the payoff vector (3, 3). We replace the entire subgame with (3, 3), as illustrated in Figure 1.6. We continue by induction. Having reduced the size of the game by successively abbreviating it by replacing subgames with their Nash equilibrium payoffs, we have a new game. We then identify the proper subgames that are closest to a terminal node of this new game. Then we replace those subgames with their Nash equilibrium payoffs. At some stage we will have reduced the game to one with a single move, as in

Figure 1.6. The Nash equilibrium of that game gives us the subgame-perfect equilibrium of the original game. The unique Nash equilibrium of Figure 1.6 has *A* choosing Right, leading to a payoff of 3 for *A*. (His payoff would only be 1 if he chose Left.) Therefore, the unique subgame-perfect equilibrium for the game of Figure 1.5 is (Right, $R \rightarrow U$).

Sources

The term *Nash equilibrium* honors the mathematician John Nash, who is the subject of the book *A Beautiful Mind* by Sylvia Nasar (1998). Lee (1988) is the basis for Section 5.4 on terrorism. The centipede game was invented by Robert Rosenthal (1981).

Links

Myerson (1999) is an excellent study of the history of Nash equilibrium in economic analysis. Baumol (2002) discusses the contribution of Adam Smith more deeply than our static version of Section 5.3. There are other situations, in addition to the centipede game, in which Nash equilibrium does not appear to offer a

good forecast of the outcome that would result when intelligent, self-motivated people interact. See Goeree and Holt (2001) for ten important cases. Cramton, Gibbons, and Klemperer (1987) is a very advanced treatment of the problem of dissolving a partnership. Subgame-perfect equilibrium is discussed at length in Binmore (1992), Gibbons (1992), Kreps (1990), and Osborne (2004). Frank (2004) offers an attractive suggestion for enriching the standard economic model in a way that is consistent with observed play of the centipede game.

Problem set

1. The utility functions of our two individuals are

 $U_1 = 100(e_1 + e_2) - 150e_1$ and $U_2 = 100(e_1 + e_2) - 150e_2$

where e_1 is the effort contributed by individual 1 and e_2 is the effort contributed by individual 2. Each individual *i* can set e_i equal to *any number* between zero and one inclusive.

- **A.** Given individual 2's choice of e_2 , whatever that might be, what is the best response of person 1? Justify your answer.
- **B.** What is the Nash equilibrium for this game?
- **2.** This time there are *n* individuals, and e_i is the effort contributed by individual *i* whose utility function is

$$U_i = \alpha(e_1 + e_2 + \cdots + e_{n-1} + e_n) - \beta e_i$$

Individual i can set e_i equal to any number between zero and one inclusive.

- **A.** Show that $e_i = 0$ is a dominant strategy for individual *i* if and only if $\alpha < \beta$.
- **B.** For what range of values of α and β will we have $U_i(1, 1, ..., 1, 1) > U_i(0, 0, ..., 0, 0)$? Justify your answer.
- **C.** If $n = 10 = \beta$, for what range of values of α is this a prisoner's dilemma game?
- **3.** *X* and *Y* are on the only candidates on the ballot in an election. Every voter prefers *X* to *Y*. Explain why we have a Nash equilibrium if everyone votes for *Y* and there are at least three voters. Is this a plausible forecast of the outcome? Are there any other Nash equilibria? (Note that each voter has a dominant strategy.)
- 4. The airline has lost the luggage of two travelers. Their luggage was identical. Each is invited to submit a claim for any *integer* amount between \$10 and \$100 inclusive. If the claims are identical then each receives a payment equal to the common claim. If the claims are not the same then the traveler submitting the smaller claim gets that amount plus \$5, and the traveler submitting the larger claim receives a payment equal to the smaller claim minus \$5. Prove that the unique Nash equilibrium has each person submitting a claim for \$10. This game was devised by Basu (1994). (It is noteworthy that experiments reveal a high concentration of claims around \$99.)

6. The Prisoner's Dilemma Game

Table	1.14
-------	------

	Player B	
	Don't confess	Confess
<i>Player A</i> Don't confess Confess	1, 1 0, 10	10, 0 5, 5

6 THE PRISONER'S DILEMMA GAME

This section discusses a simple situation in which the interplay of incentives leads to an outcome that the participants deeply regret, even though the outcome is the consequence of the pursuit of self-interest: Self-interest drives individual behavior but self-interest is self-defeating in this setting. The phenomenon under discussion, the *prisoner's dilemma* paradox, has a wide range of applications; it explains many organizational failures. It is a model of a situation in which individual incentives are *not* well aligned.

If fire is detected in a crowded building almost no one will escape alive if there is panic, and all attempt to get through the exit door at once. But if the crowd

Ninety-seven people died in a Rhode Island nightclub after a fire broke out. "There was nowhere to move" (*Boston Herald*, February 26, 2003). Twenty-one people died in a Chicago nightclub after a fight provoked a panic that resulted in the exits being so completely jammed that the people stuck there couldn't move forward or backward (*Chicago Sun-Times*, February 18, 2003). is orderly it is advantageous to any one individual to run past everyone else to get to the exit first. Panic will prevail if everyone comes to the same conclusion. However, if everyone runs to the exit then no one can gain anything by walking slowly.

The prisoner's dilemma refers to a simple game involving two players, each of whom must choose one of two options independently of the other. The game can be described abstractly but we will introduce it in its original guise: Two individuals *A* and *B* have been

arrested and charged with bank robbery. The police are convinced of their guilt but there is no admissible evidence on which they can be convicted of robbery, although they were carrying guns when caught and for this each can be sentenced to one year in jail. To obtain confessions to the crime of robbery the authorities interrogate them separately and offer each his complete freedom if he confesses to the robbery and his partner does not. The partner who does not confess will receive ten years in jail, but if *both* confess then each will receive a five-year sentence. The situation confronting each prisoner is summarized by Table 1.14.

The first number in a cell is *A*'s sentence, and the second number is *B*'s sentence. *A* and *B* cannot communicate with each other—or if they can communicate they can't make binding agreements. Suppose that *A* believes that *B* will not confess. Then *A* will receive a sentence of one year if he does not confess,

Table 1.15

	Player	B
	Cooperate	Defect
<i>Player A</i> Cooperate Defect	20, 20 30, 1	1, 30 5, 5

but he will not have to serve any time if he confesses. *A* receives a lighter sentence by confessing. However, suppose that *A* believes that *B* will confess. Then *A* will receive five years if he confesses but ten years if he doesn't. Again, *A*'s self interest is served by confessing. Whichever decision the partner in crime is expected to make, an individual does better by confessing than not confessing. They both confess and each receives a sentence of five years. If neither had confessed then each would have been free after only one year.

In this situation self-interest drives each person to take a course of action that leaves each worse off than if they had coordinated their strategies. (But notice

U.S. law allows a firm involved in a corporate conspiracy to escape punishment if it is the first to confess (*The Economist*, October 21, 2000, p. 67).

how strong the incentive is to get one's partner to agree not to confess and then, having also solemnly sworn not to confess, to confess and go free.)

Consider the general formulation of this game, with the outcomes translated to money

(or similar) payoffs, which an individual wants to *maximize*. Each person must decide whether to cooperate or to defect without knowing what choice the other will make. The payoff for each of the four possible combinations of strategies is given in Table 1.15. The first number is player *A*'s payoff, and the second number in a cell is player *B*'s payoff.

If *B* is expected to cooperate then *A* can get 20 by cooperating but 30 by defecting. If *B* is expected to defect then *A* can get 1 by cooperating and 5 by defecting. In either case the higher payoff for *A* is obtained by defecting. Defecting is a *dominant strategy*. A dominant strategy is one that is the best course of action for a decision maker regardless of the actions that others are expected to take. *B* is in exactly the same position; both will defect and each receives a payoff of 5. If each had chosen to cooperate then each would have received a payoff of 20. The equilibrium outcome is not efficient. The equilibrium, which is the outcome when both play their dominant strategies, gives each a lower payoff than when both cooperate. The incentive to defect is irresistible, however, assuming that the game is played under two conditions. First, the two players cannot undertake a binding commitment to cooperate. Second, the game is played only once.

If the players can make commitments then the incentives could be quite different. Suppose, for example, that before playing the game the two players anticipated that each would succumb to the temptation to defect and each

6. The Prisoner's Dilemma Game

signed a document that required one person to pay the other a thousand dollars if he defects. This contract could also state that it was binding on a signatory only if the other person signed. This results in a new game in which both can be expected to sign the document and cooperate. (Assume that the payoffs in Table 1.15 are dollar amounts.) Now, suppose that binding agreements are not possible but the same game is repeated a number of times by the same two players. Then we have a different game with different incentives, although the tension that we have uncovered in the "one-shot" game still plays a role in the repeated game. We can still have an equilibrium in which each person defects at each stage, but the justification of this as an equilibrium is feeble compared to the story for the one-shot game. For one thing, there are no dominant strategies in the repeated version of the prisoner's dilemma game which is discussed in more detail in Sections 7.1 and 7.2.

We now turn to a consideration of seven situations for which the prisoner's dilemma game is applicable. The first five illustrate how the prisoner's dilemma incentive structure can work to the disadvantage of society. But it can also work to society's benefit, as in the case of Sections 6.6 and 6.7. In every case there are more than two people, but the extension to the several-person case is straightforward.

DEFINITION: n-person prisoner's dilemma

A game with two or more players is a prisoner's dilemma if each has a unique dominant strategy and an inefficient outcome results when each plays his or her dominant strategy.

6.1 Economic sanctions

Shortly after Iraq invaded Kuwait in August 1990 the United Nations Security Council imposed sanctions against Iraq. Most countries endorsed the sanctions and publicly stated a commitment not to allow imports from Iraq or to permit exports to Iraq. By December, observers in the Middle East were reporting serious leakages in the blockade. Let's look at sanctions from the standpoint of the incentives facing a typical country. Oil is the chief export of Iraq. A ban on the purchase of goods from Iraq is costly to an importing country because it reduces its options for acquiring energy. The restriction on exports is costly because trade is mutually advantageous, and to the extent that a country restricts trade it obviously limits the benefits that it receives from trade. The ban on exports would be seen in the legislature and in the press of the banning country as a threat to income and employment in that country. In addition, compliance with the sanctions would have to be monitored by the central government and that involves direct costs.

On one hand, if a large number of countries joined in the imposition of sanctions then country *A* would be tempted to relax its grip to recapture some of the benefits of trade, hoping that others would maintain the sanctions with sufficient determination to allow *A* to reap the benefit of sanctions without having to pay

the cost. On the other hand, if quite a few countries allow trade to continue then country *A* will benefit little from any embargo that it imposes, because sanctions have little effect if they are not widely enforced. In short, the dominant strategy for each country is to allow its firms to disregard the sanctions. This is not an argument against multilateral sanctions. However, the prisoner's dilemma problem teaches that sanctions must be implemented with a clear understanding of the incentives facing individual countries and with the determination to use diplomacy and ongoing consultation to maintain compliance.

Although there was less than total compliance with the economic sanctions against Iraq, there was enough of an effect to cause serious hardship among the Iraqi poor. In 1995 the United Nations instituted an oil-for-food program to relieve the suffering. Iraq was allowed to export a limited amount of oil at a limited price. The revenue was paid into a United Nations escrow account, to be used only for essentials—food and medicine, in particular. The program apparently led to a wide range of abuses including smuggling, illegal commissions, bribes, and kickbacks. At least \$2 billion wound up in Saddam Hussein's pocket. More surprising are the charges that up to \$10 billion found its way into the bank accounts of officials outside of Iraq (*The Economist*, May 1, 2004, pp. 46–7).

6.2 Public opinion

It is quite costly for an individual to stay well informed on most issues that are before national legislatures. On the one hand, the cost of investing the time required to develop an intelligent opinion on each critical public event is considerable, and on the other hand, the personal benefit from the resulting improvement in the quality of public opinion is negligible because a single individual's viewpoint, whether sound or silly, has a negligible effect. Whether others are well informed or not, an individual's own utility is maximized by investing in knowledge up to the point where the benefit *to him or her* from any additional investment would be more than offset by the cost. This results in citizens generally not being well enough informed from the standpoint of their own welfare. If everyone were to invest additional time in studying current events then public opinion would induce better public decisions and that would benefit everyone.

6.3 Pollution

Suppose that consumers have a free choice between automobiles produced without emission control devices and automobiles that have equipment that eliminates most harmful exhaust but costing \$3000 more than those without. (The emission control equipment costs \$3000 per car to manufacture and install.) Consider the typical consumer's decision problem. Given the choices made by others, whatever they are, the purchase of a pollution-free car would cost the individual \$3000 extra but would not appreciably improve the quality of the air. Clearly, purchasing the cheaper, polluting automobile is a dominant strategy. Everyone makes this choice and thus automobile traffic generates substantial pollution. One could specify the payoffs so that everyone would be better off if each paid a \$3000 charge to eliminate pollution caused by automobile exhaust, but the individual incentives push the society away from this outcome.

6. The Prisoner's Dilemma Game

The Environmental Protection Agency was formed in the United States in 1970. Before that time pollution was regulated in part by private lawsuits. The prisoner's dilemma phenomenon was involved here as well. Individuals get the benefit of any pollution-reduction strategy financed by their neighbors, whether or not they themselves make a contribution. Declining to help pay the legal costs of a lawsuit is a dominant strategy for each individual. Hence, there is less than the socially optimal amount of pollution abatement when we rely exclusively on private lawsuits to regulate behavior.

6.4 Beggar-thy-neighbor policies

The great depression of the 1930s had most industrial countries in its grip, and individual nations were unable to resist the temptation to devalue their currencies. Given the exchange rates of other countries, if one country devalued its currency then its goods would be cheaper to the rest of the world and its own citizens would import less as other countries' goods rose in price in terms of the domestic currency. The result is a stimulus to the devaluing country's industries at the expense of other countries. (It was thus called a beggar-thy-neighbor policy.) But the same temptation confronts each nation. Devaluation is a dominant strategy. Each country attempts to lower the price of its currency relative to others and adopts additional measures to restrict imports. As all countries restrict imports all countries' exports dwindle and the worldwide depression deepens.

6.5 Disarmament

In this example the players are countries. Defecting in this case is a decision to arm heavily. Cooperation is the decision to maintain only a defensive posture. If a country expects others to cooperate there is a strong incentive to obtain an extra measure of security by arming heavily. If the same country expects others to arm heavily then national security demands that the country arm heavily. Arming heavily is a dominant strategy for each country. Alternatively, imagine that war has broken out between *A* and *B* and that defecting corresponds to the use of chemical weapons in combat. Without introducing any other considerations, our analysis predicts that the use of chemical weapons would be commonplace. But that's not what we observe. *Because* defecting is a dominant strategy in these situations, nations generally have a responsibility to convince belligerents that the employment of particularly heinous methods of warfare (or violations of the Geneva conventions on the treatment of prisoners of war, etc.) will be counterproductive. In other words, it has to be brought to bear on *A* and *B* that they are playing a larger game than the immediate one-shot prisoner's dilemma game.

6.6 Cartels

Producer cartels form to keep industry supply low and price high. This provides each member of the cartel with more profit than when they compete vigorously against each other. When the firms actively compete then the industry output will be high, and price low, because each firm's output will be relatively high. Cooperation in the cartel context requires a firm to stick to the cartel agreement by restricting its own supply. But if each firm does this the market price will Table 1.16

	Hold	Rosie Sell 2 shares	
Soren Hold Sell 2 shares	36, 36 40, 20	20, 40 30, 30	

be high, and if the market price is high then an individual firm has a strong incentive to increase its profit by producing more output. If every firm does this the market output will be high and the price low. This is a case where the incentive structure, which leads to an inefficient outcome from the standpoint of the *group of producers*, works to the benefit of society as a whole: Individual incentives promote competition in spite of the substantial profits awaiting the shareholders in firms that can get their competitors to agree to cooperate in restricting output. The original prisoners' dilemma, in which suspects are interrogated, is another instance in which society is served although the individuals playing the game deeply regret the outcome.

6.7 Hostile takeovers

DianeCam corporation has two owners, Soren and Rosie, each of whom owns two shares in the firm. The current market value of a share is \$10. Edie wants to acquire all four shares in the firm and then replace the current manager with a more efficient one. This will raise DianeCam's profit and hence the market value of a share from \$10 per share to \$18. Therefore, Edie could offer to buy the outstanding shares at \$15 each. This would give Soren and Rosie a nice profit. But, why would they sell if the shares will be worth even more after the takeover?

Edie can get around this difficulty by means of a *two-tier offer*: She offers to pay \$20 per share for the first two shares tendered and buy the next two at \$10 each, but if Soren and Rosie simultaneously tender two shares each then Edie will pay each owner \$20 + \$10 for two shares. Soren and Rosie now face a prisoner's dilemma problem represented by Table 1.16.

On one hand, if Soren holds onto his shares, waiting for the takeover to drive their value up to \$18, then Rosie will get \$18 per share if she also holds out, but \$20

From an economy-wide perspective, takeovers may improve the performance of managers, who risk being dismissed by a new owner if the firm has been relatively unprofitable. (See the discussion in Section 4.1 of Chapter 4.)

for each of her two shares if she tenders them to Edie. On the other hand, if Soren tenders his two shares immediately then Rosie gets \$10 for each of her two shares if she holds out, but a total of 20 + 10 if she also sells right away. (A condition of sale at the \$20 price is that DianeCam will be merged with a company Edie already owns, once Edie has 50% of the shares, and the

outstanding DianeCam shares will be purchased for \$10 each.) Whatever Soren elects to do, Rosie does better by selling her shares immediately. Similarly with

6. The Prisoner's Dilemma Game

Strate	gy pair	Gan	ne 1	Ga	me 2	Gai	me 3
$\overline{S_1}$	S_2	$\overline{U_1}$	U_2	$\overline{U_1}$	U_2	$\overline{U_1}$	U_2
U	L	-5	10	5	5	10	12
U	R	10	5	7	1	5	5
D	L	-10	5	1	7	5	40
D	R	5	-5	8	2	20	24

Table 1.17

Soren. Because selling is a dominant strategy, the takeover will be consummated. (Even if there is no takeover when they both hold out, their shares will be worth only \$10 each, so selling is still a dominant strategy.)

Sources

The prisoner's dilemma game was invented by Dresher and Flood at Rand in the 1950s. Professor Albert Tucker of Princeton University immediately recognized its great significance for social studies. The example of Section 6.7 is based on Ryngaert (1988).

Links

Poundstone (1992) is an informative book on the history of the prisoner's dilemma game. Both Osborne (2004) and Binmore (1992) provide a thorough analysis of the game. Page 28 of Osborne (2004) gives an excellent account of experiments involving the prisoner's dilemma. See Downs (1957), pages 207–219, for a thorough discussion of the public opinion "game" of Section 6.2. Frank (2004) offers an attractive suggestion for enriching the standard economic model in a way that is consistent with observed play of the prisoner's dilemma game.

Problem set

- 1. Table 1.17 gives you enough information to set up three different games. In each case player 1 has two available strategies, *U* and *D*, and player 2 also has two available strategies, called *L* and *R* in her case. The table gives you each player's payoff (or utility) for each of the four possible pairs of strategies. For each game, determine if it is a prisoner's dilemma game, and defend your answer.
- **2.** Consider a market served by two firms with identical cost functions $C_i = Q_i$, where Q_i is firm *i*'s output and C_i is the firm's total cost. The market demand curve is Q = 82 2P.
 - **A.** Determine the market output and price when the two firms form a cartel that restricts output to maximize industry profit.

Table 1.18

Game	Left	Right
Game 1 Up	15.15	2.8
Down	8, 2	10, 10
Game 2		
Up	12, 12	20, 2
Down	2, 20	5, 5
Game 3		
Up	5, 50	50, 0
Down	0, 500	10, 100
Game 4		
Up	7,7	4,10
Down	4, 10	5, 5
Game 5		
Up	100, 100	4, 102
Down	102, 4	5, 5

- **B.** Assuming that the cartel imposes a quota on each firm equal to half the industry profit-maximizing level of output, what is the firm's profit under the cartel arrangement?
- **C.** Now assume that consumers will buy only from firm *i* if it breaks the cartel agreement and charges a price of \$15 when the other firm continues to charge the cartel price. What profit will each firm receive if firm *i* maximizes profit given a price of \$15 and given the market demand curve?
- **D.** If a firm has a choice of only two strategies—charge \$15 or charge the cartel price—show that they are playing a prisoner's dilemma game.
- **3.** Determine which of the five two-person games defined by Table 1.18 are examples of the prisoner's dilemma game. In each case each individual must choose one of two strategies: *A* controls the rows and *B* controls the columns. In other words, *A* must choose between Up or Down and *B* must choose between Left or Right. Each combination of strategies determines a payoff to each person as indicated in the table: the first number in a cell is *A*'s payoff and the second number is *B*'s. Each player wants to maximize his or her payoff and the players cannot make binding contracts. (All of which says that we have the standard setting.)

Table 1.19	9
------------	---

	Ryan Cooperate	n Defect
<i>Jodi</i> Cooperate Defect	10, 10 15, 2	2, 15 5, 5

4. Consider the prisoner's dilemma of Table 1.19. Suppose that Ryan and Jodi play the game three times in succession and each knows that the game will end after three periods. Show that defecting every period is *not* a dominant strategy even though the unique (Nash) equilibrium results in each person defecting each period. (You don't have to prove that a Nash equilibrium results in each person defecting at each stage; you just have to show that the player who always defects is not employing a dominant strategy.)

7 REPETITION AND EQUILIBRIUM

A short-run decision can affect a firm or individual's long-run reputation, and that makes it easier to devise incentives under which agents can maximize individual payoffs without precipitating an inefficient outcome. Specifically, cooperation can emerge when the players have an opportunity to punish anyone who sacrifices overall group welfare by pursuing short-run personal gain. With repeated play there will be future periods in which the punishment can take place. That means that there are equilibria in which each player faces a *credible* threat of punishment should he or she deviate from the path that results in an efficient outcome. Our intuition will be confirmed by the theory when the number of repetitions is infinite. In fact, just as there are typically many efficient outcomes, there are typically many equilibrium paths if the number of repetitions is infinite. All of this depends on there always *being* a future, which is not the case in the last period if the number of repetitions is finite.

We begin with the repeated prisoner's dilemma game, but first we recall the definition of a strategy from Section 5.7.

DEFINITION: Individual strategy

At any point in the game at which the player is allowed to move, the strategy specifies an action for that player for each potential history of the game to that point—and a single action for the game's opening move.

7.1 Repeated prisoner's dilemma with terminal date

If we stick to the assumption of selfishness at every turn, then the only equilibrium when the prisoner's dilemma game is played a fixed *finite* number of times

Table 1.20

C	
U	D
С, С	ℓ,h
h,ℓ	<i>d</i> , <i>d</i>
	c, c h, ℓ

has each person defecting at each stage, provided that both players know when play will end.

At each stage the players simultaneously and independently choose between defecting (*D*) and cooperating (*C*), and payoffs are then awarded according to Table 1.20, where $\ell < d < c < h$. This game is played exactly *T* times in succession. We assume that each individual's overall payoff from the repeated game takes the form

 $\alpha_1 u_1 + \alpha_2 u_2 + \cdots + \alpha_{T-1} u_{T-1} + \alpha_T u_T$

where u_t is the player's payoff in period t and α_t is some positive weight. (The weights can be different for different players.) What will happen?

At the *T*th and last stage there is only one possible outcome: Both defect because there is no further play, and thus no opportunity for their choices to affect future payoffs, and defecting is a dominant strategy for the one-shot game. Knowing that both will inevitably defect at the last stage, independent of what has happened previously, there can be no advantage to anyone who cooperates at the second-last stage—nothing will induce the opponent to cooperate in the last round. An individual's chosen action in period T - 1 then depends solely on the payoffs in that period, and we know that defecting is a dominant strategy in that context. Therefore, both will defect in round T - 1. Knowing that both will inevitably defect in the last two rounds, independently of what has happened previously, there can be no advantage to anyone who cooperates in stage T - 2. Therefore, both will defect in round T - 2, and so on. The only equilibrium has each person defecting at each stage.

Equilibrium theorem for the finitely repeated prisoner's dilemma

Both individuals will defect at each stage if there is a fixed number of repetitions, and both payers know when the game will end. This Nash equilibrium is subgame perfect.

The equilibrium is subgame perfect because the argument of the previous paragraph works for any subgame. (Section 5.7 defines subgame perfection.)

Even though the only Nash equilibrium in the finitely repeated prisoner's dilemma game has each person defecting each period, it is not true to say that defecting each period is a dominant strategy. This is not even true for two repetitions. Suppose that T = 2, $\alpha_1 = 1 = \alpha_2$, and we have c = 20, d = 5, $\ell = 1$, and h = 30 as in Table 1.15 at the beginning of Section 6. Suppose that *A* announces

7. Repetition and Equilibrium

his intention to cooperate in the first period, and then to cooperate again in the second period *if B* has also cooperated in the first period, and to defect in the second period if *B* defected in the first period. (I don't mean to imply that this is a smart decision on *A*'s part; it may or may not be.) This is called the *tit-for-tat* strategy. If *B* defects in both periods her payoff will be 30 + 5 but if *B* cooperates in the first period and then defects in the second period her payoff will be 20 + 30. (Will that give *A* second thoughts about playing tit-for-tat?) Given *A*'s tit-for-tat strategy, the cooperate-then-defect strategy gives *B* a higher total payoff than the defect-then-defect strategy, and therefore the latter is not a dominant strategy. Defecting both periods is, however, a payoff-maximizing response of *B* to the announcement by *A* that he will defect both periods. Therefore, we have not contradicted the assertion that defection each period by each player is a Nash equilibrium. (To prove that it is the *only* Nash equilibrium you need to do more.)

DEFINITION: The tit-for-tat strategy

The individual cooperates in period 1 and for any period *t* will cooperate in that period if the opponent has cooperated in the previous period, but will defect if the opponent defected in the previous period.

We have seen that the predictions of economic theory based on Nash equilibrium are not always confirmed by experiments and observations. An important contribution to the study of the finitely repeated prisoner's dilemma game, and hence to the understanding of the conditions under which cooperation will be induced by rational self-motivated behavior, is the competition devised by Robert Axelrod in which opponents formulated strategies for playing prisoner's dilemma. The strategies competed against each other in a round robin tournament in which each match consisted of repeated play of the prisoner's dilemma game. The tit-for-tat strategy, submitted by Anatol Rapoport, was the winner. Although it did not beat any other strategy it scored highest because it was a survivor: other strategies reduced each others' scores when pitted against each other.

7.2 Infinitely repeated prisoner's dilemma

Suppose that neither player knows when the interaction is going to end. We can model this by investigating a supergame in which the one-shot prisoner's dilemma game is played period after period without end. (The one-shot game is also called the stage game.) Even when there is a finite terminal date, having an infinite number of periods in the model is a good way to embody the fact that the players don't let that terminal date influence their behavior in the early and intermediate stages. We will see that when the game is played an infinite number of times there is an abundance of equilibria. As in any dynamic game, a strategy specifies one's choice at each stage as a function of the possible previous

choices of both players, so there is a vast number of strategies and many of these are equilibria.

Players *A* and *B* simultaneously and independently choose whether to cooperate or defect in each of an infinite number of periods 1, 2, ..., *t*. An individual's preferences are captured by the discounted sum of her payoff each period. That is, her period *t* payoff u_t is discounted by the factor δ^{t-1} , where $0 < \delta < 1$. Because $\delta < 1$, the discount factor will be close to zero if *t* is very large. (We simplify by assuming the same discount factor for each individual.) The individual's overall payoff from playing the infinitely repeated prisoner's dilemma game is

$$\Sigma \delta^{t-1} u_t = u_1 + \delta u_2 + \delta^2 u_3 + \dots + \delta^{t-1} u_t + \dots$$

If the game will end in finite time, but the individual does not know the terminal date, then we can view δ^{t+1} as proportional to the probability that period *t* is the last time the game will be played. Consequently, far distant dates have a very low probability of being reached.

If $\delta = (1 + r)^{-1}$ and *r* is the (positive) rate of interest then we certainly have $0 < \delta < 1$. In fact, if the individual can borrow and lend at the rate of interest *r*, and the payoffs from the stage game are in money terms, then the individual will act so as to maximize $\Sigma \delta^{t+1} u_t$ because that maximizes the right-hand side (the wealth term) of the individual's intertemporal budget constraint, without affecting the left-hand side (the expenditure term) of that constraint.

The generic one-shot game is again represented by Table 1.20 of the previous subsection. One equilibrium pair of strategies that induces universal cooperation when δ is sufficiently close to one has each person cooperating in the first period and cooperating every period thereafter as long as his opponent cooperated in all previous periods, but defecting every period subsequent to a defection by the opponent. This is called the *grim trigger strategy*.

DEFINITION: The grim trigger strategy

The individual cooperates in period one and any period *t* if the opponent has cooperated in every previous period. The individual defects in every period following a stage in which the opponent defected.

The name derives from the fact that a defection in any period triggers a severe punishment—defection by the other player in every subsequent period.

Consider the special case c = 20, d = 5, $\ell = 1$, and h = 30. Let's see why we have a Nash equilibrium if each adopts the grim trigger strategy. Suppose that player *B* uses the grim trigger strategy but player *A* cooperates in periods 1, 2, ..., t - 1 and defects in period *t*. Then player *B* will cooperate up to and including period *t* and defect every period thereafter. Now, compare *A*'s payoffs *discounted to period t* from the grim trigger strategy with the overall payoff from the deviation. The deviation produces a payoff of 30 in period *t* and at most 5 in

7. Repetition and Equilibrium

every subsequent period. Treating period t as "now," the discounted stream of payoffs, 30, 5, 5, ..., 5, ... is no larger than

$$30 + \delta 5 + \delta^2 5 + \delta^3 5 + \dots = 30 + \frac{5\delta}{1-\delta}.$$

Let S_t denote the sum $a + a\delta + a\delta^2 + \cdots + a\delta^{t-1}$ of t terms. Then $\delta S_t = a\delta + a\delta^2 + \cdots + a\delta^{t-1} + a\delta^t$. Then $S_t - \delta S_t = a - a\delta^t$. We can solve this equation for S_t . We get

$$S_t = \frac{a - a\delta^t}{1 - \delta}.$$

If $0 < \delta < 1$ then $a\delta^t$ approaches zero as t gets arbitrarily large. In that case, S_t gets arbitrarily close to $a/(1 - \delta)$ as t gets arbitrarily large. Then we can say that $a/(1 - \delta)$ is the sum of the infinite series $a + a\delta + a\delta^2 + \cdots + a\delta^{t-1} + \cdots$ if $0 < \delta < 1$. [In the case of $\delta 5 + \delta^2 5 + \delta^3 5 + \cdots$ we have $a = \delta 5$, and thus the sum is $5\delta/(1 - \delta)$.]

The trigger strategy, which has *A* and *B* cooperating every period, yields a discounted payoff of

$$20 + \delta 20 + \delta^2 20 + \delta^3 20 + \dots = \frac{20}{1-\delta}.$$

Deviation from this can be profitable for A only if

$$30+\frac{5\delta}{1-\delta}>\frac{20}{1-\delta},$$

which is equivalent to $\delta < 0.4$. (To discount to the present multiply the payoffs discounted to period *t* by δ^{t-1} . That will lead to the same inequality.) Therefore, if $\delta \ge 0.4$ and both play the grim trigger strategy we have a Nash equilibrium.

If $\delta = (1 + r)^{-1}$ then $\delta < 0.4$ is equivalent to r > 1.5: Only when the interest rate is greater than 150% can it be profitable for a player to deviate from the trigger strategy, which induces cooperation in each period. The cooperative outcome can be sustained as long as the players are not inordinately impatient.

The grim trigger strategy equilibrium is not subgame perfect. Consider *A*'s payoff in the subgame following the choice of *D* by *A* in period *t*, with each playing *C* in each prior period, and *B* playing *C* in period *t*. The trigger strategy has *A* playing *C* in the first period of the subgame (period t + 1 of the parent game) and *B* playing *D* to punish *A* for the choice of *D* in period t + 1. Then both will play *D* in every subsequent period and hence *A*'s payoff stream from period t + 1 on will be ℓ , d, $d \dots d$, \dots If *A* were to deviate slightly and play *D* from period t + 1 on then her payoff stream would be d, d, d, \dots , d, \dots and that is better for *A* for any (positive) value of the discount factor.

We can modify the grim trigger strategy slightly to produce a subgameperfect equilibrium that sustains cooperation: Have each individual cooperate in period one and in any period *t* if the opponent has cooperated in every previous period, but have the individual defect in every period following a stage in which *either* player defected.

If $c > (h + \ell)/2$ then the cooperative outcome also results from the more conciliatory *tit-for-tat* strategy that has a player cooperating in the first period, and in every subsequent period playing whatever strategy the opponent employed in the previous period. We investigate this claim for the generic stage game.

DEFINITION: The tit-for-tat strategy

The individual cooperates in period 1 and for any period *t* will cooperate in that period if the opponent has cooperated in the previous period, but will defect if the opponent defected in the previous period.

We put ourselves in the shoes of player *A*. Assume that *B* is playing tit-for-tat. Let's see if tit-for-tat is a best response. When both play tit-for-tat each will wind up choosing *C* every period, so each will get *c* every period, and hence *A*'s overall payoff will be $c/(1 - \delta)$. Now, suppose that *A* chooses *D* in every period. Then *A* will get *h* in the first period but *d* ever after because *B* will play *D* in every period after the first. The resulting overall payoff to *A* will be

$$h+\delta d+\delta^2 d+\delta^3 d+\cdots=h+rac{\delta d}{1-\delta}.$$

The overall payoff from tit-for-tat will be higher if $c/(1 - \delta) > h + \delta d/(1 - \delta)$, and that is equivalent to $\delta > (h - c)/(h - d)$. (Note that h - c is less than h - d.) If δ is sufficiently close to 1 (i.e., if *A* is sufficiently patient) then playing tit-for-tat gives *A* a higher payoff than defecting every period—when *B* plays tit-for-tat.

Consider a different strategy for *A*: Suppose that *A* were to defect in the first period but then cooperate in period 2 and play tit-for-tat thereafter. *B* is playing tit-for-tat from the start, so *A*'s sequence of actions will be *D*, *C*, *D*, *C*,

$$h + \delta\ell + \delta^{2}h + \delta^{3}\ell + \dots = h + \delta^{2}h + \delta^{4}h + \delta^{6}h \dots + \delta\ell + \delta^{3}\ell + \delta^{5}\ell + \delta^{7}\ell \dots$$

The sum of the terms involving h is $h/(1 - \delta^2)$ and the sum of the terms involving ℓ is $\delta \ell/(1 - \delta^2)$. When do we have

$$\frac{c}{1-\delta} > \frac{h}{1-\delta^2} + \frac{\delta\ell}{1-\delta^2}$$

Multiplying both sides of the inequality by the positive number $1 - \delta_2$ reveals that it is equivalent to

$$c(1+\delta) > h + \delta\ell.$$

Because $c > (h + \ell)/2$ (by assumption), $2c > h + \ell$ and hence for $\delta < 1$ arbitrarily close to 1 we have $c(1 + \delta)$ arbitrarily close to 2c, and hence $c(1 + \delta) > h + \delta \ell$. Therefore, if player *B*'s strategy is tit-for-tat then *A* cannot do better than tit-for-tat by alternating between *D* and *C*.
7. Repetition and Equilibrium

Suppose that player *A* waits until period *t* to begin defecting every period or to begin alternating between *D* and *C*. In that case, the deviant strategies will have the same payoff for *A* as tit-for-tat up to and including period t - 1. We can then discount future-stage game payoffs back to period *t* and this will lead us back to the respective inequalities of the previous two paragraphs.

For a sufficiently high discount rate, tit-for-tat is a superior response to titfor-tat than defecting every period or alternating between D and C. What about other strategies? Continue to assume that *B* plays tit-for-tat but suppose that *A* plays C in each of the first three periods, then plays D for the next two periods, and then C every period thereafter. Compare A's overall payoff from this strategy to tit-for-tat. Each pays *c* in each of the first three periods and *c* in the seventh period and every subsequent period. Therefore, we can begin the comparison at period 4 and end it in period 6, discounting the payoffs to period 4: The deviant strategy yields $h + \delta d + \delta^2 \ell$ and tit-for-tat yields $c + \delta c + \delta^2 c$. We can make $h + \delta^2 \ell$ as close as we like to $h + \ell$ and $c + \delta^2 c$ as close as we like to 2c by taking δ sufficiently close to 1. Then with $c > (h + \ell)/2$ we have $c + \delta^2 c > h + \delta^2 \ell$ for δ sufficiently close to 1. And $\delta c > \delta d$ for any $\delta > 0$. Therefore, tit-for-tat is superior to the deviant strategy. If A were to play C for three periods then D for *n* periods and then *C* ever after we would have essentially the same argument, except that we would compare $h + \delta^n \ell$ to $c + \delta^n c$, but again with $c > (h + \ell)/2$ we have $c + \delta^n c > h + \delta^n \ell$ for δ sufficiently close to 1. It is clear that any strategy that elicits a different stream of payoffs from tit-for-tat will be inferior to tit-for-tat if the opponent plays tit-for-tat and if δ is sufficiently close to 1.

There are many other equilibria. For instance, if both players announce that they will defect at every stage regardless of their opponent's behavior then an individual's payoff will fall if he or she does not defect in each period: The opponent will defect in period t, so playing C leads to a lower payoff in that period than D, and hence to a lower overall payoff. Moreover, in this case choosing C will not induce the opponent to act in a future period in a way that enhances the player's overall payoff. Note that this argument is valid for any value of the discount factor δ . Note also that this equilibrium gives each person a payoff of d each period, whereas our first two equilibria leave each person with c at each stage.

Surprisingly, we can use a variant of the grim trigger strategy to sustain a wide range of overall payoffs at equilibrium. Let S_A be any strategy for A and let S_B be any strategy for B. Player i can threaten to punish player j if the latter doesn't follow the pattern of choices prescribed by S_j .

DEFINITION: The grim trigger strategy in general

Individual A performs the actions required by S_A in each period, as long as B has performed the actions required by S_B in each previous period. But if B deviates from S_B in some period then A will defect in every subsequent period. (The grim trigger strategy is defined analogously for player B.)

60 Equilibrium, Efficiency, and Asymmetric Information

The grim trigger strategy will induce A to behave according to S_A and B to act according to S_B provided that (i) each would prefer the resulting overall payoff to a payoff of d every period after some time t, and (ii) each is sufficiently patient. The next section clarifies condition (i), but the following example illustrates.

Example 7.1: A run of Cs and Ds

Suppose that S_A and S_B each have the player cooperating for three periods and then defecting for two periods, and then repeating the cycle indefinitely. The string of payoffs for each is *c*, *c*, *c*, *d*, *d*, *c*, *c*, *c*, *d*, *d*... If *A* deviates from S_A in a period *t* when *B* chooses *C* then *A* will get *h* instead of *c*. However, if δ is sufficiently close to one we can think of the overall payoff from S_A starting from period *t* as

$$c+c+c+d+d+c+c+c+d+d+\cdots$$

If *B* employs the grim trigger strategy we can think of the deviation as precipitating at best

$$h+d+d+d+d+d+d+d+d+d\cdots$$

for player *A*. The resulting period *t* gain of h - c from the deviation is swamped by infinite number of periods in which *A* gets *c* instead of *d* by following *S*_{*A*}. Therefore (*S*_{*A*}, *S*_{*B*}) augmented by the grim trigger strategy is a Nash equilibrium for δ sufficiently large.

To turn any Nash equilibrium based on the grim trigger strategy into a subgame-perfect equilibrium we just have to require the individual to play D every period following a deviation from the prescribed behavior S_A or S_B by *either* A or B, respectively.

7.3 Equilibrium theorem for infinitely repeated games

The argument of Section 7.2, showing that a wide range of strategy pairs can be sustained as a Nash equilibrium of the infinitely repeated prisoner's dilemma game, is easy to generalize to the infinite replication of any n-person game in which each person has a finite number of available actions.

Let's quickly review the prisoner's dilemma case: The strategy "play *D* every period" yields a payoff of at least *d* every period—the payoff will be either *d* or *h*—and thus yields an overall payoff of at least $d/(1 - \delta)$. The grim trigger strategy, which relegates a player to *d* one period after a defection *and* every subsequent period, can be used to induce each player to stick to a given strategy, provided that it yields a higher overall payoff than $d/(1 - \delta)$. (This assumes a discount factor δ sufficiently close to one.) We refer to *d* as a player's *security level* in the prisoner's dilemma game. By playing *D* an individual is assured of getting at least *d*, and she could wind up with less by playing *C*.

To generalize the argument to an arbitrary one-shot *n*-person game we need do little more than identify arbitrary player *i*'s security level. In general, it is a stage game payoff *m* such that *i* can guarantee that her payoff is at least *m*

7. Repetition and Equilibrium

by choosing some action *M*. Specifically, for each assignment *a* of actions to the other players, let max(*a*) be *i*'s payoff from her best response to *a*. Now let \hat{a} be the assignment of actions to players other than *i* that minimizes max(*a*). Then $m = \max(\hat{a})$. (We're assuming that each player has a finite set of available actions.)

DEFINITION: The individual's security level

Given a one-shot game, let *A* denote the set of all logically possible assignments of actions to everyone but *i*, and for each *a* in *A* let max(*a*) be the highest payoff that *i* can achieve when the others play *a*. Then if \hat{a} minimizes max(*a*) over all *a* in *A* we set $m = \max(\hat{a})$ and refer to it as player *i*'s security level.

As in Section 7.2 we assume that an individual evaluates the stream of payoffs resulting from an infinite number of plays of a one-shot game by discounting to the present. If the player's period *t* payoff is u_t (where t = 1, 2, 3, ...) and δ is her discount factor, then her overall payoff is

$$\Sigma \delta^{t-1} u_t = u_1 + \delta u_2 + \delta^2 u_3 + \dots + \delta^{t-1} u_t + \dots$$

The equilibrium theorem for infinitely repeated games establishes that any pattern of actions in the infinitely repeated game that allows each player to do better than her security level can be precipitated by some Nash equilibrium, provided that all players evaluate their payoff streams with a discount factor sufficiently close to one. That's because the generalized grim trigger strategy can be used to prevent anyone from deviating from the given course of action. To prove this we need to clarify the statement "doing better than." To this end we need some preliminary notation and definitions.

A strategy profile *S* for a repeated *n*-person game assigns a strategy S_i to each player *i*. Given a strategy profile *S*, for arbitrary individual *i*, and arbitrary period *t* we let U_t^t (*S*) be *i*'s payoff stream from period *t* on, discounted to period *t*, assuming that each individual *j* employs S_j . Let U_i^t be the value of the stream of *i*'s security-level payoffs, received every period, discounted to period *t*.

Equilibrium theorem for infinitely repeated games

Let *S* be a strategy profile such that $U_i^t(S) > U_i^t$ for each individual *i* for each period *t*. If each individual is sufficiently patient there is a subgame-perfect Nash equilibrium in which each individual *i* behaves according to S_i at equilibrium.

To show that *S* is sustained by a Nash equilibrium we just have each individual employ a generalized trigger strategy. Each individual *i* follows S_i provided that no individual *j* has deviated from S_i in the past. After a deviation by any

62 Equilibrium, Efficiency, and Asymmetric Information

individual, every other person employs the action that, collectively, drives the deviating individual to his or her security level and takes that action in every subsequent period, ad infinitum. This typically does not give us a *subgame*-perfect equilibrium, but it is possible to refine the strategies to give each player an incentive to punish anyone who does not do his or her part in punishing someone who deviates from the behavior prescribed for him or her by *S*, and thereby to justify the use of the adjective *subgame perfect*. (This is easy to do for the repeated prisoner's dilemma because the individual security levels emerge from the unique Nash equilibrium of the stage game.)

To bring out the significance of the equilibrium theorem, we next explore a version of the prisoner's dilemma game that gives rise to a continuum of Nash equilibria in the infinite replication game, even though the stage game has a single dominant strategy equilibrium: Each player must choose *a level of cooperation between zero and one*—not necessarily just an extreme point, zero (defect) or one (full cooperation).

Example 7.2: The continuum dilemma

In the stage game player *A* selects a fraction α ($0 \le \alpha \le 1$) and *B* selects a fraction β ($0 \le \beta \le 1$). Each person's fraction expresses the degree of cooperation chosen. The payoffs are defined so that $\alpha = 0$ is a dominant strategy in the stage game for *A*, and similarly $\beta = 0$ is a dominant strategy for *B*. Set

$$u_A(\alpha, \beta) = 2\beta - \alpha$$
 and $u_B(\alpha, \beta) = 2\alpha - \beta$.

These payoff functions can be given a simple interpretation. *A* and *B* are neighbors, and each is bothered by the amount of debris that motorists deposit as they drive by. If either *A* or *B* supplies *e* units of effort to cleaning up the trash then *each* will receive 2*e* units of utility from the improved appearance of the neighborhood. But cleanup is costly, and for each unit of effort expended by *A* there is a utility cost of 3 units. Similarly for *B*. Then if *A* devotes α units of effort to cleanup while *B* contributes β , then *A*'s utility is $2(\alpha + \beta) - 3\alpha$ and *B*'s utility is $2(\alpha + \beta) - 3\beta$. This gives us the previous payoff functions. Whatever the value of β , player *A* can increase u_A by reducing α . Therefore, $\alpha = 0$ is a dominant strategy for *A* in the stage game. Similarly, $\beta = 0$ is a dominant strategy for *B*.

What are the feasible payoff vectors for this game? They comprise the entire diamond OKLM in Figure 1.7 (including the interior). Consider point *x*, which is a convex combination of (-1, 2) and (1, 1). That is, $x = \lambda(-1, 2) + (1 - \lambda)(1, 1)$ for some value of λ between zero and unity. In plainer terms, the first component of x (*A*'s payoff) is $\lambda(-1) + (1 - \lambda)(1)$ and the second (*B*'s payoff) is $\lambda(2) + (1 - \lambda)(1)$. Can we have

$$2\beta - \alpha = -\lambda + (1 - \lambda)$$
 and $2\alpha - \beta = 2\lambda + (1 - \lambda)$?

The solution of these equations is $\alpha = 1$ and $\beta = 1 - \lambda$, and both are admissible strategies.





Consider $y = \lambda(1, 1) + (1 - \lambda)(2, -1)$. Set

 $2\beta - \alpha = \lambda + 2(1 - \lambda)$ and $2\alpha - \beta = \lambda - (1 - \lambda)$.

The solution is $\alpha = \lambda$ and $\beta = 1$; both are admissible. Verify that

$$z = \lambda(-1, 2) + (1 - \lambda)(0, 0) \text{ results from } \alpha = \lambda \text{ and } \beta = 0,$$

$$w = \lambda(2, -1) + (1 - \lambda)(0, 0) \text{ results from } \alpha = 0 \text{ and } \beta = \lambda,$$

$$p = \lambda(-1, 2) + \theta(1, 1) + (1 - \lambda - \theta)(2, -1) \text{ if } \alpha = \lambda + \theta \text{ and } \beta = 1 - \lambda,$$

$$q = \lambda(-1, 2) + \theta(2, -1) + (1 - \lambda - \theta)(0, 0) \text{ if } \alpha = \lambda \text{ and } \beta = \theta.$$

In each case $0 \le \lambda \le 1$, and for *p* and *q* we have $0 \le \theta \le 1$ and $0 \le \lambda + \theta \le 1$ as well.

To summarize, any point in the diamond OKLM in Figure 1.7 is a feasible payoff assignment for the one-shot version of the game. The equilibrium theorem says that any point (u_A, u_B) in the shaded part of the diamond, excluding the lines OH and OJ, can be sustained as a subgame-perfect Nash equilibrium in which *A* gets u_A each period and *B* gets u_B each period in the infinitely repeated game. At least, that will be the case if the discount rate is sufficiently high. We have seen why these payoffs can be supported by a Nash equilibrium. The grim trigger strategy permanently reduces an opponent to a utility of zero if he once deviates from the equilibrium degree of cooperation. It is easy to see why a point

64 Equilibrium, Efficiency, and Asymmetric Information

outside of the shaded area cannot be sustained, even with infinite replication and a high discount rate. Outside of the shaded area one person receives less than zero, but a player can always guarantee a payoff of at least zero per period by selecting a cooperation level of zero each period.

Although we can show how cooperation might be sustained in an infinitely repeated game, there are many other equilibria as well. We certainly have not been able to show that cooperation is inevitable.

7.4 Terminal date and unknown type

Section 7.1 demonstrated that if the repeated prisoner's dilemma game has a know finite terminal date then each player will defect each period. That depends on the supposition that player *B* knows that *A* will always play a best response to *B*'s strategy, and *A* knows that *B* knows this, and that *B* knows that *A* knows that *B* knows this, and so on. What if *B* believes that there is a small but positive probability that *A* is committed to the tit-for-tat strategy, even when it is not a best response to what *B* has done? We now show how this opens the door for

Table 1.21

Perio	Period 2	
Opponent	Opponent Player	
D	D	D
D	D	С
С	D	С
С	D	D
D	С	С
С	С	D
D	С	D
С	С	С

cooperation by *both* even when both players know the finite terminal date. (See Section 7.2 for a definition of tit-for-tat.)

We again begin with the generic one-shot prisoner's dilemma game of Table 1.20 in Section 7.1. Recall that $\ell < d < c < h$. To make the analysis more transparent we will not discount: Each player wants to maximize the sum of the payoffs over the lifetime of the repeated game. Player *A* is one of two types, but *B* does not know which type *A* actually is when play begins. There is a positive probability π that *A* is a cooperative type who can only play the *titfor-tat* strategy. Tit-for-tat cooperates in the first period and for every subsequent period duplicates the move made by the opponent at

the previous stage. In that case we say that *A* is type *Q*. With probability $1 - \pi$ player *A* is "rational" (type *R*), which means that *R* can play any strategy, and *R* knows that this is also true of *B*. (Section 7.1 showed that if both players are rational, and both know this, then the unique Nash equilibrium has each player defecting each period.) Player *B* is assumed to be risk neutral, which means that *B* wants to maximize $\pi q(s) + (1 - \pi)r(s)$, where q(s) is the payoff that *B* gets from strategy *s* if he's actually playing against *Q*, and r(s) is his payoff from *s* should he be playing against *R*.

With two choices available to each player there are four logically possible pairs of first-round decisions, and for each there are two possible responses for a given player. These eight cases are displayed as eight rows in Table 1.21. The first four cases will not arise when the player is type *Q* because *Q* always begins by cooperating. The tit-for-tat strategy is represented by the last two lines with *Q* as

Table 1	.22
---------	-----

	Peri	od 1	Peri	od 2	<i>B</i> 's total	
A's type	A	В	A	В	payoff	Probability
Q	С	С	С	D	c+h	π
R	D	С	D	D	$\ell + d$	$1-\pi$
Q	С	D	D	D	h+d	π
R	D	D	D	D	d + d	$1 - \pi$

the player. What will R do in equilibrium? R could begin by cooperating to fool B into thinking that she (A) is type Q. But in a two-period model this will not work. R knows that B will defect in period 2. Therefore, R will defect in both periods. Q will cooperate in the first period and select X in the second period, where X is B's first period choice. It remains to determine B's first period move X.

There are two possibilities: X = C (cooperate) and X = D (defect); for each of these *R*'s move is uniquely determined in each period and so is *Q*'s move in each period. Therefore, there are $2 \times 2 = 4$ cases, displayed as Table 1.22. If *B* cooperates in period 1 his overall payoff is $\pi(c + h) + (1 - \pi)(\ell + d)$, the sum of the column 4 payoffs weighted by the probabilities. But if *B* defects at the outset his overall payoff is $\pi(c + h) + (1 - \pi)(\ell + d)$. Cooperation in period 1 leads to a higher payoff for *B* when $\pi(c + h) + (1 - \pi)(\ell + d) > \pi(h + d) + (1 - \pi)(d + d)$, and this reduces to

$$\pi > \frac{d-\ell}{c-\ell}$$

Set $\pi^0 = (d - \ell)/(c - \ell)$. As long as $\pi > \pi^0$ the equilibrium strategies for the twoperiod game are as follows:

- Q: tit-for-tat.
- *R*: defect each period, whatever *B* does.
- *B*: cooperate in the first period, then defect, whatever *A* does.

Because *d* is smaller than *c*, the threshold π^0 decreases when ℓ increases. (If 0 < x < y < z then (y - x)/(z - x) is less than y/z.) Consider $\ell = d - \epsilon$, with ϵ positive but very small. Then π^0 will be arbitrarily small. That is, if we make ℓ sufficiently close to *d* we only need a tiny probability that *A* is tit-for-tat to sustain

Table	1.	23A
-------	----	-----

Туре	Period 1	Period 2	Period 3	Payoff
Q	С	D	D	$\ell + d + d$
В	D	D	D	h + d + d
R	С	D	D	$\ell + d + d$
В	D	D	D	h + d + d

Туре	Period 1	Period 2	Period 3	Payoff
Q	С	D	D	$\ell + d + d$
B	D	D	D	h + d + d
R	D	D	D	d + d + d
В	D	D	D	d + d + d

1.23E

some cooperation in equilibrium. That's because the cost to *B* of cooperating in the first period against someone who defects is very small if ℓ is close to *d*.

Consider a three-period replication. It is conceivable that *R* will open by cooperating to build a reputation for cooperating and so induce *B* to cooperate. We know that *R* and *B* will both defect in the last period, so there is no value to *R* in cooperating beyond the first period. Either *R* will defect every period or else *R* will cooperate in period 1 and defect in the other two periods. Therefore, the only decision to be specified for *R* is the first-period move. But if *R* defects on the first move *B* will know for sure that he is not playing against *Q* and will thus defect in each of the last two periods (because he will know that *A* will defect in each of the last two stages). Let's see if cooperation by *R* on the first move can be sustained at equilibrium. There are two possibilities, and for each of these there are four possible moves for *B*. (We know that *B* will defect on the last round.) Therefore, there are $2 \times 4 = 8$ cases to consider, represented by Tables 1.23A–1.26B. For each of the four table numbers, the *A* and *B* tables differ only with respect to *R*'s first period move.

Table 1.25A has *R* cooperating in the first period and *B* cooperating in the first two periods. Let's work out the conditions on π such that Table 1.25A is observed at equilibrium. The strategies underlying this table are as follows:

- S_R : *R* cooperates in the first period and defects in each of the other two periods, whatever *B* does.
- S_Q : *Q* cooperates in the first period and then imitates *B*'s previous move in each of the subsequent periods.
- S_B : *B* cooperates in period 1 and defects in the other two periods if *A* defects in period 1: If *A* cooperates in the first period then *B* will cooperate in the second period and defect in the last period.

Туре	Period 1	Period 2	Period 3	Payoff
Q	C	D	C	$\ell+h+\ell$
B	D	C	D	$h+\ell+h$
R	C	D	D	$\ell+h+d$
B	D	C	D	$h+\ell+d$

Table 1.24A

7. Repetition and Equilibrium

Туре	Period 1	Period 2	Period 3	Payoff
Q	С	D	С	$\ell + h + \ell$
В	D	С	D	$h + \ell + h$
R	D	D	D	d+h+d
В	D	С	D	$d+\ell+d$

Table 1.25A

Туре	Period 1	Period 2	Period 3	Payoff
Q	C	C	C	$c + c + \ell$
B	C	C	D	c + c + h
R	C	D	D	c + h + d
B	C	C	D	$c + \ell + d$

Table 1.25B

Туре	Period 1	Period 2	Period 3	Payoff
Q	C	C	C	$c + c + \ell$
B	C	C	D	c + c + h
R	D	D	D	h + h + d
B	C	C	D	$\ell + \ell + d$

Table 1.26A

Туре	Period 1	Period 2	Period 3	Payoff
Q	C	C	D	$c + \ell + d$
B	C	D	D	c + h + d
R	C	D	D	c + d + d
B	C	D	D	c + d + d

Туре	Period 1	Period 2	Period 3	Payoff
Q	С	С	D	$c + \ell + d$
В	С	D	D	c + h + d
R	D	D	D	h + d + d
В	С	D	D	$\ell + d + d$

Table 1.26B

Will it be profitable for *R* to deviate from S_R ? If *R* defects in period 1 then *B* will know in period 2 that *A* is type *R* and will defect in periods 2 and 3. Therefore, a deviation by *R* will take us to Table 1.26B. This deviation will be unprofitable for *R* if h + d + d < c + h + d, and that is equivalent to d < c, which is always the case for a prisoner's dilemma game. We don't have to consider a deviation by *Q* from S_O because, by assumption, *Q* can only play tit-for-tat.

Will *B* deviate from S_B , given that *R* plays S_R ? Note that *R* opens by playing *C*, which restricts us to the A tables. The four A tables differ only with respect to *B*'s actions (and any effect that may have on the tit-for-tat player's actions). Table 1.27 gives *B*'s payoff from each of the four A tables, given that *R* cooperates in the first period and defects in the other two, regardless of what *B* does. The third line results from S_B , and the other three lines result from the possible deviations by *B*.

We have $c + \pi (c + h) + (1 - \pi)(\ell + d) > h + 2d$ as long as

$$\pi > \frac{h+d-\ell-c}{h+c-\ell-d}.$$
[1]

We have $c + \pi (c + h) + (1 - \pi)(\ell + d) > h + \ell + \pi h + (1 - \pi)d$ as long as

$$\pi > \frac{h-c}{c-\ell} \tag{2}$$

And we have $c + \pi (c + h) + (1 - \pi)(\ell + d) > c + d + \pi h + (1 - \pi)d$ as long as

$$\pi > \frac{d-\ell}{c-\ell} \tag{3}$$

Table 1.27

B's strategy	B's payoff
From Table 23A	$\pi (h+2d) + (1-\pi)(h+2d) = h+2d$
From Table 24A	$\pi (2h+\ell) + (1-\pi)(h+\ell+d) = h+\ell + \pi h + (1-\pi)d$
From Table 25A	$\pi (2c+h) + (1-\pi)(c+\ell+d) = c + \pi (c+h) + (1-\pi)(\ell+d)$
From Table 26A	$\pi (c+h+d) + (1-\pi)(c+2d) = c + d + \pi h + (1-\pi)d$

7. Repetition and Equilibrium

Therefore, we have an equilibrium with cooperation in the early stages of the game (for one period by *R* and two periods by *B*) as long as [1], [2], and [3] hold. Notice that [3] is the condition for cooperation by *B* in the first period of a two-period game.

We have examined reputation building as a motivation for behavior that promotes social welfare. Is *R* creating a *false* reputation by cooperating on the first move? No. Defecting every period is implied by rationality only when there is a finite number of repetitions with a known terminal date and *R*'s opponent knows that he is rational. But if *B* doesn't know *A*'s type then the optimal strategy for rational *A* will be affected. In the two-period model it is only *B* who cooperates at all (for one period). But as the three-period case shows, if *B* can be induced to cooperate—because of his uncertainty about *A*'s type—then *A* has an incentive to build a reputation as a cooperative player, even if *A* is actually type *R*.

Let's apply conditions [1], [2], and [3] to the case c = 20, d = 5, $\ell = 1$, and h = 30. We get $\pi > 7/22$, $\pi > 10/19$, and $\pi > 4/19$. Therefore, *B* has to believe that the probability of *A* being tit-for-tat is greater than 10/19 to be induced to cooperate in the first two periods. However, as the number of repetitions increases the greater the long-run payoff to cooperative behavior, and hence smaller values of π will sustain cooperation. Note that $\pi > 4/19$ is sufficient to induce *B* to cooperate in the first period of the *two*-stage game, whereas $\pi > 10/19$ is the sufficient condition for (S_T , S_R , S_B) to be an equilibrium in the three-stage game. Don't be misled into thinking that cooperation—two periods instead of one period—when $\pi > 10/19$.

Sources

The repeated prisoner's dilemma competition devised by Robert Axelrod is reported in Axelrod (1984). An early version of the equilibrium theorem for infinitely repeated games was proved by Friedman (1971). The treatment of the prisoner's dilemma game when individual types are unknown is based on Gibbons (1992, p. 225). Kreps et al. (1982) actually prove that, given π , if there is a large number of periods then the players will cooperate in every period until they are close to the terminal period.

Links

For more on the prisoner's dilemma game replayed many times see Rapoport, 1989. See Calvert (1986, pp. 47–54) for related treatments of reputation in economics and politics. Osborne (2004, pp. 439–41) provides a very good assessment of Axelrod's tournament. Fudenberg and Maskin (1986) and Wen (1994) contain significant generalizations of the equilibrium theorem (called the folk theorem in the literature).

Limitations in the information processing capacity of the players can eliminate a lot of Nash equilibria of the infinitely repeated game, appearing to make cooperation more likely in the prisoner's dilemma case. In particular, see Rubinstein (1986 and 1998) and Binmore and Samuelson (1992).

70 Equilibrium, Efficiency, and Asymmetric Information

Problem set

- 1. Prove that the grim trigger strategies constitute a Nash equilibrium of the generic version of the infinitely repeated game provided that each individual is sufficiently patient.
- **2.** Let *A*'s discount rate be 0.9 and let *B*'s be 0.7. Find a condition guaranteeing that cooperation every period by both players is the outcome of a subgame-perfect Nash equilibrium when the payoffs in the stage game are given by Table 1.15 in Section 6.
- **3.** Rework the argument of Section 7.2 for the specific case c = 20, d = 5, $\ell = 1$, and h = 30.
- **4.** Does the analysis of Section 7.4 change if we replace the tit-for-tat by the strategy "cooperate every period whatever the opponent does?"

2

Basic Models and Tools

1.	Мах	kimizing a Quadratic	72
	1.1	Unconstrained maximization	72
	1.2	Constrained maximization	75
∂2 .	Ove	rview of Calculus	
	2.1	Unconstrained maximization	76
	2.2	Constrained maximization	80
	2.3	Strictly concave functions	82
	2.4	Minimization	84
	2.5	The tangency approach	84
	2.6	The total derivative and the chain rule	85
		Problem set	86
3.	Lag	rangian Multipliers	86
	∂ 3.1	The Lagrangian multiplier with a single resource constraint	87
	∂ 3.2	Remark on planning and Lagrangians	87
	∂ 3.3	Lagrangian multipliers with more than one resource constraint	e 88
	3.4	The converse of the Pythagorean theorem Problem set	96 97
4.	The	Composite Commodity Model	98
	4.1	The budget constraint and preferences	99
	4.2	The composite commodity theorem	99
5.	Qua	si-Linear Preferences	102
	5.1	Efficiency with quasi-linear utility	103
	∂ 5.2	Quasi-linear preference and demand	106
	∂ 5.3	Consumer surplus	108
6.	Dec	ision Making Under Uncertainty	112
	6.1	Asset preferences	112
	6.2	Risk aversion and risk neutrality	115

72 Basic Models and Tools

∂ 6.3	Risk aversion and a negative second derivative	118
6.4	The market opportunity line	119
6.5	The uniform probability distribution	121
∂ 6.6	The continuum case in general	122
	Problem set	122
7. Insu	rance	. 124
7.1	The complete insurance theorem	125
7.2	Sketch of the proof of the complete insurance	
	theorem	125
∂ 7.3	Calculus proof of the complete insurance theorem	126
7.4	Competitive insurance markets	130
7.5	Efficiency of competitive insurance markets with	
	full information	132
	Problem set	134

1 MAXIMIZING A QUADRATIC

There are about a hundred worked examples in this book, and many of them conclude with a simple exercise—finding the value of x that maximizes a quadratic function of the form $Px - Qx^2 + R$. This is a very simple procedure that does not require calculus. This section shows you how to determine whether a maximum exists, and if it does, how to quickly compute that maximizing value of x as a simple function of P and Q. We begin by assuming that x is unconstrained and then conclude with an examination of the solution to the constrained maximization problem, for which x must lie between the numbers a and b, inclusive.

Danger: The formula for the *root* of the quadratic equation $ax^2 + bx + c$ is

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

This gives the two values of *x* for which the value of the function is *zero*. In this section we seek to *maximize* the value of a quadratic function.

1.1 Unconstrained maximization

Consider the basic consumer decision problem, which requires the maximization of a utility function U(x, y) subject to the simple budget constraint $p_1x + p_2y = \theta$. We can turn that into an unconstrained maximization problem by solving the budget constraint for *y* as a function of *x*. Then we can substitute this expression for *y* in the utility function to get a function of a single variable *x*.

Example 1.1: Consumer choice with quadratic utility

The budget constraint is 4x + 2y = 12. Then 2y = 12 - 4x and thus y = 6 - 2x. The given utility function is

$$U(x, y) = 64x - x^2 + 3y.$$

If we substitute 6 - 2x for y in the utility function we get

$$U = 64x - x^{2} + 3[6 - 2x] = 64x - x^{2} + 18 - 6x = 58x - x^{2} + 18$$

The final expression depends on x alone because it has the budget constraint built in. If we maximize

$$f(x) = 18 + 58x - x^2$$

then we will have solved the problem of maximizing utility subject to the budget constraint.

The function $f(x) = 18 + 58x - x^2$ is called a *quadratic* because there is only one unknown, and the highest power of the unknown is a squared term. Functions of a single variable will also come up in other contexts, such as profit maximization by firms and determining the efficient level of output of a public good. The purpose of this section is to teach you how to quickly find the value of *x* that maximizes a quadratic of the form

$$Px - Qx^2 + R$$
 where $Q > 0$.

That is, P and Q are given real numbers, and Q is positive.

Example 1.2: Maximizing the function
$$25 - (x - 3)^2$$

$$f(x) = 25 - (x - 3)^2$$

Note that if *x* is not equal to 3 then $(x - 3)^2$ is positive, which means that we subtract a positive amount from 25. Therefore, the best that we can do, if we want to maximize f(x), is to make sure that $(x - 3)^2$ is zero. There is only one value of *x* for which $(x - 3)^2 = 0$, and that is x = 3. Therefore, x = 3, and only x = 3, maximizes *f*. Note that this function can be written in the form $Px - Qx^2 + R$: Because $(x - 3)^2 = x^2 - 6x + 9$ we have

$$f(x) = 25 - (x - 3)^2 = 6x - x^2 + 16.$$

That is, P = 6, Q = 1, and R = 16.

When there is one and only one value that maximizes f(x) we say that it is a unique global maximum.

DEFINITION: Global maximum

The number x^* is a global maximum of f if $f(x^*) \ge f(x)$ for every real number x. And x^* is a *unique* global maximum if $f(x^*) > f(x)$ for every real number x distinct from x^* .

The function $f(x) = 6x - x^2 + 16$ can be rewritten as $f(x) = 25 - (x - 3)^2$ to allow us to apply the argument of Example 1.2. What about other cases?

Example 1.3: Maximizing the function $6x - \frac{1}{2}x^2 + 82$

We are given $f(x) = 6x - \frac{1}{2}x^2 + 82$. But $6x - \frac{1}{2}x^2 + 82 = 82 - \frac{1}{2}(x^2 - 12x) = 82 - \frac{1}{2}(x - 6)^2 + 18$. Hence $f(x) = 100 - \frac{1}{2}(x - 6)^2$.

If $x \neq 6$ then $(x - 6)^2$ is positive, and hence a positive amount is subtracted from 100. Therefore f(x) has a unique global maximum at x = 6.

Now return to our generic function $Px - Qx^2 + R$ with Q > 0. Rewrite this as

$$f(x) = R - Q\left(x^2 - \frac{P}{Q}x\right).$$

Note that $[x - (P/2Q)]^2$ is similar to $[x^2 - (P/Q)x]$. If we distribute $[x - (P/2Q)]^2$ we get

$$\left(x - \frac{P}{2Q}\right)^2 = x^2 - \frac{P}{Q}x + \frac{P^2}{4Q^2}$$

Therefore, if in f(x) we replace $-Q[x^2 - (P/Q)x]$ by $-Q[x - (P/2Q)]^2$ we have to add $Q \times (P^2/4Q^2)$ to preserve the value of f(x). To summarize, we have

$$f(x) = R - Q\left(x^2 - \frac{P}{Q}x\right) = R + \frac{P^2}{4Q} - Q\left(x - \frac{P}{2Q}\right)^2.$$

The only part of this that is influenced by *x* is $-Q[x - (P/2Q)]^2$, and because *Q* is positive we maximize f(x) by making $[x - (P/2Q)]^2$ as small as possible. Therefore, we set x = P/2Q, which gives us a unique global maximum. It is unique because any *x* different from P/2Q will cause $[x - (P/2Q)]^2$ to be positive.

Formula for maximizaing a quadratic

If $f(x) = Px - Qx^2 + R$ and Q > 0, then for $x^* = P/2Q$ we have $f(x^*) > f(x)$ for all $x \neq x^*$.

Note that f(x) does not have a maximum if $f(x) = Px - Qx^2 + R$ and Q < 0. That's because we can write

$$f(x) = R + \frac{P^2}{4Q} - Q\left(x - \frac{P}{2Q}\right)^2$$
,



Figure 2.1

and when—*Q* is positive we can make $-Q[x - (P/2Q)]^2$ arbitrarily large by making *x* sufficiently large.

1.2 Constrained maximization

In this subsection we want to maximize $f(x) = Px - Qx^2 + R$ subject to the constraint $a \le x \le b$. Restrictions of the form $a \le x \le b$ arise naturally in consumer choice because we cannot have x < 0 nor can we have y < 0. (The budget constraint $p_1x + p_2y = \theta$ implies y < 0 when expenditure on commodity *X* exceeds income—i.e., when $p_1x > \theta$, which is equivalent to $x > \theta/p_1$. Therefore, when we find the global maximizing value x^* we have to check to make sure that the inequality $0 \le x^* \le \theta/p_1$ is satisfied at x^* .)

Because $f(x) = R + (P^2/4Q^2) - Q[x - (P/2Q)]^2$, when Q > 0 we see that, using x = P/2Q as the starting point, f(x) decreases as we increase x. That's because $Q[x - (P/2Q)]^2$ is zero when x = P/2Q and $Q[x - (P/2Q)]^2$ increases when x increases through values greater than P/2Q. And, again using x = P/2Qas the starting point, f(x) decreases as we decrease x because $Q[x - (P/2Q)]^2$ is zero when x = P/2Q and $Q[x - (P/2Q)]^2$ increases when x decreases through values less than P/2Q. We have established that the graph of f is hill shaped, with the peak occurring when x = P/2Q (Figure 2.1, with $x_4 = P/2Q$).

We have just learned that f(x) is increasing (the graph is uphill) to the left of x = P/2Q, and f(x) is decreasing (the graph is downhill) to the right of x = P/2Q. If P/2Q > b then the solution of the problem

maximize
$$f(x) = Px - Qx^2 + R$$
 subject to $a \le x \le b$

must be x = b. That follows from the fact that P/2Q > b implies that f(x) increases when x < b and x increases (Figure 2.1, with $x_2 = b$). However, if P/2Q < a then the solution to the constrained maximization problem is x = a because a > P/2Q implies that f(x) increases when x > a and x decreases (see Figure 2.1, with $x_6 = a$).

Formulas for constrained maximization of a quadratic

The given function is $f(x) = Px - Qx^2 + R$ and Q > 0. And *a* and *b* are given numbers, with a < b.

If $a \le P/2Q \le b$, then $x^* = P/2Q$ maximizes f(x) subject to $a \le x \le b$. If P/2Q > b, then $x^* = b$ maximizes f(x) subject to $a \le x \le b$. If P/2Q < a then $x^* = a$ maximizes f(x) subject to $a \le x \le b$.

∂2 OVERVIEW OF CALCULUS

This section establishes the first-order conditions for maximization of a function of one real variable, with and without constraints. The derivation is selfcontained, but some of the applications in this book assume that you know more than is presented in Section 2.1. For instance, it is taken for granted that you know the power rule: the derivative of $f(x) = x^n$ is nx^{n-1} . Also, the chain rule is used on occasion. Nevertheless, the basic theory is developed rigorously in Section 2.1 because many readers will benefit from a refresher course, particularly in view of the fact that we highlight the intuition underlying the use of calculus. We use f'(x) to denote the first derivative of f at x, although on occasion df/dx or even dy/dx, with y = f(x), make an appearance.

Consider the standard consumer choice problem:

maximize U(x, y) subject to $p_1x + p_2y = \theta$.

U is the utility function, and utility depends on the amounts *x* and *y* of the two goods consumed. The prices of goods *X* and *Y* are p_1 and p_2 , respectively, and θ is the individual's income. The budget constraint $p_1x + p_2y = \theta$ can be solved for *y* as a function of *x*:

$$y = \frac{\theta}{p_2} - \frac{p_1 x}{p_2}$$

Now, substitute this value of y into the utility function. We want to maximize

$$V(x) = U\left(x, \frac{\theta}{p_2} - \frac{p_1 x}{p_2}\right)$$

and *V* is a function of only one variable, *x*, because θ , p_1 , and p_2 are constants they are outside of the control of the consumer at the time the consumption decision is made. This means that we can apply elementary calculus to the problem and maximize V(x). We no longer have to worry about the budget constraint because that is built into *V*. With one stroke we have eliminated one variable and the budget constraint as well. Once we have obtained the number x^* that maximizes *V*, we simply use the budget constraint to solve for *y*.

2.1 Unconstrained maximization

Let f(x) represent the function to be maximized, with x a real variable. This means that x can be any real number, and for any choice of x the function f

specifies another real number f(x). Initially, we assume that there is no constraint of any kind on the range of values that x can assume. We'll show that $f'(x^*) = 0$ must hold if f is maximized at x^* .

Now let x^* represent any real number that maximizes the function f. Formally, this means that $f(x^*) \ge f(x)$ holds for every real number x. Another way of saying this is $f(x^*) \ge f(x^* + \epsilon)$ for every real number ϵ . (Just replace x by $x^* + \epsilon$, defining ϵ as the quantity $x - x^*$.) We can think of ϵ as an increment, positive or negative, taking us away from x^* . Because f is maximized at x^* , this increment, or step, cannot increase the value of f. More formally, we write

$$f(x^* + \epsilon) - f(x^*) \le 0 \quad \text{for all } \epsilon.$$
^[1]

Condition [1] is just another way of saying that f is maximized at x^* . This is pretty obvious, but we only need to pursue this a little further to get a striking and useful result.

Multiplying an inequality by a constant

Let α , β , ϵ , and ϕ be real numbers. If $\epsilon > 0$ and $\phi > 0$ then $\epsilon \times \phi > 0$. Therefore, if $\alpha > \beta$ then $\alpha - \beta > 0$, and thus $\epsilon \times (\alpha - \beta) > 0$ if $\epsilon > 0$. This implies that if $\alpha > \beta$ and $\epsilon > 0$ we have $\epsilon \alpha > \epsilon \beta$. It follows that if $\epsilon < 0$ and $\alpha > \beta$ then $-\epsilon \alpha > -\epsilon \beta$, which in turn implies $\epsilon \alpha < \epsilon \beta$. Finally, if $\epsilon > 0$ and $\alpha \ge \beta$ then $\epsilon \alpha \ge \epsilon \beta$, and if $\epsilon < 0$ and $\alpha \ge \beta$ then $\epsilon \alpha \le \epsilon \beta$.

If ϵ is positive (strictly greater than zero) then $f(x^* + \epsilon) - f(x^*)$ will still be less than or equal to zero after we divide that expression by ϵ . We state this formally as Condition [2]:

$$\frac{f(x^* + \epsilon) - f(x^*)}{\epsilon} \le 0 \quad \text{for all } \epsilon > 0.$$
^[2]

As ϵ approaches zero through positive values, the limit must also be less than or equal to zero as a consequence of Condition [2]. For future reference, we state this as Condition [3]:

The limit of
$$\frac{f(x^* + \epsilon) - f(x^*)}{\epsilon}$$
 is ≤ 0 as $\epsilon > 0$ approaches 0. [3]

Similarly, if we divide $f(x^* + \epsilon) - f(x^*)$ by any $\epsilon < 0$ the inequality sign will change direction, and so we have Condition [4]:

$$\frac{f(x^* + \epsilon) - f(x^*)}{\epsilon} \ge 0 \quad \text{for all } \epsilon < 0.$$
^[4]

As ϵ approaches zero through *negative* numbers the limit must be nonnegative because each term is nonnegative by Condition [4]. This is represented as Condition [5]:

The limit of
$$\frac{f(x^* + \epsilon) - f(x^*)}{\epsilon}$$
 is ≥ 0 as $\epsilon < 0$ approaches 0. [5]

78 Basic Models and Tools

If *f* has a derivative at *x* then, by definition, the limit of $[f(x^* + \epsilon) - f(x^*)]/\epsilon$ must be the same when ϵ approaches zero through positive values as it is when ϵ approaches zero through negative values. But then [3] and [5] can both be satisfied only if the limit is zero in both cases. In short, $f'(x^*) = 0$ is a necessary condition for *f* to have a maximum at x^* . The function *f* in Figure 2.1 is maximized at $x^* = x_4$. We see that the first derivative of *f* is zero at x_4 because the graph of *f* is perfectly horizontal at x_4 .

Necessary condition for an unconstrained maximum

If $f(x^*) \ge f(x)$ for all real numbers *x* then $f'(x^*) = 0$.

Here is an alternative derivation of the fact that $f'(x^*) = 0$ if f is maximized at x^* . (You don't need to master both treatments; just adopt the one with which you are more comfortable.) Suppose that f'(x) > 0. We show that f cannot have a maximum at x. Let δ represent f'(x). We have $\delta > 0$. Intuitively, a small move to the right will increase the value of f. It may have to be a very small move if x is close to the top of the hill, as is the case with $x = x_2$ in Figure 2.1. A move to x_7 will lower the value of f, but a sufficiently small move to the right, such as the one taking us from x_2 to x_3 , will increase f.

Here is the formal argument: The limit of $[f(x + \epsilon) - f(x)]/\epsilon$ is δ , and we assume that $\delta > 0$. For $\epsilon > 0$ sufficiently close to zero we can get $[f(x + \epsilon) - f(x)]/\epsilon$ close enough to δ to guarantee that the ratio is greater than $\frac{1}{2} \delta$. But then

$$f(x+\epsilon) - f(x) > \epsilon \times \frac{1}{2} \delta > 0.$$

This means that $f(x + \epsilon) - f(x) > 0$, or $f(x + \epsilon) > f(x)$. Then *x* does not yield the maximum value of *f*, because $f(x + \epsilon)$ is larger than f(x).

Next we show that f cannot have a maximum at x if f'(x) < 0. Let δ again represent f'(x), with $\delta < 0$ this time. Intuitively, a small move to the *left* increases the value of f. It may have to be a very small move as in the case $x = x_6$ in Figure 2.1. A move to x_1 lowers f(x), but a sufficiently small move to the left, such as the one taking us from x_6 to x_5 , will increase f(x). Consider: Because the limit of $[f(x + \epsilon) - f(x)]/\epsilon$ is δ , for $\epsilon < 0$ sufficiently close to zero we can get $[f(x + \epsilon) - f(x)]/\epsilon$ close enough to δ to guarantee that ratio is algebraically smaller than $\frac{1}{2}\delta$. Therefore, $[f(x + \epsilon) - f(x)]/\epsilon < \frac{1}{2}\delta$ for ϵ sufficiently close to zero we can get on the sign, yielding

$$f(x+\epsilon) - f(x) > \epsilon \times \frac{1}{2} \delta > 0.$$

(We have $\epsilon \times \frac{1}{2} \delta > 0$ because both ϵ and δ are negative.) But then $f(x + \epsilon) - f(x) > 0$, or $f(x + \epsilon) > f(x)$. Then *x* does not yield the maximum value of *f*, because $f(x + \epsilon)$ is larger than f(x). Therefore, if *f* is maximized at *x* we can rule out both f'(x) > 0 and f'(x) < 0.

Example 2.1: Maximizing the function $f(x) = 10x - x^2 - 25$

We want to find the point at which f is maximized. Note that $f(x) = -(x - 5)^2$, which can never be positive. When x = 5 the value of the function is zero, so that is the point at which f reaches a maximum. Every other value of x will yield f(x) < 0. So we don't need calculus in this case. But let's see how calculus brings us to the same conclusion. We need to calculate the first derivative of f.

$$f(x+\epsilon) = 10(x+\epsilon) - (x+\epsilon)^2 - 25$$
$$= 10x + 10\epsilon - x^2 - 2\epsilon x - \epsilon^2 - 25$$

Therefore,

$$f(x + \epsilon) - f(x) = 10\epsilon - 2\epsilon x - \epsilon^2$$

and hence

$$\frac{f(x+\epsilon) - f(x)}{\epsilon} = \frac{10\epsilon - 2\epsilon x - \epsilon^2}{\epsilon} = 10 - 2x - \epsilon.$$

Clearly, $10 - 2x - \epsilon$ approaches 10 - 2x as ϵ approaches zero. Therefore, f'(x) = 10 - 2x, the first derivative of f. Now, we said that f'(x) = 0 is necessary for a maximum. Set f'(x) = 0 and solve for x: We get 10 - 2x = 0, and thus x = 5.

Next we look at consumer choice.

Example 2.2: A simple consumer choice problem

We want to maximize U(x, y) = xy subject to the budget constraint 5x + 2y = 1000. The utility of a basket with *x* units of commodity *X* and *y* units of commodity *Y* is the product of the two numbers *x* and *y*. (It may help at this point to draw a typical indifference curve; say, the set of baskets that yield a utility of 12.) The price of good *X* is 5 and the price of good *Y* is 2. Income is 1000. Solving the budget constraint for *y* yields

$$y = \frac{1000}{2} - \frac{5x}{2} = 500 - 2.5x.$$

Now, substitute this value of *y* into the utility function. We want to maximize

$$V(x) = x(500 - 2.5x).$$

 $V(x) = 500x - 2.5x^2$, and thus $V'(x) = 500 - 5x.$

Then V'(x) = 0 yields 500 - 5x = 0, and thus $x^* = 100$. There is only one value of *x* that gives V' = 0. Therefore, there can be only one utility-maximizing value of *x*, namely x = 100. Now we can use the budget constraint to solve for *y*:

y = 500 - 2.5x = 500 - 2.5(100) = 500 - 250 = 250.

Therefore, the chosen basket has x = 100 and y = 250.

We can use the technique of Example 2.2 to solve for the *demand functions*. All we have to do is represent prices and income symbolically, but treat them as numbers.

Example 2.3: Deriving a demand function

U(x, y) = xy, which we maximize, subject to the budget constraint $p_1x + p_2y = \theta$, where prices and income are parameters. We will solve for the demands *x* and *y* as a function of prices and income. Then $y = \theta/p_2 - p_1x/p_2$, and we substitute this into the utility function:

$$V(x) = x \times \left(\frac{\theta}{p_2} - \frac{p_1 x}{p_2}\right).$$

We have $V = \theta x/p_2 - p_1 x^2/p_2$ and thus $V'(x) = \theta/p_2 - 2p_1 x/p_2$. When $\theta/p_2 - 2p_1 x/p_2 = 0$ we have $x = \theta/2p_1$. This is the only value of x that gives V' = 0, so the consumer choice problem has a unique solution: $x = \theta/2p_1$. From the budget constraint, $y = \theta/p_2 - p_1 x/p_2$ and if in addition $x = \theta/2p_1$ we must have $y = \theta/2p_2$. The expressions $x = \theta/2p_1$ and $y = \theta/2p_2$ are the demand functions for commodities X and Y respectively.

If we are given particular values for prices and income we can plug them into the demand functions to get the *amounts* demanded at that price and income regime. (Verify that x = 100 and y = 250 when $p_1 = 5$, $p_2 = 2$, and $\theta = 1000$.) Note that we have V''(x) < 0 for all x when V is derived from the utility function U = xy by solving the budget constraint for y and substituting.

2.2 Constrained maximization

Suppose that we want to maximize f subject to the restriction $a \le x \le b$. It is vital that you pay attention to the difference between a < x and $a \le x$ and similarly to the distinction between x < b and $x \le b$. Our first observation is that if we actually have $a < x^* < b$ at the point x^* where the constrained maximum is achieved, then $f'(x^*) = 0$ is still a necessary condition for f to be maximized at x^* . The proof of that fact is actually embedded in the discussion of the unconstrained case. If f'(x) > 0 then for $\epsilon > 0$ sufficiently close to zero we will have $[f(x + \epsilon) - f(x)]/\epsilon > 0$ and hence $f(x + \epsilon) > f(x)$. Review Section 2.1 to confirm that we will have $[f(x + \epsilon) - f(x)]/\epsilon > 0$ if we make $\epsilon > 0$ smaller still. Therefore, if f'(x) > 0 and x < b then we can find $\epsilon > 0$ small enough so that we get both $x + \epsilon < b$ and $f(x + \epsilon) > f(x)$. And we will certainly have $x + \epsilon \ge a$ if $x \ge a$. Therefore, if $a \le x < b$ and f'(x) > 0 the function x cannot be maximized at x even if we are not allowed to consider values of x larger than bor smaller than a. (In Figure 2.1, f is not maximized at x_2 , even if the constraint $x_1 \le x \le x_7$ must be observed.)

Similarly, we can show that *f* cannot be maximized at *x* if f'(x) < 0 and $a < x \le b$, even if we are not allowed to go below *a* or above *b*. (In Figure 2.1, *f* is not maximized at x_6 , even with the restriction $x_1 \le x \le x_7$.) We have proved the following: If x^* maximizes *f* subject to the constraint $a \le x \le b$ and $a < x^* < b$

actually holds, then we can't have $f'(x^*) > 0$ and we can't have $f'(x^*) < 0$. This means that $f'(x^*) = 0$ must hold. Be careful! There is nothing to guarantee that $a < x^* < b$ will actually hold at the solution value x^* . (Try maximizing f(x) = 2x subject to $0 \le x \le 100$. Clearly, the solution is $x^* = 100$, but f'(100) = 2 because f' is constant at 2.) But *if* $a < x^* < b$ *does* hold at the solution point then we must have $f'(x^*) = 0$. The function f is maximized at $x = x_4$ in Figure 2.1, with or without the constraint $x_1 \le x \le x_7$. The first derivative of f is zero at x_4 .

Now, what if we do have $x^* = a$ or $x^* = b$ at the point x^* where f is maximized, subject to the constraint $a \le x \le b$? Calculus is still a big help here, but you have to know how to use it. In general, calculus is not a formula for cranking out an answer to a problem but rather a useful device for finding the solution.

We know that if *f* is maximized at x^* and $a < x^* < b$, then $f'(x^*)$ must equal zero. Therefore, if we want to maximize *f* subject to the constraints $a \le x \le b$ then either *f* will achieve its maximum at a point where its first derivative is zero or else the solution value of *x* will be *a* or at *b*.

First-order conditions for constrained maximization

If f(x) is maximized at x^* subject to the constraints $a \le x \le b$ then either $f'(x^*) = 0$ or $x^* = a$ or $x^* = b$.

This means that in solving the constrained maximization problem we can confine our attention to a limited number of values of *x*: points where the first derivative is zero, and the values x = a and x = b. Then we can compute f(x) at these points to see which gives the highest value of *x*.

Example 2.4: Consumer choice with nonnegative consumption

Let U(x, y) = (x + 5)(y + 2), which we want to maximize subject to $Px + y = \theta$, $x \ge 0$, and $y \ge 0$. (We fix p_2 at 1 and set $P = p_1$ to simplify computation.) We have $y = \theta - Px$ from the budget equation, and substituting this into the utility function yields

$$V(x) = (x+5)(\theta - Px + 2) = (\theta + 2 - 5P)x - Px^2 + 5\theta + 10.$$

 $V'(x) = \theta + 2 - 5P - 2Px$. Then V''(x) = -2P, which is always negative. Therefore, the graph of the function *V* is hill shaped, and if V'(x) = 0 yields a unique value of *x* satisfying $0 \le x \le \theta/P$ this will be the demand for *x*. Of course, $\theta + 2 - 5P - 2Px = 0$ implies $x = [\frac{1}{2}(\theta + 2 - 5P)]/P$. If P = 1 and $\theta = 100$ then x = 97/2 = 48.5, which certainly satisfies $0 \le x \le 100$. (How much *Y* will be demanded in that case?) If P = 25 and $\theta = 100$ then V'(x) = 0 implies x = -0.46, which is inadmissible. The consumer will either set x = 0 or x = 100/25. We have $V(0) = 5 \times (100 + 2) = 510$, and $V(4) = (4 + 5) \times (100 - 100 + 2) = 18$. Therefore, the consumer will demand x = 0 units of *X* and 100 units of *Y* when $p_1 = 25$, $p_2 = 1$, and $\theta = 100$.

82 Basic Models and Tools

In solving the consumer choice problem in Example 2.2 we ignored the constraint $0 \le x \le 200$ that is required to ensure that neither *x* nor *y* is negative. Here's why we were able to do that: Note that utility is zero when *x* or *y* is zero. That is, U = xy = 0 if x = 0 or y = 0. Even a tiny amount of money spent on each good will yield a positive product *xy*, so we know that the consumer can do better than zero utility. Therefore, the utility-maximizing basket will have x > 0 and y > 0. But if *y* is positive we can't have x = 200; we must have x < 200. Therefore, the solution value x^* will have to satisfy $0 < x^* < 200$. We know that in this case we must have $V'(x^*) = 0$. We saw that only one value of *x* gives V' = 0. The solution to the consumer choice problem *must* have V' = 0. The same argument works for the derivation of the demand functions in Example 2.3. For many other utility functions that you will encounter, corner (or boundary) solutions can be ruled out a priori.

Example 2.5: Corner points need not apply

Maximize $U(x, y) = x^2 y$ subject to 5x + 2y = 60. From the budget equation we have y = 30 - 5x/2 = 30 - 2.5x. Substitute this for *y* in the utility function to obtain

$$V = x^2(30 - 2.5x) = 30x^2 - 2.5x^3.$$

We want to maximize this function of *x* subject to $0 \le x \le 60/5$. $V'(x) = 60x - 7.5x^2$ and V''(x) = 60 - 15x. Setting the first derivative equal to zero yields

$$60 - 7.5x = 0$$
 or $x = 8$.

Because U = 0 if x = 0 or y = 0, utility will be maximized at a point where x is strictly greater than 0 and strictly less than 12. Therefore, V'(x) = 0 at the solution to the consumer choice problem. Therefore, x = 8 is the optimal value of x, and the budget constraint yields y = 10.

2.3 Strictly concave functions

We now confine our attention to a special class of functions that arises most of the time in economics. We consider only functions for which the *second derivative* f''(x) is negative at all values of *x*. Such functions are called *strictly concave*.

DEFINITION: Strictly concave function

The function *f* of a single variable *x* is strictly concave if f''(x) < 0 for all *x*.

By definition, f'' is the derivative of the derivative. For $f(x) = 10x - x^2 - 25$ we have f'(x) = 10 - 2x and hence f''(x) = -2, which is negative. The second derivative tells us how the first derivative changes as *x* changes. If we always have f''(x) < 0 then f'(x) gets smaller (algebraically) as *x* increases. This has

two important implications. First, if $f'(x^*) = 0$ then there is no other value of x for which f' is zero: To the right of x^* the first derivative is negative. Why? Because f'(x) falls as x increases and it is zero at x^* , so f'(x) < 0 for all $x > x^*$. Therefore, we cannot find any $x > x^*$ for which f'(x) = 0. Now, consider $x < x^*$. The first derivative decreases as we move to the right, so it increases as we move to the left. If f' is zero at x^* and it increases as we move to the left then f'(x) is positive for all $x < x^*$. Therefore, we cannot have f'(x) = 0 for any $x < x^*$. In short, if f'' is negative at all points then there is at most one value of x for which f' is zero.

Here is the second important consequence of the fact that f'' < 0 at all points: If $f'(x^*) = 0$ we know that f' is positive to the left of x^* and f' is negative to the right of x^* . When f' is positive the value of the function f itself is increasing. We know that because f'(x) > 0 is just another way of saying that f is increasing at x. To the right of x^* we have f'(x) < 0 and hence the value of f falls as x increases beyond x^* . This is a consequence of the fact that f' is negative to the right of x^* , and f'(x) < 0 is just another way of saying that f is falling at x. (In Figure 2.1, $x^* = x_4$ and f'(x) is strictly positive to the left of x_4 and f'(x) < 0 to the right of x_4 .)

Now, let's summarize: Suppose f'' is negative everywhere and $f'(x^*) = 0$. Then f' is not equal to zero for any other value of x. Moreover, f falls as we move to the right of x^* and f rises as we move *toward* x^* from the left. This means that the graph of f is a hill with the peak at x^* , as in Figure 2.1 with $x^* = x_4$. In other words, f has a unique maximum at x^* .

Global maximization with a negative second derivative

If f''(x) < 0 for all x and $f'(x^*) = 0$ then $f(x^*) > f(x)$ for all $x \neq x^*$. In other words, f has a unique global maximum at x^* if $f'(x^*) = 0$ and f''(x) < 0 for all x.

Suppose, however, that we are restricted to the region $a \le x \le b$. If we find some x^* in that interval such that $f'(x^*) = 0$ then we are sure that is the unique solution to our problem. Why? Because $f'(x^*) = 0$ implies that $f(x^*) > f(x)$ for *all* other x, and therefore we certainly have $f(x^*) > f(x)$ for all $x \ne x^*$ satisfying $a \le x \le b$. (Caveat: This all depends on f'' < 0 holding everywhere.) Suppose, however, that the value of x for which f' is zero is outside of the interval $a \le x \le b$.

Consider first the case $f'(x^*) = 0$ and $x^* > b$. We know that f is rising to the left of x^* . Therefore, f is increasing at all x in the constraint region, because $a \le x \le b$ implies that x is to the left of x^* . Therefore, $f'(x^*) = 0$ and $x^* > b$ implies that x = b is our solution: f(b) > f(x) for all x satisfying $a \le x < b$, as illustrated in Figure 2.1 with x_1 representing a and x_2 representing b (and $x^* = x_4$). Now suppose that $f'(x^*) = 0$ and $x^* < a$. Because f'' < 0 at every point, f is falling

84 Basic Models and Tools

to the right of x^* . Therefore f is decreasing at all x in the constraint region, because $a \le x \le b$ implies that x is to the right of x^* . Therefore, $f'(x^*) = 0$ and $x^* < a$ implies that x = a is our solution: f(a) > f(x) for all x satisfying $a < x \le b$, as illustrated in Figure 2.1 with $a = x_6$ and $b = x_7$ (and $x^* = x_4$).

You will probably have encountered other techniques for generating consumer demand. For your peace of mind, we will apply each of them to the problem

maximize $U(x, y) = x^2 y$ subject to 5x + 2y = 60

to confirm that they yield the same solution. The first (in Section 2.5) is expressed in terms of the tangency of the indifference curve through the chosen bundle to the budget line.

2.4 Minimization

If x^* minimizes f(x) over all real numbers x, then by definition $f(x^*) \le f(x)$ for all x. It follows that $-f(x^*) \ge -f(x)$ for all x. In other words, if x^* minimizes f then x^* maximizes -f. It follows that the derivative of -f is zero at x^* . But for any function f, the derivative of -f is the negative of the derivative of f. Therefore, $f'(x^*) = 0$ if f is minimized at x^* . Similarly, $f'(x^*) = 0$ if $a < x^* < b$ and x^* is the solution to problem

minimize f(x) subject to $a \le x \le b$.

Finally, we say that f is *strictly convex* if its second derivative is positive at every point. But if f''(x) > 0 for all x then the function -f has a negative second derivative at every point. In that case, $-f'(x^*) = 0$ implies that $-f(x^*) > -f(x)$ for all $x \neq x^*$. It follows that $f(x^*) < f(x)$ for all $x \neq x^*$. Because $-f'(x^*) = 0$ implies $f'(x^*) = 0$, we have demonstrated that for any strictly convex function f, if $f'(x^*) = 0$ then f has a unique global minimum at x^* .

Conditions for minimization

If $f(x^*) \le f(x)$ for every real number x then $f'(x^*) = 0$. If f(x) is minimized at x^* subject to the constraints $a \le x \le b$, then either $f'(x^*) = 0$ or $x^* = a$ or $x^* = b$. If f is strictly convex and $f'(x^*) = 0$ then f has a unique global minimum at x^* . If f is strictly convex and we want to minimize f subject to $a \le x \le b$, then f'(x) = 0 and x > b implies that the solution is x = b, and the solution is x = a if f'(x) = 0 implies x < a.

2.5 The tangency approach

Recall that the marginal rate of substitution (MRS) at a point (x^0, y^0) is the absolute value of the slope of the indifference curve through (x^0, y^0) . Let *C* be that indifference curve. *C* is defined by $U(x, y) = c^0$, where c^0 is the constant $U(x^0, y^0)$, and it implicitly gives us *y* as a function of *x*. For many utility functions we can explicitly solve for *y* as a function of *x*. For other cases, we use the implicit

function theorem: the derivative dy/dx of the implicit function is the negative of the ratio of the partial derivatives of *U*.

Example 2.6: Tangency and consumer choice

We reconsider Example 2.5: Maximize $U(x, y) = x^2y$ subject to 5x + 2y = 60. Notice that utility is zero if either x = 0 or y = 0, so the solution will have x > 0 and y > 0. In that case, the economic argument based on indifference curves reveals that the MRS equals the price ratio at the chosen consumption plan. To derive the MRS we set utility equal to a constant ℓ . That gives us the equation of a generic indifference curve: $x^2y = \ell$ in the present case. We solve for y to get $y = \ell x^{-2}$. Then $dy/dx = -2\ell x^{-3}$. Now, $\ell = x^2y$. Therefore,

$$\frac{dy}{dx} = -2x^2yx^{-3} = -\frac{2y}{x}$$

The MRS is the negative of the slope of the indifference curve, and hence if $U = x^2y$ the MRS at the generic bundle (x, y) is 2y/x. The price ratio is 5/2. Therefore, the solution will satisfy 2y/x = 5/2, which implies 4y = 5x. This equation does not have a unique solution, nor should we expect one. We can't pin down the choice without the budget equation. Substituting 4y for 5x in the budget equation yields 4y + 2y = 60, which yields y = 10. Then 4y = 40 = 5x and hence x = 8. (Verify that x = 8 and y = 10 satisfies the budget equation and equates MRS and the price ratio.)

2.6 The total derivative and the chain rule

If *f* is a function of *x* and *y*, and *y* itself is a function of *x*, say y = g(x), then the chain rule gives us df/dx in terms of the partial derivatives of *f* and the derivative of *g*. Specifically

$$\frac{df}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \times \frac{dy}{dx}$$
$$= \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y} \times g'(x)$$

In words, the rate of change in f with respect to x is the rate of change of f with respect to x when y is held constant, *plus* the rate of change of f with respect to y with x held constant multiplied by the rate of change of y per unit change in x, determined by the function g. It is easy to grasp the idea by looking at linear functions.

Example 2.7: The chain rule with linear functions

Suppose f(x, y) = 2x + 5y, and y = 3x. Of course, $\partial f/\partial x = 2$ and $\partial f/\partial y = 5$, with g'(x) = 3. According to the chain rule $df/dx = 2 + 5 \times 3 = 17$. We can confirm this by substituting 3x for y in f. We get f = 2x + 5(3x) = 2x + 15x = 17x. Clearly, df/dx = 17.

Example 2.8: The chain rule and consumer choice

Maximize $U(x, y) = x^2 y$ subject to 5x + 2y = 60. We have y = 30 - 2.5x from the budget constraint, and thus dy/dx = -2.5. Because U(x, y) depends on x and y we let U_x denote the partial derivative of U with respect to x and let U_y denote the partial derivative of U with respect to y. If we think of y as a function of x, then the total derivative of U with respect to x is

$$\frac{dU}{dx} = U_x + U_y \times \frac{dy}{dx}.$$

We have $U_x = 2xy$ and $U_y = x^2$. Therefore, $dU/dx = 2xy - 2.5x^2$. Now, set this equal to zero to find a maximum:

$$2xy - 2.5x^2 = 0.$$

Dividing through by x (do we have to worry about dividing by zero?) yields

$$2y - 2.5x = 0$$
 or $4y = 5x$.

Substituting 4*y* for 5*x* in the budget equation yields y = 10, and thus x = 8.

Sources

The material in this section is very standard and is the subject of hundreds of mathematics books, including Strang (1991), Dozens more texts have been written by and for economists, featuring economic applications, including Novshek (1993) and Binmore and Davies (2001).

Problem set

- **1.** Solve for the demand functions of a consumer whose preferences can be represented by the utility function $U(x, y) = x^{\alpha} y^{\beta}$, where α and β are positive constants.
- **2.** Solve for the demand functions of a consumer whose preferences can be represented by the utility function U(x, y) = (x + 1)y.
- **3.** Solve for the demand functions of a consumer whose preferences can be represented by the utility function $U(x, y) = \sqrt{x} + y$.

3 LAGRANGIAN MULTIPLIERS

In Section 2 we were able to solve constrained maximization problems involving two variables *x* and *y* because there was only one constraint, and that could be solved to express *y* as a function of *x*. The resulting function was then substituted for *y* in the function being maximized, leaving us with an unconstrained one-variable problem. That technique suffices for all of the applications in this book. However, if you want to know more about the use of prices—that is, Lagrangian multipliers—in solving constrained maximization problems you will benefit from this section. The Lagrangian technique requires verification of a constraint qualification, but we will not address that issue because our examples will use

3. Lagrangian Multipliers

functions for which the qualification is met. Moreover, that is also the case with virtually all economic applications.

∂ 3.1 The Lagrangian multiplier with a single resource constraint

We return to Example 2.5, this time using the Lagrangian technique to obtain the solution.

Example 3.1: The Lagrangian approach to consumer choice

We want to maximize $U(x, y) = x^2 y$ subject to 5x + 2y = 60. Instead we maximize

$$\mathcal{L} = x^2 y - \lambda(5x + 2y - 60).$$

Let \mathcal{L}_x denote the partial of \mathcal{L} with respect to x and let \mathcal{L}_y denote the partial of \mathcal{L} with respect to y. Setting the first partials equal to zero yields

$$\mathcal{L}_x = 2xy - 5\lambda = 0$$
 and $\mathcal{L}_y = x^2 - 2\lambda = 0$

The first equation yields $\lambda = 2xy/5$ and substituting this value of λ into the second equation leads to

$$x^2 - 2 \times \frac{2xy}{5} = 0,$$

or 5x = 4y after dividing both sides by *x*. (We know that *x* will not be zero at the chosen consumption plan.) Of course 5x = 4y along with the budget equation yields x = 8 and y = 10.

Now, substitute the solution values of *x* and *y* into the equation $\mathcal{L}_x = 0$ to solve for λ . We get 2(8)(10) – 5 λ = 0, and hence λ^* = 32 at the solution point (*x*^{*}, *y*^{*}). To interpret λ^* , write the budget equation with income *m* as a variable:

$$5x + 2y = m$$

Solve once more for the consumer's chosen basket. We still get 5x = 4y whatever solution technique is employed. Substituting into the budget equation yields 4y + 2y = m, or y = m/6. Because 4y = 5x we have 5x = 4m/6, or x = 4m/30. This gives us the demands as a function of income: $x^* = 4m/30$ and $y^* = m/6$, given $p_1 = 5$ and $p_2 = 2$. Now, substitute these demands into the utility function. We get

$$U = \left(\frac{4m}{30}\right)^2 \times \frac{m}{6} = \frac{16m^3}{5400}.$$

Then $dU/dm = (16 \times 3m^2)/5400 = 8m^2/900$. When m = 60 this yields $dU/dm = 32 = \lambda^*$. This is a general phenomenon. The Lagrangian multiplier always gives the increase in utility per unit increase in income.

∂ **3.2** Remark on planning and Lagrangians

The function of price in a market system is in part to signal marginal values to producers and consumers. Example 3.1 illustrates the fact that Lagrangian

88 Basic Models and Tools

multipliers are also marginal values. If we maximize U(x, y) subject to the constraint px + qy = a, the value of U at the solution point (x^*, y^*) will, of course, be a function of a. If p and q are positive constants, the larger a is the larger Uwill be. Specifically, $dU/da = \lambda^*$ if λ is the multiplier associated with the constraint. That means that the Lagrangian can be interpreted as a price. This will be true even if the constraint is nonlinear. Therefore, prices are intrinsic to the solution of constrained maximization problems, even in the case of problems that appear to have nothing to do with economics. For optimization problems that arise from economic considerations, the fact that Lagrangians are marginal values is of great significance.

Suppose that U(x, y) is an economic planner's objective function, representing the social value of output in the economy, and the equation px + qy = arepresents a resource constraint on the capacity of the economy to produce xand y. Then the solution value λ^* of the Lagrangian multiplier for the constraint is the value of the scarce resource at the margin. If Δa additional units of the resource were obtained, and the maximization problem was solved again, the value of U would increase by $\lambda^* \Delta a$. Therefore, even if the planner has no intention of deferring to the market system, prices are embedded in the mathematical logic of constrained maximization. They can be used to guide the system to the socially optimal menu of goods and services—that is, the one that maximizes Usubject to resource and technology constraints. Moreover, when prices are used to guide decision making, it is far easier to design incentives to get producers and consumers to do their part in executing the optimal menu.

When there are many variables and many constraints the Lagrangian technique is by far the most efficient. And, as Example 3.1 demonstrates, once the planners start using Lagrangians they are using prices. The Lagrangian is the marginal value of an additional unit of the scarce resource that gives rise to the constraint, as we explain in greater depth in the next section.

∂ 3.3 Lagrangian multipliers with more than one resource constraint

Consider the problem

maximize f(x, y) subject to $g(x, y) \le a$ and $h(x, y) \le b$.

We will not consider functions that depend on more than the two variables *x* and *y*, nor will we have more than the two constraints *g* and *h*. The two-variable, two-constraint case will provide sufficient insight.

The function f represents the goal or objective, and we want to pick the values of x and y that maximize f. But constraints g and h restrict the values of x and y that can be selected. For instance, f might refer to the value to society of the plan (x, y) with g and h reflecting resource utilization by the plan of two inputs A and B—labor and capital, say. Then a and b denote the total amounts available of A and B, respectively. The plan (x, y) uses g(x, y) units of labor, and that cannot exceed the total amount of labor, a, in the economy. Similarly, the plan (x, y) uses h(x, y) units of capital, and the economy has only b units of capital. In another application, f(x, y) denotes a firm's profit from the production of x

3. Lagrangian Multipliers

units of commodity *X* and *y* units of commodity *Y*. The constraints represent limitations such as warehouse and transportation capacity.

DEFINITION: Resource utilization

The plan (x, y) requires g(x, y) units of resource *A* as input and h(x, y) units of resource *B*.

The solution of our constrained maximization problem can be characterized by means of two Lagrangian variables α and β associated with the respective constraints *g* and *h*. If *x*^{*} and *y*^{*} constitute a solution to the problem then there exist $\alpha \ge 0$ and $\beta \ge 0$ such that

$$\frac{\partial f(x^*, y^*)}{\partial x} - \alpha \times \frac{\partial g(x^*, y^*)}{\partial x} - \beta \times \frac{\partial h(x^*, y^*)}{\partial x} = 0$$
 [6]

and

$$\frac{\partial f(x^*, y^*)}{\partial y} - \alpha \times \frac{\partial g(x^*, y^*)}{\partial y} - \beta \times \frac{\partial h(x^*, y^*)}{\partial y} = 0.$$
 [7]

Notice that we arrive at the same two necessary conditions if (x^*, y^*) is the plan that maximizes

$$\mathscr{L} = f(x, y) - \alpha g(x, y) - \beta h(x, y),$$

provided that we treat α and β as given constants. That is, if we take the partial derivatives of \mathscr{L} and equate them to zero we get the first-order conditions [6] and [7]. The interpretation of α and β as prices is not a mere contrivance: Lagrangian variable α is a price in the sense that it is the value of a unit of the resource A underlying constraint g, and similarly for β and resource B. In other words, if additional units of A can be obtained then α is the rate at which f will increase per unit of A added. Therefore, α and β truly are *social cost prices*. (Recall the definition in the introductory section of Chapter 1.) Hence $\mathscr{L}(x, y)$ is the gross value f(x, y) of the plan (x, y) minus the cost of employing the scarce resources.

We need to *prove* that there exists α and β such that [6] and [7] have to hold if (x^* , y^*) is a solution to our original problem. Why can we use prices to characterize the solution to a problem that at the outset may have nothing to do with prices or at least is articulated without any reference to prices? The remainder of this subsection explains, but if you want to make a smaller investment of time you may be satisfied with the following numerical example.

Example 3.2: Linear functions

Maximize f(x, y) = 4x + 7y subject to $x + 3y \le 34$ and $2x + y \le 18$. Figure 2.2 shows that the solution will occur at the plan (x^*, y^*) where the lines x + 3y = 34 and 2x + y = 18 meet. Solving these two equations yields $x^* = 4$ and $y^* = 10$, yielding $4x^* + 7y^* = 86$, the maximum value of *f*. The solution is the plan (4, 10)



Figure 2.2

where the lines x + 3y = 34 and 2x + y = 18 intersect because the slope of the line 4x + 7y = 86 is in between the slopes of the lines x + 3y = 34 and 2x + y = 18. (The absolute values of the slopes of *g*, *f*, and *h* are respectively 1/3 < 4/7 < 2.) For the present problem, [6] and [7] become

$$4 = \alpha + 2\beta$$
 and $7 = 3\alpha + \beta$.

This yields $\alpha = 2$ and $\beta = 1$. Our claim is that if we obtain one more unit of resource *A* and replace the constraint $x + 3y \le 34$ with $x + 3y \le 35$ then the solution value of the objective function will increase by $\alpha = 2$. Let's confirm this. The solution is now x = 3.8 and y = 10.4 where the lines x + 3y = 35 and 2x + y = 18 meet. The value of the objective function *f* is $4 \times 3.8 + 7 \times 10.4 = 88$, an increase of 2 over the solution value of *f* for the original problem. Now let's have *b* increase by 1, with *a* at its original level of 34. Does the maximum value of the objective function increase by $\beta = 1$? The solution to this new constrained optimization problem is x = 4.6 and y = 9.8 at the intersection of the lines x + 3y = 34 and 2x + y = 19. This time, the value of the objective function is $4 \times 4.6 + 7 \times 9.8 = 87$, an increase of 1.



Figure 2.3

Consider the simple linear equation 5x + 2y = 0. It is represented in Figure 2.3 where we see that the vector of coefficients (5, 2) makes a ninety-degree angle with the line generated by those coefficients. We begin by showing that this always holds: The vector (p, q) makes a ninety-degree angle with the line px + qy = 0. Let's look at the specific case 5x + 2y = 0 first. If x = 2 then we must have y = -5 if the point (x, y) is to be on the line. (Just solve $5 \times 2 + 2y = 0$ for *y*.) Then we have a triangle with the three vertices (5, 2), (2, -5), and (0, 0) and with sides *a*, *b*, and *c* as depicted in Figure 2.3. We want to show that angle θ is a right angle, so we need to prove that $a^2 + b^2 = c^2$, the Pythagorean equality. (Section 3.4 shows why θ is a right angle *if* the Pythagorean equality holds.)

$$a^{2} = (5-0)^{2} + (2-0)^{2} = 25 + 4 = 29.$$

 $b^{2} = (2-0)^{2} + (-5-0)^{2} = 4 + 25 = 29.$
 $c^{2} = (2-5)^{2} + (-5-2)^{2} = 9 + 49 = 58.$

Therefore, $a^2 + b^2 = c^2$, and hence θ is a right angle. In general, the point (q, -p) is on the line px + qy = 0 so we have a triangle with the three vertices (p, q),





(q, -p), and (0, 0). Consult Figure 2.3 again:

$$\begin{aligned} a^2 &= (p-0)^2 + (q-0)^2 = p^2 + q^2, \\ b^2 &= (q-0)^2 + (-p-0)^2 = q^2 + p^2, \\ c^2 &= (q-p)^2 + (-p-q)^2 = q^2 - 2qp + p^2 + p^2 + 2pq + q^2 = 2(p^2 + q^2) \end{aligned}$$

Therefore, $a^2 + b^2 = c^2$, and hence θ is a right angle.

Because (p, q) makes a right angle with the line px + qy = 0, if we start at a point (x, y) on the line $px + qy = \Omega$ and move in the direction (p, q) then we are increasing the value of px + qy at the fastest rate, as illustrated by Figure 2.4. The directions A_1 and A_2 do not make right angles with the line (ℓ_0) , and they get us onto the respective level curves ℓ_1 and ℓ_2 , which are below the level curve ℓ_{pq} associated with the direction (p, q). (We have normalized the arrows so that they have the same length, say, unit length.)

Consider the generic constrained optimization problem for linear functions:

maximize $f_1x + f_2y$ subject to $g_1x + g_2y \le a$ and $h_1x + h_2y \le b$.

In this case, f_1 , f_2 , g_1 , g_2 , h_1 , and h_2 are given *constants*. We deal with the family of problems for which the solution occurs at the plan (x^*, y^*) where the lines $g_1x + g_2y = a$ and $h_1x + h_2y = b$ meet (Figure 2.5a). The vector (f_1, f_2) must lie between (g_1, g_2) and (h_1, h_2) . Otherwise (x^*, y^*) would not be the solution (see



Figure 2.5a

Figure 2.5b). But this means that (f_1, f_2) can be expressed as a linear combination of (g_1, g_2) and (h_1, h_2) and that the weights α and β will be positive (or at least non-negative). For the example of Figure 2.5a we have $(f_1, f_2) = 2(g_1, g_2) + 1(h_1, h_2)$. That is,

$$f_1 = 2g_1 + h_1$$
 and $f_2 = 2g_2 + h_2$.

(When we apply these two conditions to Example 3.2 we get $4 = 2 \times 1 + 1 \times 2$ and $7 = 2 \times 3 + 1 \times 1$.)





Figure 2.6

Consider two special cases. Case (i): The lines $f_1x + f_2y = v$ and $g_1x + g_2y = a$ coincide, where v denotes the solution value of the objective function. Then the arrows (f_1 , f_2) and (g_1 , g_2) point in the same direction, and we must have

$$f_1 = \alpha g_1 + 0h_1$$
 and $f_2 = \alpha g_2 + 0h_2$

for $\alpha > 0$. What does this tell us? Consult Figure 2.6. An increase in the *B* resource, shifting the boundary of the *h* constraint out from ℓ_b to ℓ_{b+1} , will not lead to any increase in the value of the objective function *f*. The diagram shows that there is no production plan in the expanded feasible region that puts us on a higher level curve. Clearly, the value of resource *B* is zero: additional amounts of it are not beneficial. This is why $\beta = 0$.

Case (ii): The lines $f_1x + f_2y = v$ and $h_1x + h_2y = b$ coincide, so (f_1, f_2) and (h_1, h_2) are colinear. Hence

$$f_1 = 0g_1 + \beta h_1$$
 and $f_2 = 0g_2 + \beta h_2$

for $\beta > 0$. This time an increase in the *A* resource will not increase the solution value of the objective function. You can confirm this by drawing a diagram analogous to Figure 2.6. The value to society of resource *A* is zero in this case.

Consider the typical case, with α and β both positive. Suppose that (f_1, f_2) is close to (g_1, g_2) as depicted in Figure 2.7a. Then α will be large relative to β , and this tells us that an increase in resource *A* will have a bigger impact on the objective function than an increase in resource *B*. We demonstrate this by considering in turn what happens when the amount available of input *A*


Figure 2.7a

increases from *a* to a + 1 and then when the amount of input *B* increases to b + 1. When input *A* increases to a + 1 the boundary of the *g* constraint, involving resource *A*, shifts up from ℓ_a to ℓ_{a+1} , as Figure 2.7a shows. The other boundary line is unchanged, because *b* has not changed. We can move to a higher level curve, reflecting an increase in the solution value of *f*. The optimal plan moves from S° to S^{a+1} . Figure 2.7b shows what happens when the amount available of input *B* increases to b + 1. The boundary of the *h* constraint, involving resource *B*, shifts out, from ℓ_b to ℓ_{b+1} , and ℓ_a is unchanged. We again move to a higher level curve (from S° to S^{b+1}) but the move is not nearly as great as when we get an additional unit of resource *A*. Therefore, resource *A* is substantially more valuable than resource *B*: There is a much bigger increase in the solution value of *f* when we get an extra unit of *A*. That is why α is much bigger than β . To convince yourself that α is *precisely* the rate at which the solution value of *f* increases per additional unit of resource *A*—and analogously for *B*—go back to the calculation of Example 3.2.





Basic Models and Tools

The Lagrangian variables α and β are prices in the sense that they equal the value of additional units of the respective resources. For the planning interpretation of the constrained maximization problem, the variable α is the cost imposed on society by a firm using a unit of *A*. This unit of *A* could be employed elsewhere to generate α additional units of "social welfare"—assuming that is what *f* measures. Imposing a cost on the firm of α per unit of *A* employed by the firm promotes efficiency in that it forces the firm to provide at least α units of social welfare per unit of *A* employed. Otherwise it would take a loss. The same holds for β with respect to resource *B*. (Chapter 4 investigates the problem of motivating the manager of the firm to maximize profit.) We haven't discussed incentives in this section, but we have seen that prices can in principle be used to guide an economy, or a firm within an economy, to an efficient outcome. We didn't begin with the determination to employ prices. The prices were forced on us by the mathematics.

If the functions f, g, and h are nonlinear, then the preceding argument goes through if we interpret f_1 as the partial derivative of f with respect to x, evaluated at the optimal plan, with f_2 representing the partial of f with respect to y, also evaluated at the optimal plan, and similarly for g_1 , g_2 , h_1 , and h_2 . Confirm that [6] and [7] are the first-order conditions associated with the maximization of

$$\mathscr{L} \equiv f(x, y) - \alpha g(x, y) - \beta h(x, y).$$

Recall that g(x, y) is the amount of A used up by the plan (x, y). If the price α is the cost to society of employing one unit of A, then $\alpha g(x, y)$ is the cost to society of the amount of A required by the production plan (x, y). Similarly, $\beta h(x, y)$ is the cost to society of the amount of B required by the plan (x, y). Therefore, maximization of \mathcal{L} can be interpreted as the maximization of the value to society of the plan (x, y) net of the cost to society of the resources consumed by that plan.

3.4 The converse of the Pythagorean theorem

The Pythagorean theorem proves that $a^2 + b^2 = c^2 if\theta$ is a right angle (see Figure 2.3). To prove that θ is a right angle $ifa^2 + b^2 = c^2$, drop a line from the vertex at the intersection of sides *a* and *c*, so that the line meets side *b* at a right angle (Figure 2.8). Call this line *d*, and let *e* represent the third side of the right triangle, which has as its other sides *a* and *d*. Let the same letters represent the *length* of the sides, and consider the following equations:

$$a^2 + b^2 = c^2,$$
 [8]

$$d^2 + e^2 = a^2,$$
 [9]

$$d^2 + (b - e)^2 = c^2,$$
[10]

We are given [8] and the other two are consequences of the Pythagorean theorem. Rewrite [10], after replacing d^2 with $a^2 - e^2$ (from [9]) and c^2 with $a^2 + b^2$ (from [8]). We get

$$a^2 - e^2 + b^2 - 2be + e^2 = a^2 + b^2$$
,

3. Lagrangian Multipliers





which reduces to 2be = 0. We are given $b \neq 0$, and hence e = 0. Therefore, *a* and *d* coincide. It follows that θ is a right angle. (Could θ be greater than ninety degrees, in which case the line *d* would be to the *left* of side *a*? In that case, draw a new diagram by extending the base *b* of the original triangle to the left so that it meets the line *d*, which is perpendicular to the extended base line. Equations [8] and [9] still hold but [10] is replaced by $d^2 + (b + e)^2 = c^2$, and the three equations still yield 2be = 0.)

Source

Lagrangian theory with inequality constraints is the creation of Kuhn and Tucker (1950). The Lagrangian multipliers are often called Kuhn-Tucker multipliers.

Links

A good introductory treatment of the general problem of maximizing a function of an arbitrary number of variables subject to an arbitrary number of constraints can be found in Chapter 12 of Weintraub (1982). Chapters 5 and 6 of Novshek (1993) provide a more thorough account, as does Chapter 6 of Binmore and Davies (2001). Koopmans (1957) remains a superb elucidation of the lessons of constrained optimization theory for resource allocation.

Problem set

- **1.** Given the constants f_1 and f_2 , use an algebraic argument to show that if we take a step of unit length in any direction from the point (x, y) then we will obtain the greatest increase in the value of $f_1x + f_2y$ if we move in the direction (f_1, f_2) .
 - **A.** $(f_1, f_2) = (0, 1), (g_1, g_2) = (2, 2), \text{ and } (h_1, h_2) = (1, 0).$ Draw a diagram to show that (f_1, f_2) does not lie between (g_1, g_2) and (h_1, h_2) . Now use algebra to show that we cannot have $(f_1, f_2) = \alpha(g_1, g_2) + \beta(h_1, h_2)$ for nonnegative α and β .
 - **B.** Repeat A with $(f_1, f_2) = (4, 1), (g_1, g_2) = (1, 2), and <math>(h_1, h_2) = (2, 1).$
 - **C.** $(f_1, f_2) = (2, 2), (g_1, g_2) = (0, 1), \text{ and } (h_1, h_2) = (1, 0).$ Draw a diagram to show that (f_1, f_2) lies between (g_1, g_2) and (h_1, h_2) , and then find $\alpha > 0$ and $\beta > 0$ such that $(f_1, f_2) = \alpha(g_1, g_2) + \beta(h_1, h_2).$
 - **D.** Repeat C with $(f_1, f_2) = (2, 1), (g_1, g_2) = (1, 2), \text{ and } (h_1, h_2) = (4, 1).$
- 2. Consider the standard consumer choice problem:

maximize $U(x, y) \equiv xy$ subject to the budget constraint $x + 4y \le 24$.

Of course, *U* is the utility function. Utility depends on the amounts *x* and *y* consumed of the two goods. The price of *X* is \$1 and the price of *Y* is \$4. Income is \$24. The constraint is the budget line.

- **A.** Draw the indifference curve through (4, 3) and the indifference curve through (4, 5). Try to be reasonable accurate.
- **B.** Use calculus to find the basket that maximizes utility subject to the budget constraint. To turn this into a one-variable problem, first express the budget constraint as an equality (why?), solve it for *y* as a function of *x*, and then substitute this expression for *y* into the utility function.
- **C.** Let g(x, y) = x + 4y represent the left-hand side of the budget constraint. Compute the following four partial derivatives: $\partial U(x, y)/\partial x$, $\partial U(x, y)/\partial y$, $\partial g(x, y)/\partial x$, and $\partial g(x, y)/\partial y$. Remember: The partial derivative of f with respect to x, denoted by $\partial f/\partial x$, is obtained by treating y as a constant. (For instance, if $f(x, y) = x^2 + yx + y^2$ then $\partial f/\partial x = 2x + y$. Similarly, the partial derivative of f with respect to y, denoted by $\partial f/\partial y$, is obtained by treating x as a constant.
- **D.** Evaluate the partial derivatives of part C at the chosen consumption plan. Find a positive number α such that

$$\frac{\partial U(x^0, y^0)}{\partial x} = \alpha \times \frac{\partial g(x^0, y^0)}{\partial x} \text{ and } \frac{\partial U(x^0, y^0)}{\partial y} = \alpha \times \frac{\partial g(x^0, y^0)}{\partial y}$$

where (x^0, y^0) represents the chosen consumption plan from part B.

E. Now solve this problem:

maximize
$$U(x, y) \equiv xy$$
 subject to $x + 4y = \Omega$.

Note that we have just replaced income in the budget constraint with the variable Ω . The chosen basket (x^* , y^*) will now be a function of Ω . Now substitute x^* and y^* into the utility function U = xy to get utility as a function of Ω . Now, take the derivative of this function (with respect to Ω) and evaluate it at $\Omega = 24$. The number that you get will equal the value of α from part *D*.

4 THE COMPOSITE COMMODITY MODEL

This section justifies the two-commodity model of consumer choice. To do so we must test it against the complete model with a large number of goods—an arbitrary number, in fact.

In the contrived, composite commodity model *X* is a conventional good, which we also refer to as the zeroth good, with *x* denoting the amount of commodity *X* demanded. Assuming that the prices of all goods other than *X* are constant, we let *y* denote total expenditure on all goods other than *X*. The second good, *Y*, is called a *composite commodity*. The consumer actually has a total of n + 1 commodities form which to choose.

4. The Composite Commodity Model

4.1 The budget constraint and preferences

Let p_0 be the price of *X* and let θ denote the individual's income. The budget constraint is "total expenditure = income," and this is equivalent to

expenditure on X + expenditure on all other goods = income

hence $p_0 x + y = \theta$.

Therefore, we can express the individual's budget constraint in terms of the conventional good *X* and the contrived, composite commodity *Y*.

What about individual preferences? Let (x', y') and (x'', y'') be two different commodity bundles in the composite model. We say that the individual prefers (x', y') to (x'', y'') if and only if the individual prefers (x', b') to (x'', b'') where b' is the most desirable basket of the *n* goods other than X that the individual can buy with y' dollars, given that he or she will consume x' units of X, and b'' is the most desirable basket of goods other than X that the individual can buy with y'' dollars, given that he or she will consume x'' units of X. Note that b' itself is a collection of *n* commodities (involving everything but X) and so is b''. The individual's primitive preferences have been compressed into a preference for bundles in the composite commodity model. We lose a lot of information in the process. The bundle (x^*, y^*) that is the most preferred of all the bundles satisfying the budget constraint identifies expenditure y* on all goods other than X, but we have no idea how y* is distributed across the individual commodities. However, the composite commodity model does tell us that the individual demands exactly x* units of commodity X.

4.2 The composite commodity theorem

Section 4.1 showed rigorously that the budget constraint can be expressed in terms of *x* and *y*, given the price p_0 of *X* and the individual's income θ . It also suggested that the consumer's preferences can be squeezed into this mold. In this section we give a rigorous proof that the resulting preferences can be used to identify the amount of commodity *X* that the individual would actually demand in the real world of n + 1 goods.

Let x_c denote the amount of commodity c consumed, for c = 0, 1, 2, ..., n, and let p_c denote its price. Let $u(x_0, x_1, ..., x_n)$ represent the individual's utility function. The consumer will choose the consumption plan $(x_0, x_1, ..., x_n)$ that maximizes utility subject to the budget constraint. We refer to this as problem *B*.

DEFINITION: The basic problem

Problem *B*: Find the values $x_0, x_1, x_2, \ldots, x_n$ that maximize $u(x_0, x_1, x_2, \ldots, x_n)$ subject to $p_0x_0 + p_1x_1 + p_2x_2 + \cdots + p_nx_n \le \theta$.

We want to solve this and show that the solution value of x_0 is equal to the solution value of x in the composite commodity model—assuming that we use the same income and the same prices in both cases.

Before we can even state the maximization problem for the composite commodity model we have to derive the two-commodity utility function U from the primitive utility function u of the basic problem. Simply put, U(x, y) is the utility from consuming x units of X along with the best combination of the other goods and services that costs y dollars.

DEFINITION: The contrived utility function U

 $U(z_0, y)$ is the value of $u(x_0, x_1, x_2, ..., x_n)$ when we choose $x_0, x_1, x_2, ..., x_n$ to maximize $u(x_0, x_1, x_2, ..., x_n)$ subject only to the restrictions

 $x_0 = z_0$ and $p_1 x_1 + p_2 x_2 + \dots + p_n x_n \le y$.

As we will see, the basic problem is closely related to the maximization of U subject to the restriction that expenditure on X and Y cannot exceed θ .

DEFINITION: The contrived problem

Problem *C*: Choose α and β to maximize $U(\alpha, \beta)$ subject to $p_0\alpha + \beta \le \theta$.

The solutions to problems B and C are related in the following way.

The composite commodity theorem

If $(z_0, z_1, z_2, ..., z_n)$ is a solution to problem *B* then for $\alpha = z_0$ and $\beta = p_1z_1 + p_2z_2 + \cdots + p_nz_n$, the two-commodity bundle (α, β) is a solution to problem *C*. Conversely, if (α, β) is a solution to problem *C* there is some solution $(z_0, z_1, z_2, ..., z_n)$ to problem B such that $z_0 = \alpha$ and $p_1z_1 + p_2z_2 + \cdots + p_nz_n = \beta$. Moreover, for any solutions $(z_0, z_1, z_2, ..., z_n)$ and (α, β) to the respective problems *B* and *C* we have $u(z_0, z_1, z_2, ..., z_n) = U(\alpha, \beta)$.

Proof

Part 1. We show that if $(z_0, z_1, z_2, ..., z_n)$ is any solution to problem *B* then (α, β) solves problem *C* if $\alpha = z_0$ and $\beta = p_1 z_1 + p_2 z_2 + \cdots + p_n z_n$.

Suppose, then, that $(z_0, z_1, z_2, ..., z_n)$ is a solution to problem *B*. Set $\alpha = z_0$ and $\beta = p_1 z_1 + p_2 z_2 + \cdots + p_n z_n$. We have $U(\alpha, \beta) \ge u(z_0, z_1, z_2, ..., z_n)$ by definition of *U*, because $\alpha = z_0$ and $p_1 z_1 + p_2 z_2 + \cdots + p_n z_n \le \beta$ certainly hold. Clearly, $p_0 z_0 + \beta \le \theta$ holds. Therefore, (α, β) satisfies the constraints of problem *C*. Suppose that (α^*, β^*) maximizes *U* subject to the two constraints of problem *C*. Then $U(\alpha^*, \beta^*) \ge U(\alpha, \beta)$ because (α, β) is feasible for problem *C*. By definition of *U*, there is some $(x_0, x_1, x_2, ..., x_n)$ such that

$$U(\alpha^*, \beta^*) = u(x_0, x_1, x_2, \ldots, x_n),$$

with $x_0 = \alpha^*$ and $p_1 x_1 + p_2 x_2 + \dots + p_n x_n \le \beta^*$.

4. The Composite Commodity Model

But $p_0x_0 + p_1x_1 + p_2x_2 + \cdots + p_nx_n \le p_0\alpha^* + \beta^* \le \theta$. This means that $(x_0, x_1, x_2, \ldots, x_n)$ is feasible for problem *B*. Therefore

$$u(z_0, z_1, z_2, \ldots, z_n) \geq u(x_0, x_1, x_2, \ldots, x_n)$$

because $(z_0, z_1, z_2, ..., z_n)$ gives the maximum value of *u* subject to the feasibility constraints of problem *B*. We have proved the following:

$$u(z_0, z_1, z_2, \dots, z_n) \ge u(x_0, x_1, x_2, \dots, x_n) = U(\alpha^*, \beta^*) \ge U(\alpha, \beta)$$

$$\ge u(z_0, z_1, z_2, \dots, z_n).$$

This can only hold if all of the inequalities are satisfied as *equalities*. Then $U(\alpha, \beta) = U(\alpha^*, \beta^*)$. Note that (α^*, β^*) is the name we have given to an arbitrary solution to problem *C* and (α, β) satisfies the constraint of problem *C*. Therefore, (α, β) is also a solution to problem *C*. We have thus proved the first part of our claim.

Proof

Part 2. We show that if (α, β) is a solution to problem *C* then there is some solution $(z_0, z_1, z_2, ..., z_n)$ to problem *B* such that $z_0 = \alpha$ and $p_1 z_1 + p_2 z_2 + \cdots + p_n z_n = \beta$.

Suppose that (α, β) is a solution to problem *C*. By definition of *U* there is some $(z_0, z_1, z_2, ..., z_n)$ such that $\alpha = z_0, \beta \ge p_1 z_1 + p_2 z_2 + \cdots + p_n z_n$, and

$$u(z_0, z_1, z_2, \ldots, z_n) \ge u(x_0, x_1, x_2, \ldots, x_n)$$

for all $(x_0, x_1, x_2, ..., x_n)$ such that $x_0 = \alpha$ and $p_1x_1 + p_2x_2 + \cdots + p_nx_n \le \beta$. Because (α, β) satisfies the constraint of problem *C* we have

$$p_0 z_0 + p_1 z_1 + p_2 z_2 + \dots + p_n z_n \le p_0 \alpha + \beta \le \theta$$

and hence $(z_0, z_1, z_2, ..., z_n)$ satisfies the constraint of problem *B*. Next we show that $(z_0, z_1, z_2, ..., z_n)$ is actually a solution to problem *B*. Let $(x_0, x_1, x_2, ..., x_n)$ be any consumption plan satisfying the constraints of problem B. Set $\alpha^* = x_0$ and $\beta^* = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n$. Then

 $p_0 \alpha^* + \beta^* = p_0 x_0 + p_1 x_1 + p_2 x_2 + \dots + p_n x_n \le \theta.$

Therefore, by definition of *U* we have

$$u(x_0, x_1, x_2, \ldots, x_n) \leq U(\alpha^*, \beta^*).$$

Note that $p_0\alpha^* + \beta^* \le \theta$, and therefore $U(\alpha^*, \beta^*) \le U(\alpha, \beta)$, because (α, β) solves problem *C* and (α^*, β^*) is feasible for *C*. Therefore, we have established the following:

$$u(x_0, x_1, x_2, \ldots, x_n) \leq U(\alpha^*, \beta^*) \leq U(\alpha, \beta) = u(z_0, z_1, z_2, \ldots, z_n).$$

Therefore, $u(x_0, x_1, x_2, ..., x_n) \le u(z_0, z_1, z_2, ..., z_n)$ for any values $x_0, x_1, x_2, ..., x_n$ that satisfy the constraint for problem *B*. This proves that $(z_0, z_1, z_2, ..., z_n)$ is a solution to problem *B*.

Proof

Part 3. We have to show that the maximum utility for problem *B* equals the maximum utility for problem *C*. We have already done that because part 1 established that $u(z_0, z_1, z_2, ..., z_n) = U(\alpha, \beta)$ holds for any two solutions $(z_0, z_1, z_2, ..., z_n)$ and (α, β) of the respective problems.

In some applications, X is also a composite commodity: In an economic analysis of health care, x would be total expenditure on health care and y would be expenditure on everything else. You can see that the justification for employing a composite commodity Y would also be valid for X when x is expenditure on health care, or education, or food, and so forth.

Source

The composite commodity theorem was discovered independently by Hicks (1939) and Leontief (1936).

5 QUASI-LINEAR PREFERENCES

Having simplified things by reducing the number of commodities to two, we now show how a simple family of utility functions can be used to bring additional clarity. Suppose the individual's utility function U(x, y) has the special form B(x) + y. This function is linear in y but not necessarily in x.

DEFINITION: Quasi-linear function

A quasi-linear function of two variables *x* and *y* has the form B(x) + y, where *B* can be any function of *x*.

Quasi-linear preferences endow economic models with some very nice properties. They will be used to uncover basic principles at relatively low cost.

We assume throughout this section that *Y* is a private good that is divisible. This means two things: (i) all individuals care only about their own consumption of *Y* (but not anyone else's) and prefer more *Y* to less; and (ii) any amount of any individual *i*'s consumption of *Y* can be transferred to any individual *j*. Assumption (i) implies that if *x* is unchanged but individual *i*'s consumption of *Y* increases then individual *i*'s utility increases, regardless of how anyone else's consumption of *Y* changes. Assumption (ii) implies that if individual *i* has a positive amount ϵ of the private good, however small, any positive fraction of ϵ can be transferred from individual *i* to someone else. One of the advantages of assuming quasi-linear preferences is that efficiency is equivalent to maximization of the sum of individual utilities (subject to the limitations inherent in resource constraints, etc.). This is demonstrated in two subsections. Subsection 5.1 offers a short, easy proof, but it does not incorporate the constraint that an individual's consumption of *Y* cannot fall below zero. Using calculus, and assuming that consumption is strictly positive, Subsection 1.2 of Chapter 8

5. Quasi-Linear Preferences

establishes the equivalence of efficiency and total utility maximization when preferences are quasilinear.

5.1 Efficiency with quasi-linear utility

Assuming that *Y* is a divisible private good and that individual preferences are quasilinear, we show that an allocation is efficient if and only if it maximizes total utility. We have already seen that for *any* model, any outcome that maximizes total utility is efficient (Section 4 of Chapter 1). Without quasi-linear preferences, an outcome can be efficient without maximizing total community utility, as Example 5.3 demonstrates. The next example highlights the role of the divisibility assumption.

Example 5.1: Efficiency without divisibility

There are two feasible outcomes, A and B, and two individuals whose utility functions are displayed in Table 2.1. A is efficient, because a move to B would lower person 2's utility. But A certainly does not maximize total utility. If a divisible private good were available, some of it could be transferred from person 1 to person 2 at outcome B to increase U_2 . And *if* both U_1 and U_2 were quasilinear the transfer could be accomplished in a way that left both 1 and 2 with more utility than they would have at A. But this transfer would create a new outcome C, contradicting the fact that only two outcomes are feasible in the present case. Therefore, in this example there is no divisible commodity in the background.

This section explains why, when every individual's utility function is quasilinear, and *Y* is a divisible private good, every efficient allocation maximizes total utility. We do this by showing how everyone's utility can be increased at any allocation that does *not* maximize total utility.

Each individual *i*'s utility has the form $U_i(x, y_i) = B_i(x) + y_i$. Therefore, if y_i changes to $y_i + \Delta y_i$ but *x* remains the same, then the change in the individual's utility is

$$\Delta U_i = B_i(x) + y_i + \Delta y_i - [B_i(x) + y_i] = \Delta y_i.$$

In brief, if *x* does not change, then for each individual *i* we have $\Delta U_i = \Delta y_i$. In words, the change in individual *i*'s utility is equal to the change in *i*'s consumption

Table 2.1

Outcome	U_1	U_2
A	2	20
B	100	15

of *Y* if *x* is unchanged. Now, suppose that outcomes *F* and *G* are both feasible, but total utility is higher at *G* than at *F*. Then we can create outcome *H* from *G* by redistributing commodity *Y* without changing the value of *x*. Because *x* does not change, there is no transfer of resources from the production of *Y* to the production of *X*. Consequently, the total amount of *Y* available for consumption will be the same at *G* and *H*, and thus if we create *H* from *G* by

redistributing Y the new outcome H will be feasible. And because total utility is higher at G than at F the redistribution can be done in a way that increases everyone's utility.

Example 5.2: Three individuals

Outcomes *F* and *G* are given, and the individual utility levels realized at each are specified Table 2.2. *F* is not efficient, but outcome *G* by itself does not demonstrate that because persons 2 and 3 have lower utility at *G* than at *F*. However, if we create *H* from *G* by setting $\Delta x = 0$, $\Delta y_1 = -12$, $\Delta y_2 = +8$, and $\Delta y_3 = +4$ then the sum of the changes in *Y* is zero. Therefore, *H* is feasible. Because $\Delta x = 0$ we have $U_1(H) = U_1(G) + \Delta y_1 = 29 - 12 = 17$. And $U_2(H) = U_2(G) + \Delta y_2 = 14 + 8 = 22$. Finally, $U_3(H) = U_3(G) + \Delta y_3 = 28 + 4 = 32$. Outcome *H* gives everyone more utility than *F*, and thus we have demonstrate that *F* is not efficient.

Example 5.2 does not specify the consumption y_1 of individual 1 at *G*. Therefore, we cannot be sure that $y_1 - 12$ is positive, or at least zero. Our argument was perfectly rigorous, provided that $y_i \ge 0$ is not required. In many models, the original consumption levels of the private good are assumed to be high enough so that there is no danger of driving someone's consumption of that good below zero. We continue to ignore the constraint $y_i \ge 0$ until Section 1.2 of Chapter 8.

Efficiency theorem for quasi-linear utility functions

If y_i is allowed to have any value (positive, negative, or zero) then an allocation is efficient if and only if it maximizes total utility.

There will be more than one efficient allocation because if (x, y) maximizes total utility then it is efficient. Then (x, y') is efficient for any y' such that $\sum_{i \in N} y'_i = \sum_{i \in N} y_i$. That's because x does not change and total consumption of Y does not change, and thus total utility is still maximized. Thus (x, y') must be efficient.

	Table	2.2
--	-------	-----

	1	2	3	Total
$egin{array}{l} U_i(F) \ U_i(G) \ U_i(H) \end{array}$	15	20	29	64
	29	14	28	71
	17	22	32	71

Here is the proof of the efficiency theorem for an arbitrary number n of individuals. The individuals are indexed by i = 1, 2, ..., n. If X is a private good, then x specifies an assignment of some amount X to each individual. If X is a public good, then x denotes the level of that good provided to all. It is also possible that x denotes some mix of public and private goods. The following argument works for any interpretation of the variable x.

Each individual *i* has a utility function of the form $U_i(x, y_i) = B_i(x) + y_i$. We will show that an outcome cannot be efficient if it does not maximize total utility.

5. Quasi-Linear Preferences

Let (\tilde{x}, \tilde{y}) be a feasible allocation. If it does not maximize the sum of individual utilities then there is another feasible allocation (x, y) such that

$$\sum_{i \in N} [B_i(x) + y_i] > \sum_{i \in N} [B_i(\tilde{x}) + \tilde{y}_i.$$
[11]

Define a new allocation (x^0, y^0) such that $x^0 = x$ and

$$y_i^0 = B_i(\tilde{x}) + \tilde{y}_i - B_i(x) + \frac{1}{n} \times \Sigma_{h \in N} [B_h(x) + y_h - B_h(\tilde{x}) - \tilde{y}_h].$$

Note that

$$\Sigma_{i\epsilon N} y_i^0 = \Sigma_{i\epsilon N} B_i(\tilde{x}) + \Sigma_{i\epsilon N} \tilde{y}_i - \Sigma_{i\epsilon N} B_i(x) + \Sigma_{i\epsilon N} B_i(x) + \Sigma_{i\epsilon N} y_i - \Sigma_{i\epsilon N} B_i(\tilde{x}) - \Sigma_{i\epsilon N} \tilde{y}_i = \Sigma_{i\epsilon N} y_i$$

In words, the allocation (x^0, y^0) is created from (x, y) by leaving the value of x unchanged but redistributing commodity Y. This means that (x^0, y^0) is feasible. But for each i in N

$$B_{i}(x^{0}) + y_{i}^{0} = B_{i}(x) + B_{i}(\tilde{x}) + \tilde{y}_{i} - B_{i}(x) + \frac{1}{n} \times \Sigma_{h \in N}[B_{h}(x) + y_{h} - B_{h}(\tilde{x}) - \tilde{y}_{h}]$$

and thus

$$B_{i}(x^{0}) + y_{i}^{0} = B_{i}(\tilde{x}) + \tilde{y}_{i} + \frac{1}{n} \times \Sigma_{h \in N}[B_{h}(x) + y_{h} - B_{h}(\tilde{x}) - \tilde{y}_{h}]$$
[12]

for each individual *i*.

Statement [11] implies that $\Sigma_{h \in N}[B_h(x) + y_h - B_h(\tilde{x}) - \tilde{y}_h]$ is a positive number, and thus [12] implies that $B_i(x^0) + y_i^0 > B_i(\tilde{x}) + \tilde{y}_i$ holds for each individual *i*. We conclude that (\tilde{x}, \tilde{y}) is not efficient.

Our proof clearly depends on the divisibility of Y—think of Y as money but X could be available only in discrete units, although the argument applies equally well when X is divisible. We conclude this subsection by showing that the quasi-linear assumption is crucial.

Example 5.3: Counterexample when one of the utility functions is not quasi-linear

There are two individuals, 1 and 2. $U_1(x, y_1) = \frac{1}{2} x + y_1$ and $U_2(x, y_2) = xy_2$. Person 1's utility function is quasi-linear, but 2's is not. The feasible outcomes are values of x, y_1 , and y_2 such that $x + y_1 + y_2 = 4$. Outcome A has x = 2, $y_1 = 1$, and $y_2 = 1$. Then A is feasible. It is also efficient. To prove that A is efficient we begin by trying to increase U_2 without changing U_1 . To keep U_1 constant we will have to have $\frac{1}{2}x + y_1 = 2$. We also have to satisfy the feasibility requirement $x + y_1 + y_2 = 4$. If we subtract $\frac{1}{2}x + y_1$ from the left-hand side of the last equation and we subtract 2 from the right-hand side we get $\frac{1}{2}x + y_2 = 2$. This means that $y_2 = 2 - \frac{1}{2}x$. Therefore, we want to maximize $U_2 = xy_2$ subject to $y_2 = 2 - \frac{1}{2}x$. That is equivalent to maximizing $x(2 - \frac{1}{2}x) = 2x - \frac{1}{2}x^2$. You can use calculus to get the solution value x = 2. Alternatively, note that $2x - \frac{1}{2}x^2 = 2 - \frac{1}{2}(x - 2)^2$. To maximize $2 - \frac{1}{2}(x - 2)^2$ we have to set x = 2; otherwise we will be subtracting a positive number from 2 to get the value U_2 , and that is less than $U_2 = 2$, which we get when x = 2. We have x = 2 and $y_2 = 2 - \frac{1}{2} x$. Therefore, $y_2 = 1$. If x = 2, $y_2 = 1$, and feasibility requires $x + y_1 + y_2 = 4$, we must have $y_1 = 1$. That is precisely outcome *A*. We have demonstrated that if U_1 must equal its value at *A* then any feasible outcome other than *A* must yield a lower value of U_2 than the value of U_2 at *A*. This implies that, starting from *A*, we cannot increase U_2 without lowering U_1 . It also implies that, starting from *A*, we cannot increase U_1 without lowering U_2 . (Why?) Therefore, *A* is efficient. But *A* does not maximize $U_1 + U_2$ over the set of feasible allocations. At *A* we have $U_1 + U_2 = 2 + 2 \times 1 = 4$. If *B* has $x = 2 = y_2$ and $y_1 = 0$, then at *B* we have $U_1 + U_2 = 1 + 2 \times 2 = 5$. *A* is efficient, but it does not maximize the sum of utilities. This does not depend on the fact that y_1 is 0 at *B*: Set x = 2, $y_1 = \epsilon$, and $y_2 = 2 - \epsilon$ to create outcome *C*. Then *C* is feasible. We can take $\epsilon > 0$ sufficiently small so that $U_1 + U_2$ is as close as we like to 5.

Our proof did not acknowledge the possibility that when Δy_i is negative the resulting consumption of the private good by individual *i*, which is $y_i + \Delta y_i$, could be negative. Fortunately, there are many applications in which we do not need to worry about the constraint $y_i \ge 0$, simply because there is reason to believe that no one's consumption of the private good will be driven to zero. For completeness and rigor, Section 1.2 of Chapter 8 explicitly imposes $y_i \ge 0$ for each individual *i*.

∂ **5.2** Quasi-linear preference and demand

We now focus on a single individual, so we can drop the subscript *i*. The individual's utility depends only on his or her own consumption of *X* and *Y*. The utility function has the quasi-linear form U(x, y) = B(x) + y. We also assume diminishing marginal utility, which means that B''(x) < 0 at all *x*. We show (in Section 5.3) that the function *B* can be recovered from the individual's demand function for *X*.

In this section we demonstrate that, beyond a minimum income level, when the individual's income increases he or she will not increase the consumption of *X*, assuming that prices do not change. Assuming that the price of *X* is not so high that the individual demands zero units of *X*, maximization of utility subject to the budget constraint implies that B'(x) equals the price ratio. Let x^* be the value of *x* for which this holds. As income increases, we will still have $B'(x^*)$ equal to the price ratio, so x^* will still be the individual's demand for *X*. (*This claim is true only for quasi-linear preferences.*)

Formally, we maximize B(x) + y subject to $p_1x + p_2y = \theta$, and $x \ge 0$ and $y \ge 0$. We can solve the budget constraint for y. We have $y = (\theta - p_1x)/p_2$. Therefore we maximize

$$V(x) = B(x) + \frac{\theta - p_1 x}{p_2},$$

5. Quasi-Linear Preferences

a function of *x*, subject to $0 \le x \le \theta/p_1$. (Note that $y = (\theta - p_1x)/p_2$ will be nonnegative if and only if $x \le \theta/p_1$.) Assume that $0 < x^* < \theta/p_1$ at the maximizing value of *x*. Then the first derivative of *V*(*x*) must equal zero at x^* . That is, $B'(x^*) - p_1/p_2 = 0$, which of course implies $B'(x^*) = p_1/p_2$. The solution x^* will be unique because B''(x) < 0 at all *x*. Moreover, $B'(x^*) = p_1/p_2$ will still hold if income increases and prices do not change. Therefore, the demand for *X* does not change when income changes.

Income effect with quasi-linear utility functions

If U = B(x) + y, and both *x* and *y* are positive at the chosen consumption plan, then the demand for *x* will not increase when income increases.

We can view this in terms of the tangency condition for consumer choice: The indifference curve is tangent to the budget line at the chosen bundle. This means that the marginal rate of substitution (MRS) equals the price ratio. To determine the MRS we start with the fact the utility is constant along an indifference curve. Therefore, the equation of an indifference curve is $B(x) + y = \ell$, where ℓ is a constant. We have $y = \ell - B(x)$. The derivative of this function is -B'(x), which is thus the slope of the indifference curve at the point (x, y). Because the MRS is the absolute value of the slope of the indifference curve, the MRS is B'(x). The MRS is independent of y so, with x on the horizontal axis, the MRS is constant along any vertical line. If we have a consumer optimum (x^*, y^*) that does not occur at a corner point of the budget region we will have MRS = p_1/p_2 , and that can occur at only one point on the budget line. As income increases and the budget line shifts out parallel to itself the new optimum will also occur at a point where MRS = p_1/p_2 . The MRS doesn't change. This can only happen on the vertical line through x^* : There is no change in the demand for X. (The demand for X does change when p_1 or p_2 changes.)

Now, suppose that $B'(x) = p_1/p_2$ implies x < 0. Because B'' < 0 we have $B'(0) < p_1/p_2$ and thus $V'(0) = B'(0) - p_1/p_2 < 0$. Because V'' = B'' < 0, we have V'(x) < V'(0) < 0 for all x > 0, and hence V(x) < V(0) for all x > 0. The solution to the constrained utility maximization problem is x = 0. If θ (income) increases we will still have $B'(0) < p_1/p_2$, and thus x = 0 will still solve the constrained utility-maximization problem. There is no income effect on the demand for *X* in this case as well.

There can be an income effect on the demand for *X* only if $B'(\theta/p_1) > p_1/p_2$, and this inequality will fail for θ large enough, because B' falls as *x* increases. Let θ^* satisfy $B'(\theta^*/p_1) = p_1/p_2$. For $\theta > \theta^*$ there is no income effect on the demand for *x*, and even when $\theta \le \theta^*$ there is no income effect if $B'(0) < p_1/p_2$. But beyond a minimum income level θ^* there is no income effect for sure.

Example 5.5: Consumer choice with quasi-linear utility

 $U(x, y) = \ln(x + 1) + y$. Now maximize utility subject to $p_1x + p_2y = \theta$. (If y is a composite commodity set $p_2 = 1$.)

Because the budget constraint implies $y = (\theta - p_1 x)/p_2$ we can maximize $V(x) = \ln(x + 1) + (\theta - p_1 x)/p_2$ subject to $0 \le x \le \theta/p_1$. The first derivative is $(x + 1)^{-1} - p_1/p_2$ and the second derivative is $-(x + 1)^{-2}$, which is always negative. If $V'(x^*) = 0$ and $0 \le x^* \le \theta/p_1$ then x^* is our solution (see Section 2.2). Then we will have $(x^* + 1)^{-1} - p_1/p_2 = 0$, which implies $x^* = p_2/p_1 - 1$. If $p_2/p_1 - 1 < 0$ or $p_2/p_1 - 1 > \theta/p_1$ we know that $p_2/p_1 - 1$ cannot be the demand for *X*. In either case the consumer will demand either zero or θ/p_1 units of *X*. *V*" is negative everywhere so (from Section 2.2) if V'(x) = 0 implies $x > \theta/p_1$ then $x = \theta/p_1$ maximizes *V* subject to $0 \le x \le \theta/p_1$. If V'(x) = 0 implies x < 0 then x = 0 is our solution. We can now display the demand function for *x*:

$$\begin{aligned} x(p_1, p_2, \theta) &= 0 & \text{if } \frac{p_2}{p_1} < 1, \\ x(p_1, p_2, \theta) &= \frac{p_2}{p_1} - 1 & \text{if } 1 \le \frac{p_2}{p_1} \le 1 + \frac{\theta}{p_1} \\ x(p_1, p_2, \theta) &= \frac{\theta}{p_1} & \text{if } \frac{p_2}{p_1} > 1 + \frac{\theta}{p_1}. \end{aligned}$$

By solving the budget constraint $p_1x + p_2y = \theta$ for *y* we can obtain directly the demand function for *Y*:

$$y(p_1, p_2, \theta) = \frac{\theta}{p_2} \quad \text{if } \frac{p_2}{p_1} < 1,$$

$$y(p_1, p_2, \theta) = \frac{\theta + p_1 - p_2}{p_2} \quad \text{if } 1 \le \frac{p_2}{p_1} \le 1 + \frac{\theta}{p_1}$$

$$y(p_1, p_2, \theta) = 0 \quad \text{if } \frac{p_2}{p_1} > 1 + \frac{\theta}{p_1}.$$

Fix p_1 and p_2 and allow θ to vary. If p_1 is larger than p_2 then we have $x(p_1, p_2, \theta) = 0$ for all values of θ . The income effect on the demand for X is zero. If $p_2/p_1 > 1 + \theta/p_1$ then all income is spent on X, and this continues to be the case as θ rises until it reaches $p_2 - p_1$. At this point we have $x = \theta/p_1 = (p_2 - p_1)/p_1 = p_2/p_1 - 1$. As θ rises beyond $p_2 - p_1$ all additional income is spent on commodity Y. Note that if either inequality $1 \le p_2/p_1$ or $p_2/p_1 \le 1 + \theta/p_1$ holds then it continues to hold as θ rises. In summary, for $\theta \ge p_2 - p_1$ or $p_1 > p_2$ there is no increase in the demand for X when income increases. There is an income effect on the demand for X only when $p_1 < p_2$ and even then only in the extreme case of incomes less than $p_2 - p_1$ (or less than $1 - p_1$ if y is a composite commodity and $p_2 = 1$).

∂ **5.3** Consumer surplus

Now we show that if utility is quasi-linear then the demand function for commodity *X* can be used to estimate the utility function. Specifically, if U(x, y) = B(x) + y then the demand curve for *X* can be used to recover the benefit function

B(x). For convenience we assume that B(0) = 0. Quasi-linear utility also means that the area under the demand curve and above the line $P = p_1$, where p_1 is the given price of *X*, is equal to the utility gain from being able to purchase *X* at the price p_1 . We refer to this utility increase as the *consumer surplus*.

DEFINITION: Consumer surplus

The consumer surplus is $U(x, \theta - p_1 x) - U(0, \theta)$, where θ is the individual's income and *x* is the amount of commodity *X* that maximizes *U* subject to the budget equation $p_1 x + y = \theta$. (We simplify by setting $p_2 = 1$.)

Because $p_2 = 1$, the budget equation implies that $y = \theta - p_1 x$. Therefore, $U(x, \theta - p_1 x) - U(0, \theta)$ is the utility from being able to purchase *X* at the price p_1 minus utility when only commodity *Y* is available (in which case the budget equation implies $y = \theta$).

Let's see how the consumer surplus can be recovered from the demand function for *X*. We begin by deriving that demand function. Maximize B(x) + y subject to $p_1x + y = \theta$ and $x \ge 0$ and $y \ge 0$. Equivalently, maximize $B(x) + \theta - p_1x$ subject to $0 \le x \le \theta/p_1$. We assume a range of prices such that the demand for *X* satisfies $0 < x < \theta/p_1$, which implies that the first derivative of $B(x) + \theta - p_1x$ equals zero. That is, $B'(x) = p_1$ is satisfied at the solution to the consumer decision problem. It follows that if we plot $B'(x) = p_1$ on a diagram with *x* on the horizontal axis and p_1 on the vertical axis we will portray the individual's demand curve for *X*. Given p_1 , the curve shows us the value of *x* for which $B'(x) = p_1$, and that is in fact the demand for *X* at the price p_1 . (Strictly speaking, B'(x) is the *inverse* demand function. However, when we plot the graph of B'(x) we can interpret it as the demand curve by taking a given price and finding the value of *x* on the graph at that price. It will be the quantity demanded at that price because it will be the quantity *x* at which B'(x) equals the given price.)

We can't observe the function *B* directly, but we can observe prices and quantities, so we can estimate the demand curve. Let P(x) be the function represented by that demand curve. Suppose, for convenience, that *P* and *B'* are identical. (The demand curve has been estimated with precision.) Using the fundamental theorem of calculus we have

$$B(x) = \int_0^x B'(t) \, dt = \int_0^x P(t) \, dt$$

Because $\int_0^x P(t) dt$ is the area under the curve P(t) from 0 to *x*, we can use the observable demand curve to compute B(x).

Assuming for convenience that B(0) = 0, the consumer surplus is $B(x) + \theta - p_1x - [B(0) + \theta]$, which equals $B(x) - p_1x$. The consumer surplus $B(x) - p_1x$ is the area under the demand curve from 0 to *x* minus p_1x . Now, $p_1 \times x$ is the area of a rectangle with height p_1 and length *x* and thus is the area under the line $P = p_1$ between 0 and *x*. We have shown that the consumer surplus is the area under the demand curve from 0 to *x* minus the area under the horizontal line

 $P = p_1$ between 0 and *x*. In other words, the consumer surplus is area below the demand curve and above the price line, and between 0 and *x*.

Measuring consumer surplus with quasi-linear utility functions

If U = B(x) + y and $p_2 = 1$, then the surplus from consuming *X* at price p_1 is equal to the area under the demand curve for *X* and above the horizontal line drawn p_1 units above the horizontal axis.

Example 5.6: Consumer surplus with quasi-linear utility

 $U(x, y) = \ln(x + 1) + y$, as in Example 5.5. We saw that the individual's demand function is $x = P^{-1} - 1$ if $p_2 = 1$ and P denotes p_1 . (Again, we are assuming a range of prices for which the amount of commodity X demanded is strictly between 0 and θ/p_1 .) Of course, $x = P^{-1} - 1$ implies $P = (x + 1)^{-1}$. This is the inverse demand function, which we want to integrate to determine the area under the demand curve. $\int P dx = \int (x + 1)^{-1} dx = \ln(x + 1) + c$, for arbitrary constant c. Note that $B(x) = \ln(x + 1)$, and if B(x) = 0 we must have c = 0. We have recovered the function B(x) from the demand curve. The consumer surplus is B(x) - Px, which equals the area under the demand curve and above the horizontal line at height P.

We conclude this section by showing that if each individual *i* has a quasilinear utility function then the area under the *market* (or total) demand curve is equal to the aggregate consumer surplus and hence is equal to the total utility realized by the community when each individual is able to purchase *X* at a price of p_1 . Individual *i*'s utility function is $U_i = B_i(x_i) + y_i$, where x_i is the amount of *X* consumed by household *i* and y_i is the amount of *Y* consumed by *i*. The function B_i can be different for different individuals, hence the *i* subscript.

Individual *i*'s consumer surplus is

$$\int_0^x P_i(t)\,dt - Px_i,$$

the area below *i*'s demand curve P_i and above the horizontal line at height *P*. But we can also integrate along the vertical axis: The area under the individual demand curve and above the horizontal line at *P* is

$$\int_P^\infty x_i(\rho)\,d\rho$$

where x_i is consumer *i*'s quantity demanded as a function of the price ρ . Total market demand *q* is the sum of the individual demands, so we can write $q(\rho) = \sum_{i \in N} x_i(\rho)$, where *N* is the set of consumers. Therefore,

$$\sum_{i \in N} \int_{P}^{\infty} x_{i}(\rho) \, d\rho = \int_{P}^{\infty} q(\rho) \, d\rho.$$

5. Quasi-Linear Preferences

But $\int_{P}^{\infty} q(\rho) d\rho$ can be expressed as $\int_{0}^{q} P(t) dt$, where P(q) is the inverse market demand curve—that is, the price at which a total of q units would be demanded in total by all consumers. Clearly, $\sum_{i \in N} Px_i = Pq$. Therefore,

$$\sum_{i \in N} \int_{P}^{\infty} x_{i}(\rho) \, d\rho = \int_{0}^{q} P(t) \, dt - Pq$$

Therefore, the area under the market demand curve and above P is the sum of the areas under the individual demand curves above P. Because the sum of the areas under the individual demand curves is equal to the total utility, we can say that total utility is exactly equal to the area under the market demand curve when each individual's utility function is quasi-linear. Similarly, the total consumer surplus equals the area under the market demand curve and above P.

Measuring total consumer surplus with quasi-linear utility functions

If each individual's utility function has the form $U_i = B_i(x) + y$, and $p_2 = 1$, then the total surplus from consuming *X* at price *P* is the area under the market demand curve for *X* and above the horizontal line drawn *P* units above the horizontal axis.

Example 5.7: Total consumer surplus with quasi-linear utility

To simplify the calculations we will assume *n* identical consumers, each with the utility function $U(x, y) = \ln(x + 1) + y$, as in Example 5.5. When each individual begins with x = 0 and then is able to purchase *X* at price *P*, the individual increase in utility is $\ln(x + 1) - Px$. Therefore, the total increase in utility over the entire community is $n\ln(x + 1) - nPx$. We will show that this equals the area between the market demand curve and the horizontal line at height *P*.

The individual demand function is $x = P^{-1} - 1$ (Example 5.5), so market (or total) demand is *n* times that. If *q* denotes market demand we have

$$q = nP^{-1} - n$$
 and thus $P = n(q + n)^{-1}$.

The second equation is the inverse market demand function, which we want to integrate. We have $\int P \, dq = \int n(q+n)^{-1} \, dq = n \ln(q+n) - n \ln n$. (By subtracting *n* ln *n* we get an area of zero when *q* is zero.) Therefore, the area under the market demand curve and above the horizontal line at *P* is

$$n\ln(q+n) - n\ln n - Pq.$$

(The area from 0 to *q* between the horizontal axis and the line of height *P* is a rectangle of height *P* and width *q*, and thus has area *Pq*.) Now, q = nx, and thus $n \ln(q + n) - n \ln n = n \ln(nx + n) - n \ln n = n \ln[(nx + n)/n] = n \ln(x + 1) = nB(x)$. Therefore, the area under the market demand curve and above the horizontal line at height *P* is equal to n[B(x) - Px], the total consumer surplus. (Compare with Example 5.6.)

112 Basic Models and Tools

Source

The efficiency condition of Section 5.1 first appeared in Samuelson (1954).

Links

See Campbell and Truchon (1988) for a general characterization of efficiency with quasi-linear preferences covering allocations for which $y_i = 0$ for some individuals *i*. See also Conley and Diamantaris (1996). See Katzner (1970, p. 152) for the general result on demand functions and consumer surplus when preferences are not necessarily quasi-linear.

6 DECISION MAKING UNDER UNCERTAINTY

Most of the models in this book either employ a framework in which there is no uncertainty or assume that there are only two possible random events, "bad" and "good," and that the decision maker knows the probability of each event. Therefore, we begin with a study of choice under uncertainty when an action leads to one event with probability π and an alternative event with probability $1 - \pi$. Of course, $0 \le \pi \le 1$. The bad event leads to a low payoff x, and the good event yields a high payoff y. We typically think of x and y as changes in the decision maker's wealth in the respective events. To keep the terminology simple we refer to the prospect of getting x with probability π and y with probability $1 - \pi$ as an asset, even though there will be other applications, such as the purchase of insurance.

DEFINITION: Asset

An asset is any opportunity that yields a specified low payoff *x* with probability π and a specified high payoff *y* with probability $1 - \pi$.

We allow *x* to be negative, because in some cases we want the payoffs to be reported net of the purchase price of the financial instrument.

6.1 Asset preferences

An individual with a current wealth of θ is confronted with a choice between a safe asset (money) that preserves his or her wealth at θ with certainty and a risky asset (an investment) that reduces his or her wealth to *x* with probability π but will cause his or her wealth to increase to *y* with probability $1 - \pi$. Of course, $x < \theta < y$. One important element—but not the only element—of the decision process is the expected payoff. The expected monetary value of an asset is the weighted sum of the monetary payoffs, where each payoff's weight is its probability.

DEFINITION: Expected monetary value (EMV)

If *x* dollars is received with probability π and *y* is received with probability

 $1 - \pi$ then the expected monetary value of the asset is $\pi x + (1 - \pi)y$.

Example 6.1: EMV when the bad outcome is a burglary

An individual's current wealth of \$100 will be reduced to \$40 if he or she is robbed, and that will happen with probability 0.3. Then the EMV of wealth (without insurance) is $0.3 \times 40 + 0.7 \times 100 = 82$.

A risky asset would leave the individual with *x* with probability π and *y* with probability $(1 - \pi)$, with $x < \theta < y$. For most of us there is a value of π sufficiently close to 0 (perhaps *extremely* close) that would induce us to choose this risky asset. And there would be a value of π sufficiently close to 1 that would prompt us to choose the safe asset, with a guaranteed θ . But what about more realistic, intermediate, values of π ? Clearly, the decision would depend on the magnitudes θ , *x*, *y*, on the probability π , and on the individual's preferences under uncertainty. For a wide range of circumstances it is possible to model an individual's preferences by means of a utility-of-wealth function U(w), where *w* is the market value of the individual's wealth. The utility function represents the individual's preferences in the sense that he or she would prefer the risky asset if and only if

$$\pi U(x) + (1 - \pi)U(y) > U(\theta).$$

The expected utility (EU) of an asset is the weighted sum of the payoff utilities, where each weight is the probability of the associated payoff.

DEFINITION: Expected utility (EU)

If *U* is the utility-of-wealth function and *x* dollars is received with probability π and *y* is received with probability $1 - \pi$ then

$$EU = \pi U(x) + (1 - \pi)U(y).$$

Example 6.2: EU when the bad outcome is a burglary

The individual's utility-of-wealth function is $U(w) = 4\sqrt{w}$. Then for the situation of Example 6.1

$$EU = 0.3 \times 4\sqrt{40} + 0.7 \times 4\sqrt{100} = 35.6.$$

It is possible to prove that, under some fairly mild assumptions on the nature of individual preference under uncertainty, for each preference scheme there is a utility-of-wealth-function such that the individual will always choose the asset that leads to the highest expected utility of wealth. In other words, the individual acts so as to maximize his or her expected utility.

DEFINITION: Expected utility maximization

The market will provide the individual with a range of affordable assets, and the individual will choose the one that yields values of *x* and *y* that maximize $\pi U(x) + (1 - \pi)U(y)$ over all affordable assets.

This means that the individual can be represented as an expected utility maximizer. The quantity $\pi U(x) + (1 - \pi)U(y)$ is called the expected utility of an investment that results in a wealth level of *x* with probability π and a wealth level of *y* with probability $1 - \pi$. In general, given a choice between an investment *I* that yields *x* with probability π and yields *y* with probability $1 - \pi$ and an investment *J* that yields *a* with probability ρ and *b* with probability $1 - \rho$, the individual will choose *I* if

$$\pi U(x) + (1 - \pi)U(y) > \rho U(a) + (1 - \rho)U(b)$$

and will choose J if

$$\rho U(a) + (1 - \rho)U(b) > \pi U(x) + (1 - \pi)U(y).$$

We are *assuming* expected utility maximization but, as we have said, it is possible to deduce this property from mild assumptions about individual preference. We do not present the proof in this book, however.

Example 6.3: Individuals with different preferences make different choices

Dale's utility-of-wealth function is $U(w) = 10\sqrt{w}$ and Joanne's is U(w) = 2w. Each has to choose between a safe asset *A* that leaves the individual with \$196 for sure, and a risky asset *B* that yields \$36 with probability $\frac{1}{2}$ and \$400 with probability $\frac{1}{2}$. For Dale we have

$$EU(A) = 10\sqrt{196} = 140$$

and

$$EU(B) = \frac{1}{2} \times 10\sqrt{36} + \frac{1}{2} \times 10\sqrt{400} = 130.$$

For Joanne,

$$EU(A) = 2 \times 196 = 392$$

and

$$EU(B) = \frac{1}{2} \times 2 \times 36 + \frac{1}{2} \times 2 \times 400 = 436.$$

Dale chooses A but Joanne chooses B.

Ì

6. Decision Making Under Uncertainty

Even in an uncertain environment outcomes can be tested for efficiency. We apply the standard definition (of Chapter 1, Section 4) but we used expected utility as the individual's payoff. (If your town built so many parking garages that you always found a spot, no matter where or when you arrived, the outcome would not be efficient. Can you explain why?)

Source

Von Neumann and Morgenstern (1944) introduced the notion of expected utility. They also showed how it could be used to represent a very wide family of preference schemes.

Links

The axioms that imply expected utility maximization are introduced and explained in Chapter 5 of Kreps (1988), where a fairly elementary proof that the axioms imply EU maximization can also be found. A more general result is Herstein and Milnor (1953). Chapter 6 in Mas-Colell, Whinston, and Green (1995) also contains a very general theorem and proof, along with a discussion of the associated economics.

6.2 Risk aversion and risk neutrality

Suppose that you are offered a choice between your annual salary of \$40,000 for sure, and a chance of getting double that salary with probability $1/_2$ accompanied by an equal chance of winding up with zero. Most of us would choose the sure thing because the two options have the same expected monetary value of \$40,000, but the plunge from \$40,000 to zero is far more devastating than a drop from \$80,000 to \$40,000.

Example 6.4: A chance to double your salary

If the individual's utility-of-wealth function is $U(w) = 4\sqrt{w}$ then the EU of \$40,000 for sure is $4\sqrt{40,000} = 800$. The EU of a gamble that yields \$80,000 with probability $\frac{1}{2}$ or zero with the same probability is

$$\frac{1}{2} \times 4\sqrt{80,000 + \frac{1}{2} \times 4\sqrt{0}} = 565.7.$$

This individual prefers \$40,000 for sure because it yields a higher level of expected utility than the gamble.

We say that individuals are *risk averse* if they prefer having *w* for sure to an uncertain wealth level with an expected monetary value that is no higher than *w*. If asset *A* yields a high outcome *y* with probability 1/2 and a low outcome *x* with probability 1/2, and asset *B* yields $y + \delta$ with probability 1/2 and $x - \delta$ with probability 1/2 and $\delta > 0$, then *B* is unambiguously the riskier asset. The two have the same mean, but *B*'s payoffs have a wider spread than *A*'s.

DEFINITION: Risk aversion

Individuals are risk averse if for any payoff w they would always prefer w dollars for sure to an asset under which their wealth would have an expected monetary value of w but which would leave them with less than w with positive probability. In general, if assets A and B provide the individuals with the same expected monetary value of wealth, but A is unambiguously less risky, risk-averse individuals will always prefer A to B.

The individual would prefer θ for sure if the opportunity of obtaining a higher level of wealth brought with it the chance of winding up with a lower level of wealth *and* θ is at least as high as the average (expected) wealth associated with the gamble.

We conclude this section by showing that risk aversion is equivalent to diminishing marginal utility of wealth: We let MU_w denote the marginal utility of wealth at the level w. Suppose that we have $MU_x > MU_y$ for any choice of x and y such that x < y. In other words, the marginal utility of wealth is always positive, but it is smaller at higher levels of wealth. Consider an asset that pays $w - \delta$ with probability $\frac{1}{2}$ and $w + \delta$ with probability $\frac{1}{2}$, where $\delta > 0$. The EMV of this asset is w. Let L denote the potential utility loss, $U(w) - U(w - \delta)$, and let G denote the potential utility gain, $U(w + \delta) - U(w)$. Diminishing marginal utility of wealth implies that L is larger than G, because L involves a change in wealth at a lower level than G. Therefore, $\frac{1}{2}G < \frac{1}{2}L$, which can be written

$$\frac{1}{2}[U(w+\delta) - U(w)] < \frac{1}{2}[U(w) - U(w-\delta)].$$

Add $\frac{1}{2}U(w) + \frac{1}{2}U(w - \delta)$ to both sides of this inequality. We get

$$\frac{1}{2}[U(w+\delta) + U(w-\delta)] < \frac{1}{2}[U(w) + U(w)] = U(w)$$

Therefore, the EU of an asset that yields $w + \delta$ with probability $\frac{1}{2}$ and $w - \delta$ with probability $\frac{1}{2}$ is less than the EU of w dollars for sure. This is a consequence of diminishing marginal utility of wealth, and it holds for every wealth level w and every positive δ . Therefore, diminishing marginal utility of wealth implies risk aversion.

Figure 2.9 portrays the graph of a utility-of-wealth function with diminishing marginal utility. Diminishing MU causes the graph to bow upward, so that the straight line connecting any two points on the graph lies entirely below the graph, except at the endpoints. Let *x* be the wealth level at the left end with *y* denoting the wealth at the other end. The wealth level at the halfway point on the line is $\frac{1}{2}x + \frac{1}{2}y$. Note that this is the expected monetary value of an asset that leaves wealth at *x* with probability $\frac{1}{2}$ and at *y* with probability $\frac{1}{2}$. The coordinate on the vertical axis for the halfway point on the straight line is the average utility, $\frac{1}{2}U(x) + \frac{1}{2}U(y)$. Of course, this is the EU of an asset that leaves wealth at *x* with probability $\frac{1}{2}$ and at *y* with probability $\frac{1}{2}$. Because of the curvature of the graph of *U*, $U(\frac{1}{2}x + \frac{1}{2}y)$ is greater than $\frac{1}{2}U(x) + \frac{1}{2}U(y)$. In other words, the



Figure 2.9

utility from $\frac{1}{2} x + \frac{1}{2}y$ for sure is greater than the expected utility of an asset with an EMV of $\frac{1}{2} x + \frac{1}{2}y$, if the probability of receiving *x* is positive. Therefore, diminishing marginal utility of wealth implies risk aversion.

It is easy to show that risk aversion implies diminishing marginal utility of wealth. We just have to press "rewind." By definition of risk aversion, we have

$$\frac{1}{2}[U(w+\delta)+U(w-\delta)] < U(w)$$

for every choice of positive w and δ because the EMVs are equal. This implies

$$\frac{1}{2}[U(w+\delta) + U(w-\delta)] < \frac{1}{2}[U(w) + U(w)]$$

and hence $U(w + \delta) - U(w) < U(w) - U(w - \delta)$.

But $U(w + \delta) - U(w)$ is proportional to the marginal utility of wealth at the wealth level w, and $U(w) - U(w - \delta)$ is proportional to the marginal utility of wealth at the lower level $w - \delta$. Therefore, risk aversion implies that the marginal utility of wealth is lower at higher levels of wealth.

A *risk-neutral* individual is insensitive to the degree of risk. He or she will always chooses the asset with the higher EMV and will be indifferent between two assets with the same EMV, even if one has a much bigger spread between the two payoffs.

DEFINITION: Risk neutrality

Individuals are risk neutral if for any two assets *A* and *B* they prefer *A* to *B* if and only if *A* has a higher expected monetary value than *B*.

In general, to determine which asset an individual will choose we have to determine how each asset will affect an individual's wealth and then calculate the resulting expected utility. The asset that results in the highest EU from the final wealth portfolio will be the one that is chosen. However, in the case of a risk-neutral decision maker, we just have to calculate the expected monetary value of each asset. The one with the highest EMV will be chosen. That follows from the fact that

$$\begin{split} \label{eq:EMV} EMV \mbox{ (original wealth + new asset)} &= EMV \mbox{ (original wealth)} \\ &+ EMV \mbox{ (new asset)}. \end{split}$$

The asset with the highest EMV will be the asset that leads to the highest EMV of the new wealth portfolio.

We have shown—albeit informally—that risk aversion is equivalent to diminishing marginal utility of wealth. The next subsection uses elementary calculus to establish this rigorously.

∂ 6.3 Risk aversion and a negative second derivative

If $y \ge x$ then a risk-averse individual is one who would prefer an asset *A* that yielded *y* with probability $\frac{1}{2}$ and *x* with probability $\frac{1}{2}$ to an asset *B* that yielded $y + \delta$ with probability $\frac{1}{2}$ and $x - \delta$ with probability $\frac{1}{2}$, as long as δ is positive. Notice that *A* and *B* have the same EMV, $\frac{1}{2}x + \frac{1}{2}y$. For a risk-averse individual asset *B* will have a lower expected *utility* because there is a greater spread between the bad outcome and the good outcome. The definition of risk aversion leads directly to a proof that risk-averse individuals have utility functions with negative second derivatives. (For the converse, a utility-of-wealth function with a negative second derivative everywhere implies risk aversion; see the sketch of a proof employing Figure 2.9 in Section 6.2.)

Suppose that y > x and $\delta > 0$. By the definition of risk aversion $\frac{1}{2}U(y) + \frac{1}{2}U(x) > \frac{1}{2}U(y+\delta) + \frac{1}{2}U(x-\delta)$ and therefore $U(x) - U(x-\delta) > U(y+\delta) - U(y)$, which implies

$$\frac{U(x) - U(x - \delta)}{\delta} > \frac{U(y + \delta) - U(y)}{\delta}$$

because δ is positive. As δ approaches zero the left-hand side of this inequality approaches U'(x) and the right-hand side approaches U'(y). Therefore, we have proved that $U'(y) \leq U'(x)$ holds whenever y > x. But we can do better.

Suppose that U'(x) = U'(y) and x < y. Then U'(x) = U'(z) = U'(y) for $x \le z \le y$ because we have just proved that U' cannot increase as wealth increases. That is, $U'(x) \ge U'(z) \ge U'(y) = U'(x)$ implies U'(x) = U'(z) = U'(y). Consider the asset that yields $x^0 = x + \frac{1}{4}(y - x)$ with probability $\frac{1}{2}$ and $y^0 = y - \frac{1}{4}(y - x)$ with probability $\frac{1}{2}$. For a risk-averse individual this must have a higher expected utility than the asset that yields *x* and *y* each with probability $\frac{1}{2}$ because the latter has the same expected monetary value as the former but a lower bad outcome and a higher good outcome. Therefore, for $\delta = \frac{1}{4}(y - x)$ we have

$$\frac{1}{2}U(x) + \frac{1}{2}U(y) < \frac{1}{2}U(x+\delta) + \frac{1}{2}U(y-\delta)$$

6. Decision Making Under Uncertainty

and thus

$$U(y) - U(y - \delta) < U(x + \delta) - U(x).$$

But this is inconsistent with U' being constant on the range of values between x and y. The inconsistency arises from the fact that constant U' implies $U(y) - U(y - \delta) = \delta U'(x) = U(x + \delta) - U(x)$. We must conclude that U'(x) > U'(y) actually holds for a risk-averse person whenever x < y.

The risk-aversion theorem

An individual with a twice differentiable utility-of-wealth function *U* is risk averse if and only if the second derivative of *U* is negative at every point.

Because a risk-averse individual gets higher expected utility from asset *A* than asset *B* if they have the same EMV but *A* is unambiguously less risky, it is clear that a risk-averse individual will pay a premium—large or small, depending on preference—to avoid risk. This is one of the foundations of the insurance industry. (The other is the law of large numbers.) In fact, the prominence of insurance in almost all aspects of our economy is strong evidence for the prevalence of risk aversion. Individuals have even been known to buy insurance against the possibility that an existing insurance opportunity will disappear.

Because a risk-neutral person is indifferent between two assets with the same EMV, we have

$$\frac{1}{2}U(x) + \frac{1}{2}U(y) = \frac{1}{2}U(x+\delta) + \frac{1}{2}U(y-\delta)$$

for all values of *x*, *y*, and δ . Therefore,

$$\frac{U(y) - U(y - \delta)}{\delta} = \frac{U(x + \delta) - U(x)}{\delta}$$

for all $\delta \neq 0$, and thus U'(x) = U'(y) for all x and y. If the first derivative is constant the function U must be of the form $U(x) = \alpha x + \beta$. If utility is increasing in wealth we must have $\alpha > 0$. Therefore, maximizing expected utility is equivalent to maximizing expected monetary value in the case of a risk-neutral individual.

6.4 The market opportunity line

In the absence of uncertainty the individual's consumption plan (x, y) must be chosen from a budget line determined by equating expenditure and income. When the individual chooses in an uncertain environment the market also determines the combinations of x and y from which the decision maker is able to choose. Although these pairs (x, y) can't always be represented by a straight line, almost all of the examples in this book are elementary enough to be so depicted. Hence, for convenience we refer to the *market opportunity line*.

DEFINITION: Market opportunity line

The market opportunity line is the set of pairs (x, y) from which the individual is allowed to choose.

Example 6.5: An investment

An individual with current wealth of \$160 has an opportunity to invest in a project that will be successful with probability 0.7, in which case the individual will receive \$4 for every dollar invested. There is a probability 0.3 that the project will fail and the individual will get back only twenty cents on the dollar. Let *C* be the amount invested. If the project fails the individual's wealth will be 160 - C + 0.2C = 160 - 0.8C. If the project were to succeed, the individual's wealth will be 160 - C + 4C = 160 + 3C. Therefore, x = 160 - 0.8C and y = 160 + 3C. From the first of these equations we have C = 200 - 1.25x. Now substitute the right-hand side of this equation for *C* in y = 160 + 3C. We get

$$y = 160 + 3(200 - 1.25x) = 760 - 3.75x.$$

Finally, the market opportunity line can be expressed as 3.75x + y = 760.

Example 6.6: Insurance

An individual with current wealth of \$100 will have 70% of it stolen with a probability of 0.3. He can purchase insurance for forty cents per dollar of coverage. If *C* is the amount of coverage purchased, then the individual's wealth will be x = 30 + C - 0.4C if there is a burglary, and y = 100 - 0.4C if there is no burglary. Because x = 30 + 0.6C we have C = (1/0.6)x - 50. Therefore,

$$y = 100 - 0.4\left(\frac{1}{0.6}x - 50\right) = 120 - \frac{2}{3}x.$$

The market opportunity line is (2/3)x + y = 120. Note that this equation is satisfied when the individual buys no insurance, in which case x = 30 and y = 100.

When π is the probability of receiving x and $1 - \pi$ is the probability of receiving y, *and* the market opportunity line has the form $\pi x + (1 - \pi)y = \theta$, then we say that the odds are *fair*.

DEFINITION: Fair odds line

A fair odds line is any market opportunity line that can be written in the form

$$\pi x + (1 - \pi)y = \theta,$$

where π is the probability of actually receiving x, $1 - \pi$ is the probability of actually getting y, and θ is some constant.

6. Decision Making Under Uncertainty

If $2\pi x + 2(1 - \pi)y = \theta$ we still have fair odds because we can divide both sides of the equation by 2 to get $\pi x + (1 - \pi)y = \theta/2$. In fact, if the market opportunity equation is $p_1x + p_2y = \theta$ then we have fair odds if and only if $p_1/p_2 = \pi/(1 - \pi)$. In that case, multiply both sides of $p_1x + p_2y = \theta$ by π/p_1 and use the fact that $p_2\pi/p_1 = 1 - \pi$ to prove that the equation can be expressed in the form $\pi x + (1 - \pi)y = \theta\pi/p_1$.

Note that we do not have fair odds in Example 6.5 because $p_1/p_2 = 3.75/1 = 3.75$ but $\pi/(1 - \pi) = 0.3/0.7 = 0.429$. Nor do we have fair odds in Example 6.6 because $p_1/p_2 = 2/3$ but $\pi/(1 - \pi) = 0.3/0.7 = 0.429$.

Example 6.7: Insurance with fair odds

An individual with current wealth of \$100 will have 70% of it stolen with probability 0.2. Every dollar of premium paid to the insurance company results in \$5 being received in case of an accident. If *P* is the premium paid, then x = 30 + 5P - P and y = 100 - P. This last equation yields P = 100 - y and thus x = 30 + 4P = 30 + 4(100 - y). Therefore, the market opportunity line is x + 4y = 430. The ratio of probabilities is 0.2/0.8, which is also equal to $\frac{1}{4}$, the ratio of the coefficients of the market opportunity line. The individual faces fair odds.

The next two sections relax the assumption that there are only two possible outcomes. In fact, we now assume an infinite number of possibilities.

6.5 The uniform probability distribution

Suppose that the random variable *x* could turn out to be any of the real numbers between zero and one inclusive. Assume further that each value is as likely as any other. This characterizes the *uniform probability distribution*. Because there are an infinite number of values between 0 and 1, the most useful way of applying the uniform distribution is in terms of the probability that *x* is between 0 and β , for a given value of β (not exceeding 1). For the uniform probability distribution, the probability that $0 \le x \le \beta$ is β itself. The probability that $0 \le x < \beta$ is also β . Now, consider the uniform probability distribution on the interval [*a*, *b*], which is the set of *x* such that $a \le x \le b$.

DEFINITION: The uniform probability distribution

If *x* is a random draw from the interval [*a*, *b*] then the probability that $a \le x \le \beta$ is $(x - a)/(\beta - a)$ if *x* is governed by the uniform probability distribution on [*a*, *b*].

For instance, if you are submitting a bid in a sealed-bid auction with one other participant, and you view your opponent's bid as a random draw from the uniform probability distribution on the interval [0, 100] then the probability that you will win with your bid of *x* is the probability that the other person's bid is below *x*, which is x/100.

122 Basic Models and Tools

$\partial 6.6$ The continuum case in general

Suppose that the random variable *x* is drawn from the interval [*a*, *b*] of real numbers, where [*a*, *b*] denotes the set of all numbers *x* such that $a \le x \le b$. We use a probability density function f(x) to determine the probability that *x* belongs to the subinterval [α , β]. The function *f* is integrable, and given real numbers α and β such that $a \le \alpha < \beta \le b$,

$$\operatorname{Prob}[\alpha < x < \beta] = \int_{\alpha}^{\beta} f(x) \, dx$$

Of course, $\operatorname{Prob}[\alpha < x < \beta]$ denotes the probability that $x > \alpha$ and $x < \beta$ both hold. In the continuum case, $\operatorname{Prob}[\alpha < x < \beta] = \operatorname{Prob}[\alpha < x \le \beta] = \operatorname{Prob}[\alpha \le x < \beta]$.

Given the utility-of-wealth function *U*, the expected utility of the random variable *x* over the interval (α , β) is defined by

$$\int_{\alpha}^{\beta} U(x) f(x) \, dx.$$

In the case of the uniform probability distribution on [a, b], f(x) = 1/(b - a) for all *x*. Hence

$$\operatorname{Prob}[\alpha < x < \beta] = \int_{\alpha}^{\beta} \frac{1}{b-a} \, dx = \frac{\beta}{b-a} - \frac{\alpha}{b-a} = \frac{\beta-\alpha}{b-a}$$

Therefore, if a = 0 then x is drawn from the uniform probability distribution on the set of numbers between 0 and b, in which case the probability that xis less than β is β/b . Consequently, if x is drawn from the uniform probability distribution on the set of numbers between 0 and 1 then the probability that xis less than β is β itself. Finally, for the uniform probability distribution on [a, b], we have f(x) = 1/(b-a), and thus the expected utility of the random variable x over the interval (α, β) with respect to the utility function U is

$$\int_{\alpha}^{\beta} \frac{U(x)}{b-a} \, dx.$$

Link

See Sheffrin (1993, p. 51) for an instance of the purchase of insurance against the possibility that an existing insurance opportunity will disappear.

Problem set

- 1. The individual's utility-of-wealth function is $U(w) = \sqrt{w}$ and current wealth is \$10,000. Is this individual risk averse? What is the maximum premium that this individual would pay to avoid a loss of \$1900 that occurs with probability $\frac{1}{2}$? Why is this maximum premium not equal to half of the loss?
- **2.** An individual has a utility-of-wealth function $U(w) = \ln(w + 1)$ and a current wealth of \$20. Is this individual risk averse? How much of this wealth will this person use to purchase an asset that yields zero with probability $\frac{1}{2}$, and with probability $\frac{1}{2}$ returns \$4 for every dollar invested? (When the asset pays off, a \$1 investment returns \$3 net of the original outlay.)

6. Decision Making Under Uncertainty

- **3.** The utility-of-wealth function is $U(w) = \ln(w+1)$ and the individual's current wealth is θ . As a function of π , r, and θ how much of this wealth will the individual invest in a project that yields zero with probability π , and with probability 1π pays rC dollars to an investor who has sunk C dollars into the project?
- **4.** For $U(w) = \sqrt{w}$ prove that $U(\frac{1}{2}x + \frac{1}{2}y) > \frac{1}{2}U(x) + \frac{1}{2}U(y)$ for $x \neq y$.
- **5.** For $U(w) = \sqrt{w}$ prove that $U(\pi x + (1 \pi)y) > \pi U(x) + (1 \pi)U(y)$ for $0 < \pi < 1$ and $x \neq y$.
- **6.** Diane has a utility-of-wealth function $U(w) = \sqrt{w}$ and a current wealth of \$2000.
 - **A.** Will Diane invest in a scheme that requires an initial capital outlay of \$2000 and returns nothing with probability of 1/2 (i.e., the initial outlay is lost and there is no revenue) and returns \$6000 with probability of 1/2?
 - **B.** Cathy is identical to Diane in every respect. If Diane and Cathy can share the investment (this is called *risk spreading*) will they do so? In this case sharing means that each puts up \$1000 and they split the proceeds of the investment.
- **7.** Leo, who has a utility-of-wealth function $U(w) = \ln(w + 20)$, has \$100 of income before tax and is taxed at a rate of 40% of earned income. If he is caught underreporting his income he will have to pay the taxes owed and in addition will pay a fine of \$1 for every dollar of income he failed to report. How much income will he conceal (i.e., fail to report) if the probability of being caught is 0.2? (Let *C* denote the amount of income concealed.)
- **8.** Teri, who has a utility-of-wealth function $U(w) = \ln(w + 100)$, would have an *after-tax* income of \$100 *if* she reported all her income. She is taxed at a rate of 50% of earned income (just to keep the calculations simple). If she is caught underreporting her income she will have to pay the taxes owed, of course, but in addition she will pay a fine of *F* dollars for every dollar of income she failed to report.
 - **A.** How much income will she conceal (i.e., fail to report) if F = 2 and the probability of being caught is 0.10? Let *C* denote the amount of income concealed.
 - **B.** Determine *C* as a function of the fine *F* and the probability of being caught ρ . Show that *C* falls when either *F* or ρ increases.
- **9.** Spencer, who has a utility-of-wealth function $U(w) = \sqrt{w}$, has an initial wealth of \$52. He has an opportunity to invest in a project that will cause him to lose his capital with probability 0.75, but with probability 0.25 will provide a net return of \$4 for every dollar of capital he puts up. How much will he invest? (Let *A* denote the amount invested—i.e., the amount of capital he puts up. He loses *A* if the project fails, but if it succeeds it will pay him \$5 gross for every dollar invested.)

- **10.** Determine the market opportunity line for questions 2, 3, 7, 8A, and 9.
- **11.** A standard measure of the degree of risk aversion at wealth level *w* is -U''(w)/U'(w). (It is called the *Arrow-Pratt* measure.)
 - **A.** Show that if $U(w) = w^{\alpha}$, where α is a positive constant less than one, then risk aversion decreases as *w* increases.
 - **B.** Show that if $U(w) = K e^{-\alpha w}$, where $\alpha > 0$, then degree of risk aversion is independent of the level of wealth.

(The idea of the Arrow-Pratt measure is that the faster the marginal utility of wealth declines, the more risk averse the individual is. The first derivative of U is used as the denominator because we would otherwise have a change in the measure if we replaced U with λU for $\lambda > 0$. That would be undesirable because the underlying preferences can be represented by either U or λU : If the EU of asset A exceeds that of asset B according to U then the EU of A will be greater than the EU of B according to λU , and vice versa.)

- **12.** The host of the television game show *Wheel of Fortune* tells contestants that it would be irrational to risk \$15,000 for a \$12,000 car because the probability of losing everything is one-half. The "asset" in this case yields zero with probability $\frac{1}{2}$ and \$27,000 with probability $\frac{1}{2}$. If the individual stands pat and takes the safe asset, he or she will have \$15,000 for sure.
 - A. Explain why any risk-averse individual would stand pat.
 - **B.** Find a utility-of-wealth function *U* such that the EU of the risky asset is greater than the expected utility of \$15,000 for sure. Of course, *U* will not have diminishing marginal utility.

(I have oversimplified the position in which the contestants find themselves, but not in ways that vitiate the point that the gamble is not irrational for *some* people.)

7 INSURANCE

In this section we work out the equilibrium of a competitive insurance market when the probability of an accident is the same for any two individuals and no individual has an opportunity to reduce the probability of an accident by devoting effort to preventive care. These two extreme assumptions establish a benchmark case. We relax the latter in Section 9 of Chapter 3 when we take account of the fact that if everyone takes preventive care there will be far fewer accidents and hence a higher level of expected utility for everyone. However, with insurance coverage no *individual* has incentive to invest in preventive care. Section 7 of Chapter 5 examines a competitive insurance market when different individuals have different probabilities of an accident. That information is hidden from the insurance companies and that can also result in inefficiency. But in this chapter everyone has the same probability of an accident, and that probability is independent of any choice made by any individual. We begin by showing that a risk-averse decision maker will set x = y when confronted with fair odds.

7. Insurance

7.1 The complete insurance theorem

When risk-averse individuals choose under fair odds they will set x = y. This will allow us to identify their decision quickly: If the market opportunity line is $\pi x + (1 - \pi)y = \theta$, then x = y implies $\pi x + (1 - \pi)x = \theta$, so $x = \theta = y$.

Complete insurance theorem

Arisk-averse individual will maximize expected utility by setting x = y whenever the odds are fair.

We say that we have *compete insurance* whenever x = y because the individual's wealth is the same in either event. Without complete insurance, the low payoff (which occurs with probability π) will be lower than the payoff that occurs with probability $1 - \pi$.

7.2 Sketch of the proof of the complete insurance theorem

We test the bundle (x, y) to see if it maximizes expected utility subject to the fact that *x* and *y* must be on the opportunity line. With *x* as the initial level of wealth, let MU_X denote the increase in *U* per unit increase in wealth in the "accident" state. Similarly, starting from the level y, let MU_Y denote the increase in *U* per unit increase in wealth in the "no accident" state.

Under fair odds the individual's choice of *x* and *y* must satisfy $\pi x + (1 - \pi)y = \theta$. We can solve this for *y*. We get

$$y = \frac{\theta}{1-\pi} - \frac{\pi}{1-\pi}x.$$

Therefore, if we change *x* and *y* by Δx and Δy , respectively, we must have $\Delta y = -[\pi/(1-\pi)]\Delta x$. The resulting change in EU will be

$$\Delta EU = \pi \times MU_X \times \Delta x + (1 - \pi) \times MU_Y \times \Delta y.$$

Replace Δy in this last expression by $-[\pi/(1-\pi)]\Delta x$. We get

$$\Delta EU = \pi \times MU_X \times \Delta x + (1 - \pi) \times MU_Y \times -\frac{\pi}{1 - \pi} \Delta x$$
$$= \pi \Delta x (MU_X - MU_Y).$$

Therefore, a fair odds market opportunity line implies that $\Delta EU = \pi \Delta x (MU_X - MU_Y)$. If $MU_X > MU_Y$ then we can increase ΔEU by setting $\Delta x > 0$, in which case both $\pi \Delta x$ and $MU_X - MU_Y$ will be positive, and thus ΔEU will also be positive. Therefore, if $MU_X > MU_Y$ we have not maximized EU. If $MU_X < MU_Y$ then we can increase ΔEU by setting $\Delta x < 0$, in which case both $\pi \Delta x$ and $MU_X - MU_Y$ we have not maximized EU. If $MU_X < MU_Y$ then we can increase ΔEU by setting $\Delta x < 0$, in which case both $\pi \Delta x$ and $MU_X - MU_Y$ will be negative, and thus ΔEU will be positive. Therefore, EU is not at its maximum if $MU_X \neq MU_Y$. We have established that MU_X must equal MU_Y at the values of x and y that maximize EU. Diminishing marginal utility of wealth implies that $MU_X > MU_Y$ if x < y and $MU_X < MU_Y$ if x > y. Therefore, if $MU_X = MU_Y$ holds if and only if x = y. Consequently, maximization of EU

126 Basic Models and Tools

subject to fair odds implies that x = y. When we set x = y in the opportunity line $\pi x + (1 - \pi)y = \theta$ we get $x = \theta = y$.

∂ 7.3 Calculus proof of the complete insurance theorem

We want to maximize $EU = \pi U(x) + (1 - \pi)U(y)$ subject to $\pi x + (1 - \pi)y = \theta$. First we solve the market opportunity equation for *y*. We get $y = \theta/(1 - \pi) - [\pi/(1 - \pi)]x$. Then $dy/dx = -\pi/(1 - \pi)$. The variable *y* appears in the expression for EU, but we will treat it as a function of *x*. Then we can set

$$V(x) = \pi U(x) + (1 - \pi) U(y)$$

and use the chain rule to find the first derivative of V. We have

$$\frac{dV}{dx} = \pi U'(x) + (1 - \pi)U'(y) \times \frac{dy}{dx} = \pi U'(x) + (1 - \pi)U'(y) \times -\frac{\pi}{1 - \pi}$$
$$= \pi U'(x) - \pi U'(y).$$

The first derivative must be equal to 0 at a maximum of *V*. But V' = 0 and $V' = \pi U'(x) - \pi U'(y)$ imply U'(x) = U'(y). Finally, because U'' < 0 at every point (by the risk-aversion assumption) we can have U'(x) = U'(y) if and only if x = y. (If U'' < 0 at every point then the first derivative of *U* falls as the wealth level increases. Therefore, if x < y then U'(x) > U'(y), and if x > y we have U'(x) < U'(y).)

To confirm that V' = 0 takes us to a *maximum*, let's compute the second derivative of *V*. With fair odds the first derivative of *V* is $\pi U'(x) - \pi U'(y)$. Then

$$V'' = \pi U''(x) - \pi U''(y) \times \frac{dy}{dx} = \pi U''(x) - \pi U''(y) \times -\frac{\pi}{1 - \pi}$$
$$= \pi U''(x) + \frac{\pi^2}{1 - \pi} U''(y).$$

Because π and $(1 - \pi)$ are both positive and U'' is negative at every point (by the risk-aversion assumption) we have V''(x) < 0 for all x. Therefore, V' = 0 at the point where V achieves its unique global maximum.

Example 7.1: A specific utility-of-wealth function

The utility-of-wealth function is $U(w) = \ln(w + 1)$. The individual will receive x with probability $\frac{1}{4}$ and y with probability $\frac{3}{4}$. We are told that $\frac{1}{4}x + \frac{1}{4}y = 10$ is this individual's market opportunity equation. Therefore, he or she faces fair odds. Now, maximize $\frac{1}{4}\ln(x + 1) + \frac{3}{4}\ln(y + 1)$ subject to $\frac{1}{4}x + \frac{3}{4}y = 10$. From the market opportunity line we have $y = \frac{40}{3} - \frac{1}{3}x$. Then we wish to maximize

$$V(x) = \frac{1}{4}\ln(x+1) + \frac{3}{4}\ln\left(\frac{40}{3} - \frac{1}{3}x + 1\right).$$

We have

$$V'(x) = \frac{1}{4}(x+1)^{-1} + \frac{3}{4}(y+1)^{-1} \times \left(-\frac{1}{3}\right) = \frac{1}{4}(x+1)^{-1} - \frac{1}{4}(y+1)^{-1}.$$

7. Insurance

Note that V'' < 0 at all points. Therefore, if V'(x) = 0 yields nonnegative values of x and y the equation V'(x) = 0 will characterize the solution to our problem. But V'(x) = 0 gives us $\frac{1}{4}(x+1)^{-1} = \frac{1}{4}(y+1)^{-1}$. Multiply both sides by 4(x+1)(y+1). This yields y + 1 = x + 1, and hence x = y. Substituting x for y in the market opportunity equation yields x = 10. Thus x = 10 = y and we have complete insurance.

The next three examples demonstrate that different risk-averse individuals will purchase different levels of insurance coverage if the odds are not fair, but the same individuals will choose complete insurance when the odds are fair.

Example 7.2: Insurance without fair odds

Rosie's utility-of-wealth function is $U(w) = \sqrt{w}$. Her current wealth is \$100, but with probability 0.3 an accident will reduce her wealth to \$30, as summarized by Table 2.3. Suppose that a dollar of insurance coverage costs forty cents. Therefore, if Rosie has an accident and has *C* dollars of coverage her wealth will be

$$x = 30 + C - 0.4C = 30 + 0.6C.$$

(She gets a claim check for *C* dollars but still has to pay her premium in a year when she has an accident.) With *C* dollars of insurance she will have wealth of

$$y = 100 - 0.4C$$

if she doesn't have an accident. Let's calculate the market opportunity line: We have y = 100 - 0.4C and thus C = 250 - 2.5y. Now substitute 250 - 2.5y for *C* in the expression x = 30 + 0.6C. We get $x = 30 + 0.6 \times (250 - 2.5y) = 30 + 150 - 1.5y$. That is, x = 180 - 1.5y. Then

$$x + 1.5y = 180$$

is the equation of the market opportunity line. We do not have fair odds because the ratio of the *x* coefficient to the *y* coefficient is 2/3 but the ratio of probabilities is 3/7, which is smaller than 2/3.

Table 2.3

State	Probability	Wealth
No accident	0.7	100
Accident	0.3	30

Let's determine how much insurance Rosie will purchase. We want to maximize

$$EU = 0.3\sqrt{x} + 0.7\sqrt{y} = 0.3\sqrt{30 + 0.6C} + 0.7\sqrt{100 - 0.4C},$$

128 Basic Models and Tools

a function of C. The first derivative is

$$\frac{0.3 \times 0.6}{2\sqrt{30 + 0.6C}} + \frac{0.7 \times -0.4}{2\sqrt{100 - 0.4C}}$$

We maximize EU by setting the first derivative equal to 0. (Confirm that the second derivative is negative for all $C \ge 0$.) After setting the first derivative equal to 0 and multiplying both sides of the equation by 200 we get

$$\frac{18}{\sqrt{30+0.6C}} - \frac{28}{\sqrt{100-0.4C}} = 0.$$

This implies $9\sqrt{100 - 0.4C} = 14\sqrt{30 + 0.6C}$. We square both sides and solve for *C*, yielding $C^* = 14.8$. If Rosie has an accident her wealth will be $x^* = 30 + 0.6 \times 14.8 = 38.88$. If there is no accident her wealth will be $y^* = 100 - 0.4 \times 14.8 = 94.08$. With insurance, Rosie's EU will be $0.3\sqrt{38.88} + 0.7\sqrt{94.08} = 8.66$. Without insurance her EU is $0.3\sqrt{30} + 0.7\sqrt{100} = 8.64$.

Let's see what a different individual will choose with the same market opportunity.

Example 7.3: The same odds but a different utility-of-wealth function

Soren's utility-of-wealth function is $U(w) = \ln(w + 1)$. Except for the utility-ofwealth function, the data are the same as for Example 7.2: Soren's wealth will be \$100 with probability 0.7 and \$30 with probability 0.3. A dollar of insurance coverage costs forty cents. To determine how much insurance Soren will purchase we maximize

$$EU = 0.3 \ln(x+1) + 0.7 \ln(y+1) = 0.3 \ln(30+0.6C+1) + 0.7 \ln(100-0.4C+1).$$

The first derivative is

$$\frac{0.3 \times 0.6}{31 + 0.6C} + \frac{0.7 \times -0.4}{101 - 0.4C}.$$

(Confirm that the second derivative is negative for all $C \ge 0$.) When we set the first derivative equal to 0 and solve for *C* we get $C^* = 39.58$. This is substantially more coverage than Rosie would purchase under the same terms. Evidently, Soren is more risk averse than Rosie. If Soren has an accident his wealth will be $x^* = 30 + 0.6 \times 39.58 = 53.75$. If there is no accident his wealth will be $y^* = 100 - 0.4 \times 39.58 = 84.17$. With insurance, Soren's EU will be $0.3 \ln(53.75 + 1) + 0.7 \ln(84.17 + 1) = 4.31$. Without insurance his EU is $0.3 \ln(30 + 1) + 0.7 \ln(100 + 1) = 4.26$.

Both individuals buy some insurance, even though it lowers the EMV of their wealth. With x = 30 + 0.6C and y = 100 - 0.4C we have EMV = 0.3(30 + 0.6C) + 0.7(100 - 0.4C) = 79 - 0.1C. Then without insurance we have

7. Insurance

C = 0 and EMV = 79. But if *C* is positive then EMV is less than 79. Risk-averse individuals typically buy some insurance even when it lowers the EMV of their wealth because the reduction in exposure to risk more than makes up for the loss in EU resulting from the loss in EMV. (Of course, under fair odds the EMV of wealth is the same with insurance as without.)

Now, suppose that the cost of a dollar of insurance coverage falls to thirty cents. This leads to fair odds. Let's check. We now have

$$x = 30 + C - 0.3C = 30 + 0.7C$$
 and $y = 100 - 0.3C$.

From the second equation we get C = 100/0.3 - y/0.3. When we substitute the right-hand side for *C* in the equation x = 30 + 0.7C we get

$$x = 30 + \frac{70}{0.3} - \frac{0.7y}{0.3}.$$

Multiply both sides by 0.3, resulting in 0.3x = 79 - 0.7y, or 0.3x + 0.7y = 79. We do indeed have fair odds.

Example 7.4: The chosen coverage under fair odds

As shown in Table 2.4, the individual faces the same risk as in the previous two examples. A dollar of insurance coverage now costs thirty cents. Rosie will now maximize

$$EU = 0.3\sqrt{x} + 0.7\sqrt{y} = 0.3\sqrt{30 + 0.7C} + 0.7\sqrt{100 - 0.3C}.$$

The first derivative is

$$\frac{0.3 \times 0.7}{2\sqrt{30 + 0.7C}} + \frac{0.7 \times -0.3}{2\sqrt{100 - 0.3C}}.$$

Note that the derivative is 0 when C = 70. Rosie chooses \$70 of coverage. In that case, $x = 30 + 0.7 \times 70 = 79$ and $y = 100 - 0.3 \times 70 = 79$. Rosie chooses complete insurance. Note also that the expected value of her wealth

Table 2.4

State	Probability	Wealth
No accident	0.7	100
Accident	0.3	30

is \$79 with insurance, and it is also \$79 without insurance $(0.3 \times 30 + 0.7 \times 100 = 79)$. With complete insurance Rosie's EU is $0.3\sqrt{79} + 0.7\sqrt{79} = 8.89$.

To determine how much insurance Soren will purchase under fair odds we maximize

$$EU = 0.3 \ln(x+1) + 0.7 \ln(y+1) = 0.3 \ln(30 + 0.7C + 1) + 0.7 \ln(100 - 0.3C + 1).$$

130 Basic Models and Tools

The first derivative is

$$\frac{0.3 \times 0.7}{31 + 0.7C} + \frac{0.7 \times -0.3}{101 - 0.3C},$$

which will be zero when C = 70. The two risk-averse individuals make the same choice when the odds are fair. With complete insurance, Soren's EU is $0.3 \ln(79 + 1) + 0.7 \ln(79 + 1) = 4.38$.

7.4 Competitive insurance markets

Assume a large number *n* of individuals and that an individual's wealth will fall from *z* to *a* with probability π . Each has the utility-of-wealth function U(w), and we assume diminishing marginal utility of wealth because we assume that each person is risk averse. Because there are *n* identical individuals, we can treat the *n* experiences as the result of *n* statistically independent experiments in which the probability of failure is π in each case. The law of large numbers assures us that the actual number of failures will be very close to the expected number πn with very high probability. In that case we save ourselves the trouble of saying that our results hold with probability extremely close to one by claiming that there will be exactly πn accidents. In that case there will be πn individuals with wealth *a* and $n - \pi n$ individuals who do not suffer an accident and hence whose wealth is *z*. Therefore, the community's actual wealth will be $\pi na + (n - \pi n)z$.

Suppose that an individual pays a premium of p per dollar of *net* coverage. That is, if coverage of c is purchased then the individual's wealth will be a + c with probability π and z - pc with probability $(1 - \pi)$. (Note that the individual pays the premium pc whether or not there is an accident.) In short, x = a + c and y = z - pc. Because the individual ultimately cares about x and y we let the terms of the policy be implicit and refer to an insurance policy as a pair (x, y). We can always use (x, y) to derive the terms of the policy because c = x - a and p = (z - y)/c. In other words, given the pair (x, y), the total premium is z - y (the difference between wealth if there is no accident and no insurance and wealth if there is insurance but no accident). The net coverage (the claim check minus the premium) is the difference between wealth if there is an accident with insurance and wealth if there is an accident but no insurance. The premium per dollar of net coverage is, of course, the total premium divided by net coverage.

DEFINITION: The simple model

Without insurance, all individuals have wealth *a* if they have an accident and *z* otherwise. With insurance, an individual's wealth is *x* in case of an accident and *y* otherwise.

Suppose that everyone buys the same policy. Then everyone will have the same pair (x, y), with x denoting wealth in case of an accident and y representing wealth if there is no accident. Because "exactly" πn individuals have an accident
7. Insurance

(and thus $n - \pi n$ individuals do not have an accident) when everyone buys the policy (*x*, *y*) the community's wealth will be

$$\pi n \times x + (n - \pi n) \times y$$

But we have also calculated total community wealth as $\pi n \times a + (n - \pi n) \times z$. The insurance industry cannot create wealth; it can only redistribute it from those who have not suffered an accident to those who have. Therefore,

$$\pi nx + (n - \pi n)y \le \pi na + (n - \pi n)z.$$

Now suppose that we actually have $\pi nx + (n - \pi n)y < \pi na + (n - \pi n)z$. This means that insurance companies have taken in more money in premiums than they paid out in claims. This is not consistent with equilibrium *if* we assume that there is vigorous competition among insurance providers and that administration costs are zero. (The latter is adopted just for convenience.) Then when more money is collected in premiums that is paid out in claims then insurance companies are making a positive economic profit. Some insurance companies will lower their premiums to attract customers away from other companies and thus increase their profit. Hence, we can't be at equilibrium if $\pi nx + (n - \pi n)y < \pi na + (n - \pi n)z$. Because we have already established that $\pi nx + (n - \pi n)y \leq \pi na + (n - \pi n)z$ must hold, we conclude that we have

$$\pi nx + (n - \pi n)y = \pi na + (n - \pi n)z.$$

Now, divide both sides of this equation by *n*, yielding

$$\pi x + (1 - \pi)y = \pi a + (1 - \pi)z.$$

Because this equation embodies fair odds, and individuals are risk averse, expected utility is maximized, subject to this equality, by setting x = y. Now, $\pi x + (1 - \pi)x = x$. Therefore, when we replace y by x in the equation $\pi x + (1 - \pi)y = \pi a + (1 - \pi)z$ we get $x = \pi a + (1 - \pi)z = y$. This is the equilibrium of the competitive insurance industry. Any other policy on the fair odds line would give individuals less expected utility, and thus would leave room for an insurer to offer a policy resulting in $x = \pi a + (1 - \pi)z$ and y just slightly less than that. This would be preferred by consumers (as long as y is not too far below $\pi a + (1 - \pi)z$) and would be profitable for the insurer. That tells us that the original situation could not have been an equilibrium. And if a policy does not leave the consumer on the fair odds line it is either not feasible or it fails to distribute all of the money collected in claims, and either case is inconsistent with equilibrium.

Example 7.5: A simple case

The probability of an accident is $\pi = \frac{1}{4}$ for each person. Then $1 - \pi = \frac{3}{4}$, the probability that the individual does not have an accident. Each person's wealth is a = 4 if there is an accident and no insurance, and z = 12 if there is no accident and no insurance. In a competitive insurance market the individual's opportunity equation is

132 Basic Models and Tools

Because complete insurance under competitive conditions yields x = y, we have x = 10 = y. The insurance premium is *z* minus wealth when there is insurance but no accident. Specifically, the premium is 12 - 10 = 2. In case of an accident individuals would receive claim checks for \$8, and their wealth would be a +claim check – premium = 4 + 8 - 2 = 10.

We would expect that, even with complete insurance, the individual's wealth would be lower if the probability of an accident were higher. The next example illustrates.

Example 7.6: A higher probability of an accident

The probability of an accident is 1/2, which is also the probability that there is no accident. We again assume that wealth is a = 4 if there is an accident and no insurance and z = 12 if there is no accident and no insurance. The competitive equilibrium per capita opportunity equation is

$$\frac{1}{2}x + \frac{1}{2}y = \frac{1}{2} \times 4 + \frac{1}{2} \times 12 = 8.$$

Then x = y gives us x = 8 = y. The insurance premium is now 12 - 8 = 4, and if there is an accident individuals receive claim checks for \$8, and their wealth would be 4 + 8 - 4 = 8.

7.5 Efficiency of competitive insurance markets with full information

To establish a benchmark case, we are assuming away all hidden action and hidden characteristic problems. Specifically, we suppose that no individual can reduce the probability of an accident by devoting effort to prevention— no hidden action—and also that the probability of an accident is the same for everyone—no hidden characteristics. We continue to assume a large number n of risk-averse individuals, each with a probability π of his wealth declining from z to a. We have seen that the competitive equilibrium results in $x = y = \pi a + (1 - \pi)z$ for each individual. We now prove that this outcome is efficient. We do so by showing that any outcome that gives everyone higher expected utility than the competitive equilibrium is not feasible because it requires more total wealth than the community has available.

Suppose that each individual's expected utility is higher at (x', y') than it is at the competitive equilibrium. Because the latter maximizes individual EU on the market opportunity line $\pi x + (1 - \pi)y = \pi a + (1 - \pi)z$, an allocation delivering even higher EU must be above that line. Therefore,

$$\pi x' + (1 - \pi)y' > \pi a + (1 - \pi)z.$$

If we multiply both sides of that inequality by *n* we get $n\pi x' + (n - n\pi)y' > n\pi a + (n - n\pi)z$. Because $n\pi$ individuals will have an accident, $n\pi x' + (n - n\pi)y'$ is total community wealth as a result of giving each person (x', y'), and $n\pi a + (n - n\pi)z$ is actual community wealth. Then $\pi x' + (1 - \pi)y' > \pi a + (1 - \pi)z$ tells us that we *can't* give everyone (x', y') because it requires more wealth to

7. Insurance

be distributed than the community actually has. (Insurance, whether it is provided by private firms or the government, cannot create wealth; it can only redistribute it.) Therefore, there is no feasible outcome that gives everyone higher expected utility than the competitive equilibrium.

The preceding argument is incomplete because it ignores the possibility of giving individuals different levels of insurance.

Example 7.7: A highly unlikely scenario

Suppose that a = 40 and z = 60 with $\pi = \frac{1}{2}$. Consider two policies, A = (46, 56) and B = (50, 54). If everyone got policy A the outcome would not be feasible because the expected value of per capita wealth with insurance would be $\frac{1}{2} \times 46 + \frac{1}{2} \times 56 = 51$, which exceeds $50 = \frac{1}{2} \times 40 + \frac{1}{2} \times 60$, the expected value of per capita wealth without insurance. Similarly, B is not feasible if everyone gets B because its expected value is $\frac{1}{2} \times 50 + \frac{1}{2} \times 54 = 52$. Suppose, however, that half the people get A and the other half get B. The "actual" number of accidents will be $\frac{1}{2} \times n$. Suppose that all the people with A have an accident, but none of the people with B suffer an accident. Then the actual wealth per capita will be $\frac{1}{2} \times 46 + \frac{1}{2} \times 54 = 50$, and that *is* feasible. But we can discount this challenge to the competitive equilibrium because, with a large number of individuals, the probability of it happening is virtually zero. That is, if we give $\frac{1}{2}n$ individuals policy A and the rest policy B we can't count on only those holding policy A to have an accident.

There is something seriously wrong with the argument of Example 7.7. The outcome is feasible only if a very specific—and very improbable—pattern of accidents occurs. If no one holding the policy that pays a high claim has an accident then the premiums can be sufficient to cover the claims paid out. Feasibility calculations should not be so contrived. To impose a more meaningful test we will say that a mix of policies is feasible if the *expected* amount of revenue from premiums is at least as large as the expected amount paid out in claims. We conclude this section by showing that there is no feasible set of policies—according to this new definition—that would give everyone more expected utility than the competitive equilibrium. We do that by showing that if a set of policies *S does* give everyone more expected utility than the competitive equilibrium, then *S* must not collect enough premium revenue on average to pay the claims that will be paid out on average.

Suppose that n_1 individuals get (x_1, y_1) , a different group of n_2 persons get (x_2, y_2) , and n_3 get (x_3, y_3) , and so on. Suppose that there are k different groups. Of course, $n_1 + n_2 + \cdots + n_k = n$. Suppose that the expected utility of each individual in group i is higher at (x_i, y_i) than it is at the competitive equilibrium. Because the latter maximizes individual expected utility on the market opportunity line $\pi x + (1 - \pi)y = \pi a + (1 - \pi)z$, an allocation delivering even higher expected utility must be above that line. Therefore,

$$\pi x_i + (1 - \pi) y_i > \pi a + (1 - \pi) z.$$

134 Basic Models and Tools

If we multiply both sides of that inequality by n_i we get $n_i \pi x_i + (n_i - n_i \pi)y_i > n_i \pi a + (n_i - n_i \pi)z$. Because $n_i \pi$ individuals will have an accident, the number $n_i \pi x_i + (n_i - n_i \pi)y_i$ is total group *i* wealth as a result of giving (x_i, y_i) to each person in group *i*, and $n_i \pi a + (n_i - n_i \pi)z$ is actual group *i* wealth. Therefore, the total wealth allocated to all individuals (over all groups) exceeds actual total wealth summed over all individuals. This tells us that we *can't* give (x_i, y_i) to everyone in group *i* for all groups because it requires more wealth to be distributed than the community actually has. Note that

$$n_i \pi x_i + (n_i - n_i \pi) y_i > n_i \pi a + (n_i - n_i \pi) z$$

is equivalent to

$$n_i \pi (x_i - a) > (n_i - n_i \pi)(z - y).$$

The second inequality says that the total net claim paid to the individuals in group i who have an accident exceeds the total money collected in premiums from the people in group i who do not suffer an accident. Therefore, there is no feasible outcome that gives everyone higher expected utility than the competitive equilibrium, establishing that the competitive equilibrium is efficient.

Sufficient condition for efficiency

If there is a large number of individuals, each with probability π of having wealth level *a* and probability $(1 - \pi)$ of wealth *z* then the outcome at which each person has $x = \pi a + (1 - \pi)z = y$ is efficient.

Example 7.8: Expected premiums and expected claims for Example 7.7

Suppose *a* = 40 and *z* = 60 with $\pi = \frac{1}{2}$. Half of the *n* individuals get *A* = (46, 56) and the other half get *B* = (50, 54). Therefore, *A* pays a net claim of 46 - 40 = 6, and *B* pays a net claim of 50 - 40 = 10. A's premium is 60 - 56 = 4 and *B*'s premium is 60 - 54 = 6. The expected number of accidents in each group is $\frac{1}{4}n$. The expect amount of premium income from the individuals who don't suffer an accident is $\frac{1}{4}n \times 4 + \frac{1}{4}n \times 6 = 2.5n$. The total expected value of claims paid out to those who do have an accident is $\frac{1}{4}n \times 6 + \frac{1}{4}n \times 10 = 4n$. Expected claims paid out (4*n*) exceeds expected premium income (2.5*n*).

Problem set

- 1. Prove that the competitive equilibrium is efficient (not just weakly efficient) by showing that if it is possible to give one person higher EU without lowering anyone else's EU then it is possible to give everyone higher EU.
- **2.** Note that individual utility is quasi-linear in this section. Prove that an outcome is efficient if and only if it maximizes total utility assuming that leisure

7. Insurance

 ℓ is transferrable across individuals. (Leisure is not really transferrable. I can't give you five hours of my leisure on Saturday, allowing you to consume twenty-nine hours of leisure that day. But we can justify the fiction if we allow the possibility that one individual can work for another. If I work for you for five hours on Saturday, my utility will fall by five and yours will increase by five.)

The remaining questions pertain to Dana, whose utility-of-wealth function is U(w) = 1 - 1/(w + 1), and Tyler, whose utility-of-wealth function is $U(w) = \ln(w + 1)$. Each has a current wealth of \$100 and in each case there is a probability of 0.2 that \$60 of it will be destroyed in an accident.

- **3.** Are Dana and Tyler each risk averse? Explain.
- **4.** Insurance can be purchased for twenty-five cents per dollar of coverage. (If the individual has an accident and has \$30 of coverage then that individual will get a claim check for \$30.)
 - A. Derive the market opportunity line. Does it exhibit fair odds? Explain.
 - **B.** How much insurance coverage will Dana purchase? Consequently, what will Dana's wealth be if she has an accident, and what wealth will she have if she does not have an accident?
 - **C.** How much insurance coverage will Tyler choose? Consequently, what will his wealth be if he has an accident, and what wealth will he have if he does not have an accident?
- **5.** Insurance can be purchased for forty cents per dollar of coverage. (If the individual has an accident and has \$30 of coverage then that individual will get a claim check for \$30.)
 - A. Derive the market opportunity line. Does it exhibit fair odds? Explain.
 - **B.** How much insurance coverage will Dana purchase? Consequently, what will Dana's wealth be if she has an accident, and what wealth will she have if she does not have an accident?
 - **C.** How much insurance coverage will Tyler choose? Consequently, what will his wealth be if he has an accident, and what wealth will he have if he does not have an accident?
- **6.** Insurance can be purchased for twenty cents per dollar of coverage. (If the individual has an accident and has \$30 of coverage then that individual will get a claim check for \$30.)
 - A. Derive the market opportunity line. Does it exhibit fair odds? Explain.
 - **B.** How much insurance coverage will Dana purchase? Consequently, what will Dana's wealth be if she has an accident, and what wealth will she have if she does not have an accident?
 - **C.** How much insurance coverage will Tyler choose? Consequently, what will his wealth be if he has an accident, and what wealth will he have if he does not have an accident?

Hidden Action

1.	Reso	Durce Allocation Operation Problem set Problem set	. 139 142
2.	Mar	ketable Pollution Rights	. 143 151
3.	Ince Indu	ntive Regulation of the Telecommunications	. 152
4.	The	Savings and Loan Debacle	. 155 163
5.	Pers	onal Bankruptcy	. 164
6.	Man	datory Retirement	. 165
	6.1	Posting a bond	166
	6.2	The formal argument	167
	6.3	The intertemporal budget constraint Problem set	171 173
7.	Ten	ure and the Performance of Professors	. 174
8.	Pay	and Performance in U.S. Public Schools	. 177
9.	Mor	al Hazard and Insurance	. 179
	9.1	Overview	181
	9.2	The formal model	186
	9.3	The binary choice model of moral hazard	188
	∂ 9.4	A continuum of effort supply levels	190
	∂ 9.5	Incomplete insurance	192
		Problem set	194

Hidden Action

This chapter and the next investigate the extent to which an agent can be motivated to act in the principal's interest when the principal cannot determine whether the agent has in fact taken the appropriate action. Agents' behavior is problematic because their goal is to maximize their own utility. The next chapter is devoted to the specific hidden action problem of motivating workers and management in a firm, with the latter receiving most of our attention. This chapter examines a wide variety of other issues. In many of these the principal is a surrogate for society as a whole, and the principal's utility is maximized when the agents—the producers and consumers—are all motivated to do their part in contributing to an efficient outcome.

As with all hidden information problems, there are hidden characteristic elements as well as the hidden action element. In fact, some of the topics could have been presented as hidden characteristic problems. For instance, we could study resource allocation from the standpoint of inducing consumers to reveal their hidden preferences and firms to reveal their hidden production technologies so that an efficient outcome can be identified. However, the approach taken in the first section is that of inducing each individual to choose a bundle of goods and services at which his or her marginal rate of substitution equals that of the other consumers. Similarly, when discussing pollution abatement in Section 2, we begin with the fact that the adjustment cost of an individual firm-the firm's characteristic-is hidden from the government. If the firm were simply asked to report its adjustment cost we would have a hidden characteristic problem, belonging in Chapter 5. However, we instead look at an incentive scheme that harnesses the profit motive to induce firms to coordinate their actions so that the adjustment burden falls on the firms that can reduce pollution at the lowest cost to consumers.

Although the incentives governing the behavior of workers and management in firms is the subject of the next chapter, we briefly examine a simple firm to introduce you to the central problem. Hidden action problems are complicated by the presence of uncertainty. If your car breaks down a week after you bring it home from the repair shop you do not know whether you are the victim of bad luck or shirking by the mechanic. This makes it hard to design efficient, incentive-compatible contracts. Consider the case of a principal who owns farmland and hires a worker-the agent-to operate the farm. Suppose that the landlord charges a fixed rent and hence allows the agent to keep all the proceeds of the farm over and above the rental payment. Then the agent has maximum incentive to run the farm efficiently. That's because the agent is the residual claimant under the rental contract: Once the rent is paid every additional dollar of profit goes directly into the agent's pocket. This should result in the maximum possible payoff for both the principal and the agent. The rent can be set at a level that leaves each with a higher return that can be achieved through any alternative contract-that is, sharing arrangement-that results in a lower profit.

We have implicitly assumed, however, that the agent is risk neutral. The farm's profit is affected by uncertainty in many ways. The weather, the activity of pests, and so forth can be viewed as random variables from the standpoint of both

the principal and the agent. It is reasonable to assume that the farm worker is risk averse. Because the profit derived from the farm has a random component, when the agent is the residual claimant, he not only has maximum incentive, he also has maximum exposure to risk. The rent will have to be reduced sufficiently to keep the agent on the farm and prevent him from taking a fixed-wage job in the city. The owner of the land will find that she can get more income by sharing the risk with the agent—by offering a contract that reduces the rental payment when events beyond the control of the agent reduce the harvest.

We still haven't incorporated all the ways in which uncertainty affects the nature of the contracts that principals offer their agents. There are things that the agent can do to mitigate the effect of harmful events. For instance, keeping fences in good repair makes it unlikely that the crop will be trampled by the neighbor's cattle. By devoting effort to keeping rodents out of the barn, less seed will be consumed by these intruders. In some cases a contract can be structured so that the agent's rent is reduced in bad times only if he has devoted effort to keeping pests away. However, if the owner

When you buy a new car the warranty is kept in force only if the periodic maintenance has been done properly. It is fairly easy for the manufacturer to determine whether the maintenance was proper. of the land does not live on the farm then she can't observe whether the agent has expended effort to that effect. In that case, the contract cannot be written so that the agent's return is conditional on his supply of effort. In other cases, the agent's effort is observable by the principal but not by a third party such as a

judge. For instance, if the owner lives on the farm she can see whether the agent puts effort into keeping rodents out of the barn. But the owner's observations would not be admissible in court, which would require independent evidence of shirking by the agent. Therefore, the conditional contract could not be enforced, so it would not be written in the first place. When the agent's effort can't be observed by the principal or by a judge or jury, a fixed rent would leave the worker with maximum exposure to risk but minimum incentive to shirk. However, a fixed wage would provide the worker with maximum shelter from risk but minimum incentive to work. The principal will maximize her return from the contract by finding the optimal trade-off between providing incentive and providing shelter from uncertainty. In this case farming often takes the form of sharecropping, particularly in developing countries: The worker gets a fixed share of the farm's profit, say one-half. If profit falls by 100 then the worker's share falls by 50, not by 100, so there is some insulation from uncertainty. When profit increases by 120 as a result of the agent's effort the agent's income increases by 60, so there is some incentive for the agent to supply effort, although it is not the maximum incentive.

The effect of risk on efficiency is not treated explicitly until Section 9, although uncertainty plays a supporting role in Sections 3 through 6.

Source

The rationale for sharecropping is based on Stiglitz (1974), a seminal contribution to the theory of incentives.

1. Resource Allocation

1 RESOURCE ALLOCATION

One of the elements of efficient resource allocation is ensuring that a consumer does not purchase a commodity beyond the point where it adds very little to the buyer's welfare when it could provide substantial additional benefit to someone else. This section considers the possibility of giving individuals an incentive to consume only up to a certain point—a point that depends on the preferences of others. The aim is to arrive at an efficient allocation of consumer goods. (Chapter 10 considers production and consumption simultaneously.)

To get a handle on the conditions for efficiency in the allocation of consumer goods we'll begin with the simplest case of two individuals, A and B, and two commodities, 1 and 2. Suppose that A's marginal rate of substitution (MRS) is 2 at his current consumption plan, and B's MRS is 1/2 at his current plan. That means that if A sacrificed less than 2 units of commodity 2 but received an additional unit of commodity 1 she would be better off as a result. If B lost a unit of the first good he would wind up better off if he received more than half a unit of the second good as compensation. Then if we arrange for A to give B one unit of the second good in return for one unit of the first good they would both wind up better off. In general, if $MRS_A > MRS_B$ then A and B could each gain by trading, provided that A exported the second good and imported the first good, and B did the reverse, and the amount of commodity 2 exchanged per unit of commodity 1 were between MRS_A and MRS_B . Similarly, they could strike a mutually advantageous trade if $MRS_A < MRS_B$. Efficiency requires equality of the marginal rates of substitution for any two individuals and any two commodities that each actually consumes.

Example 1.1: Two consumers with unequal marginal rates of substitution

A's utility function is $U_A = x^2 y$, and $U_B = xy^2$, where *x* is the amount of the first good consumed and *y* is the amount of the second good. Suppose that each person is currently consuming 4 units of each good. Then $U_A = 64 = U_B$. We don't actually have to compute each MRS here to construct a trade that increases the utility of each. Note that *A*'s utility function puts extra weight on the first good and *B*'s puts extra weight on the second good. In other words, the first good gets more weight in *A*'s preference scheme and the second good gets more weight in *B*'s preference scheme. Surely $MRS_A > MRS_B$ and both would be better off if *A* gave one unit of good 2 to *B* in return for one unit of the first good. Let's check: $U_A(5, 3) = 5^2 \times 3 = 75$ and $U_B(3, 5) = 3 \times 5^2 = 75$. The trade increases the utility of each.

If an economic system is not efficient, then there are equilibria that could be improved to the extent of making some people better off without adversely affecting anyone else. This would be a serious waste because it is extremely costly to identify the individuals in question and to bring about the necessary changes in economic activity. The economic system should not burden public policy makers with this kind of adjustment.

It is easy to show that at an equilibrium of the private ownership market economy is efficient if one person's consumption does not directly affect the welfare of another. Before giving the brief (but rigorous) proof of that claim we give the intuition: Consider two people, A and B, and two goods. Suppose that each consumes some commodity 1 and some commodity 2 at equilibrium. Then each individual's MRS will equal the price ratio P_1/P_2 , the price of good 1 divided by the price of good 2. The price ratio plays a central role in A's determination of her preferred consumption plan. But the price ratio equals B's MRS, so without realizing it A is taking B's preferences into consideration when determining her own consumption plan. It's as though A says, "I've studied economics. When my MRS is greater than the price ratio my MRS is greater than B's. Then because I place a higher intrinsic value on commodity 1, I am justified in consuming more of it. I'm not wasting resources. But I don't want to consume up to the point where my MRS is below the price ratio. If that happened I would be wasting resources. I would be consuming units of the good that have less intrinsic value to me than they do to person B." In fact, it is in A's self-interest not to consume good 1 beyond the point where her MRS equals the price ratio. The prices transmit information to A about the preferences of other consumers, and the budget constraint gives A the incentive to take that information into consideration when planning her consumption. This results in an efficient allocation of resources.

Now here's the general proof. Consider two individuals: *A*, who lives in Allentown (Pennsylvania), and *B*, who lives in Bozeman (Montana). They haven't met, and because the market economy is *decentralized*, with no central agency making sure that individuals do get together when something mutually advantageous might ensue, we have to ask if it is possible for *A* and *B* to trade in a way that would leave both better off. We're assuming that the system has reached an equilibrium before the trade takes place because we're testing the market economy for efficiency. Let's suppose that we have found a mutually advantageous trade. This assumption will quickly be shown to be untenable.

The trade must be balanced if we want to leave the consumption of others unchanged. (We want to increase U_A and U_B without harming anyone else.) The trade will be balanced if every increase in *A*'s consumption comes at the expense of *B* and vice versa. Let *a* denote the list of exports and imports for individual *A*. For instance, if a = (+7, -3, -6, ...) then *A* receives (imports) 7 units of the first good from *B*, but delivers (exports) 3 units of the second good and 6 units of the third good to *B*, and so on. We'll let *b* represent the list of *B*'s exports and imports. Hence, our example requires b = (-7, +3, +6, ...) because *B* exports 7 units of the first good to *A* and imports 3 units and 6 units, respectively, of the second and third goods. In brief, b = -a. Let *pa* denote the value of all *A*'s imports minus the value of all *A*'s exports, calculated using equilibrium prices. Similarly, *pb* is the value of all *B*'s imports minus the value of all *B*'s exports.

Suppose this trade makes both *A* and *B* better off. Because the trade makes *A* better off, we must have pa > 0. If $pa \le 0$ then these changes would already

1. Resource Allocation

have been incorporated into *A*'s consumption plan at equilibrium. For instance, if a = (+7, -3, -6) and each good costs \$2 (so pa < 0) then by reducing consumption of good 2 by 3 units and reducing consumption of good 3 by 6 units, individual *A* would have reduced her expenditure by enough to enable her to purchase 7 units of the first good. We're claiming that these changes would leave her better off, and that she could afford to make the changes on her own. (If the cost of the goods that *A* is acquiring is \$14, and the cost of the goods that *A* is giving up is \$18, then *A* can effect the change unilaterally.) That contradicts the notion that at equilibrium individuals have maximized their utility subject to the budget constraint. It follows that pa > 0.

Similarly, because the changes specified by *b* leave individual *B* better off they must not have been affordable when *B* chose his consumption plan. In other words, pb > 0. We also have a + b = 0 because every unit of a commodity imported by *A* is exported by *B* and vice versa. But a + b = 0 is inconsistent with *a* and *b* both having a positive market value: That is, we can't simultaneously satisfy a + b = 0, pa > 0, and pb > 0. For instance, if a = (+7, -3, -6), b = (-7, +3, +6), and each good costs \$2 at equilibrium, then pb = +4 but pa = -4. We are forced to abandon the supposition that there is a trade between *A* and *B* that would leave both better off than they are at the market equilibrium.

Once the market system reaches equilibrium, if someone in Allentown telephones everyone in Bozeman, hoping to find an individual with whom to strike a mutually advantageous trade, he or she will be disappointed. (How do we account for eBay then? Preferences have changed: People are trading things they no longer want. There is also a lot of retail activity on eBay—it is part of the market process.)

In any market, the price has three functions:

- 1. *Rationing*: The price adjusts until demand equals supply, which means that for every unit of the good that someone wants to buy there is a unit that someone wants to sell and vice versa.
- 2. *Information transmission*: The equilibrium price ratio transmits information to each consumer about the marginal rate of substitution of others.
- 3. *Incentive compatibility*: The budget constraint gives all individuals the incentive to take that information into consideration when planning their consumption.

We can easily extend this argument to any number of consumers. Let's organize a trade involving *n* individuals. Let t_1 be the list of exports and imports for individual 1, with t_2 denoting the list of exports and imports for individual 2, t_3 the list of exports and imports for individual 3, and so on. If this trade makes individual *i* better off, then we have $pt_i > 0$. If the trade makes everyone better off than under the market equilibrium, we have $pt_i > 0$ for each individual *i*. But if we add over all *n* individuals we get $pt_1 + pt_2 + \cdots + pt_n > 0$. This tells us that the total value of imports exceeds the total value of exports. That is inconsistent with the fact that for every unit of a good imported by someone there is a unit of the same good exported by someone. (We are not changing

any production plans at this stage. That will be considered in Chapter 10.) Therefore, it is not possible to have *any* number of individuals trade among themselves in a way that leaves everyone better off than they are at the market equilibrium.

Would it be possible to change the consumption of a subset of the individuals in a way that doesn't raise or lower the level of well-being of any member of the subset but do so in a way that generates a surplus that can be used to make someone else better off? No. Section 4 of Chapter 1 showed that if the menu of produced goods can be reallocated in a way that makes *some* individuals better off, and leaves others with the same level of welfare as before, then it is possible to make *everyone* strictly better off. (Just have the individuals who gain share some of the gain with the rest.) But we have just proved that there is no feasible outcome that makes everyone better off.

We said at the outset that the key assumption is that one person's consumption does not have a direct effect on another's welfare. Where did this assumption get used? If, say, person *C*'s welfare were affected by the consumption of *A* and *B*, we could construct an example of an economy such that, starting at equilibrium, we could have *A* and *B* trade in a way that neither increased nor decreased the welfare of either but that increased the welfare of *C*. We can't say that *C* would have brought about the change as part of his or her own consumption plan, because the increase in *C*'s utility requires *A* and *B* to act in very specific ways.

If we use the term *private good* to refer to a commodity that is immune to external effects, then we have shown that there is no change in the market equilibrium allocation of the produced private goods that could be used to make some individuals better off without harming others.

DEFINITION: Private good

A commodity is private if a change in the amount of it consumed by one individual has no effect on the welfare of any other individual whose own consumption does not change.

Links

Chapter 4 of this book examines the problem of giving a firm's manager the incentive to choose the production plan that contributes to efficiency. Koopmans (1957) is a classic and very readable exposition of the connection between efficiency and the competitive market system. Pages 1–126 are especially recommended.

Problem set

The two cases presented in Table 3.1 give you the amounts of two commodities *X* and *Y* consumed by two individuals *A* and *B*. Individual *A*'s utility function is $U_A(x, y) = xy$ and thus *A*'s MRS at commodity bundle (x, y) is y/x. Individual *B*'s

Table	3.	1
-------	----	---

	Cas Comn	se 1 nodity		Cas Comm	e 2 Iodity
Person	X	Y	Person	X	Y
А	4	8	А	2	2
В	6	2	В	8	8

utility function is $U_B(x, y) = x^2 y$ and the MRS at (x, y) is 2y/x. Answer each of the following questions for each case.

- **1.** Report the utility level and the MRS for each individual at the given commodity bundle.
- **2.** Construct a trade between *A* and *B* that *increases* the utility of each. Make sure that the trade is balanced—that is, the total consumption of *X* remains at 10 and the total consumption of *Y* also remains at 10.
- **3.** Report the exchange rate for your trade of question 2. (The exchange rate is the amount of commodity *Y* exchanged per unit of commodity *X*.) Notice that the exchange rate is between the two MRSs.)
- **4.** Using the exchange rate of question 3, construct a trade that *reduces* the utility of both *A* and *B* by making the trade "too big"—that is, with a large amount of exports and imports.
- **5.** Assume the exchange rate of question 3, but with *trade flowing in the opposite direction* (have *A* export the good that he or she imported in your answer to question 2). For each of the two cases displayed in Table 3.1, show that both individuals have *lower* utility than they had to start with (in question 1) as a result of trade.

2 MARKETABLE POLLUTION RIGHTS

Incentive regulation allows the regulated party a choice from a menu that is governed by the central authority. It has replaced command and control regulation in many cases. If the menu items are cleverly chosen then the outcome will be superior to what can be achieved by command and control regulation. That is because the regulated agent has better information than the

Command and control regulation can be quite daunting: The U.S. Department of Defense requires thirty-five pages of small print to define a T-shirt to guide private firms supplying that garment to military personnel (Stiglitz, 1993). central authority. After all, the local decision makers are on the scene day after day. They typically have far more at stake, which gives a strong motivation for acquiring information. A successful incentive scheme can tap this information by harnessing the self-interested behavior of the regulated agent.

Table 3.2.	Firm A
------------	--------

SO ₂ reduction	Revenue	Cost	Profit
50%	2000	1700	300
100%	2000	2300	300

This section presents a scheme for giving producers an incentive to cooperate in achieving the least-cost method of reducing pollution.

The production of electricity releases large amounts of sulphur dioxide (SO_2) into the air as coal is burned to produce the steam that runs the turbines. The SO_2 by-product can be reduced only if the firms that make electricity install new equipment. Suppose that a government agency wants to reduce the amount of SO_2 in the air in a way that minimizes the value of new equipment used in the production of electric power. The agency would have to acquire specific information about the production technology of individual electricity firms to identify the firms that can adjust at lowest cost. Without proper incentives, these firms will not disclose the information willingly and accurately. If a firm can lower its burden of adjustment by misleading the regulators, it can be expected to do so. Here is a simple example.

Example 2.1: Three ways to hit the target

There are two firms *A* and *B* who each have been dumping 30 tons of SO₂ into the air per year. The government wants to reduce total SO₂ output by 50%, and we consider three possibilities: *A* and *B* each reduce their SO₂ by 50%, *A* reduces its SO₂ by 100% while *B* continues to dump 30 tons per year, or *B* reduces its SO₂ by 100% while *A* continues to dump 30 tons per year. To determine which of the three is most advantageous for consumers we need the data in Tables 3.2 and 3.3. Table 3.2 shows that the 100% reduction in SO₂ output would require *A* to significantly modify its technology, and that would add so much to cost that *A* would incur a \$300 loss, instead of the \$300 profit that it would realize with the 50% reduction. Table 3.3 reveals that even a 50% reduction in SO₂ output would leave *B* with a loss, because it is much more costly for *B* to adjust its production recipe. Perhaps the adjustment cost will have to fall entirely on the shoulders of firm *A*. However, *A* can be expected to be greatly overstate its adjustment cost in an attempt to avoid this.

Table	3.3.	Firm	В
-------	------	------	---

SO ₂ reduction	Revenue	Cost	Profit
0%	2100	1300	800
50%	2100	2400	300

2. Marketable Pollution Rights

Suppose that the government required each firm of Example 2.1 to reduce its SO₂ output by 50%, without attempting to discover what an individual firm's adjustment cost is. Firm A would have to reduce its output from 30 to 15 tons but would be allowed to dump 15 tons of SO₂, and similarly with firm B. Suppose that the government also allowed one firm to sell its pollution entitlement of 15 tons to another firm. If A sold its entitlement to B for \$700 it would have to reduce its SO₂ output by 100%, and its profit would be 2000 + 700 - 2300 = 400. (The \$400 amount equals the loss of \$300 from the last row of the first table, plus the \$700 from the sale of the right to dump 15 tons.) This is more than the \$300 profit that it could make by keeping its entitlement to dump 15 tons. Therefore, A has an incentive to sell its pollution entitlement. And firm B has an incentive to buy it: *B* takes a loss of \$300 if it tries to reduce its SO₂ output by 50% (15 tons). However, if *B* pays \$700 to firm *A* for the right to dump 15 tons of SO₂ it will be able continue production without doing any adjusting at all. (It will have its own entitlement to dump 15 tons plus the 15-ton entitlement obtained from *A.*) Firm *B*'s profit would then be 2100 - 1300 - 700 = 100. (The \$100 amount equals the \$800 profit from the first row of the second table minus the \$700 payment.)

If the government asked each firm to report its adjustment cost to identify the least-cost way of reducing total SO₂ output by 50% it wouldn't get anything close to truthful revelation from individual firms. However, by allowing one firm to sell its pollution entitlement to another, the low-cost firm has an incentive to assume all of the adjustment cost, and the other has an incentive to pass the adjustment burden onto the former. Note that if each firm were forced to reduce its SO₂ output by 50% the two firms' costs would total \$4100. But with firm *A* shouldering the entire burden of adjustment the total cost is only \$3600. The incentive regulation scheme—that is, marketable pollution permits—significantly reduces the value of resources that have to be diverted from the production of other goods and services to enable the electricity generation industry to modify its technology and reduce SO₂ emissions.

This example illustrates how the U.S. program of marketable pollution permits has worked to reduce the sulphur dioxide production by the electric utility industry by 50% since 1990. In fact there is a large number of firms producing electricity, and each is given a number of pollution permits. Each permit entitles the bearer to dump one ton of sulphur dioxide into the air, and the total number of permits issued by the government equals the target level of SO₂ output. (The difference between the previous year's SO₂ output and the total number of permits issued in the current year is the target SO₂ reduction.) The individual firm can buy any number of entitlements and can sell as many of its own permits as it chooses. The selling price of a permit is not determined by negotiation between two firms but by a competitive market.

Suppose that at the current market price P_t of pollution permits the adjustment costs for firms in general are so high that the demand for permits exceeds the supply. The price will be bid up (to P_{t+1}). A firm that could reduce its SO₂ output by one ton at a cost *C* that is greater than P_t but less than P_{t+1} would now want to sell a pollution permit and increase its profit by $P_{t+1} - C$. At the lower price P_t the firm would have been better off to hold the permit (because

 $C > P_t$). Any firm with a cost $C > P_{t+1}$ would have demanded a permit at the lower price and will continue to demand it at the new higher price. Only the firms with $P_t < C < P_{t+1}$ will switch from demanders to suppliers of permits. Therefore, only the low-cost firms will, on their own initiative, modify their production process to reduce their SO₂ output. (Firms with $C < P_t$ will have already assumed that role. Firms with $C > P_{t+1}$ will continue to demand permits.)

Here is a simple proof that the equilibrium price of pollution rights will result in target pollution abatement being reached at the minimum cost. Let P denote the equilibrium price of the right to dump one ton of SO₂. Suppose that firm X dumps x tons of SO₂ at equilibrium. Let M_X^{ADJ} equal the cost of dumping x tons minus the cost at x + 1. (Cost decreases as x increases because it is costly to reduce SO₂ emissions.) Suppose that $P < M_X^{ADJ}$. Then the market is not in fact at equilibrium. Firm X can increase its profit by reducing its abatement effort by one ton, emitting x + 1 tons of SO₂ and buying a pollution permit at a cost of *P*. The increase in profit will be $M_X^{ADJ} - P$. Therefore, $P > M_X^{ADJ}$ must hold at equilibrium for every firm X. Now, suppose firm Y uses exactly q pollution permits at equilibrium. Let M_V^{POL} be the addition to Y's cost were it to use only q-1 permits and assume the cost of reducing its SO₂ output by one additional ton. (The costs are calculated without taking into consideration any purchase or sale of pollution permits.) If $P > M_V^{POL}$ we can't be at equilibrium because firm Y could increase its profit by incurring the cost M_V^{POL} of increasing its abatement by one and selling one permit (or eliminating the need to buy one permit) at a price of *P*. Therefore, $P < M_V^{POL}$ must hold at equilibrium for every firm *Y*.

We have demonstrated that $M_x^{ADJ} < P < M_Y^{POL}$ must hold at equilibrium for any two firms *X* and *Y* (including the case X = Y). Then the equilibrium price of a pollution permit must fall between the highest M_x^{ADJ} over all firms *X* and the lowest M_Y^{POL} over all firms *Y*. Because $M_x^{ADJ} < P$ for any firm *X* that has reduced its SO₂ output by one additional ton, and $P < M_Y^{POL}$ for any firm *Y* that has used a pollution permit, we see that market forces ensure that the adjustment is made by a firm (such as *X*) when that can be accomplished at a lower cost to society than when it is made by some other firm (such as *Y*). The price transmits information to each firm about the adjustment costs of other firms, and the profit motive gives each firm the incentive to use that information in the socially optimal way. Any change in the SO₂ adjustment pattern away from equilibrium would shift the adjustment burden from some firm *X* with $M_x^{ADJ} < P$ to some firm *Y* with $P < M_Y^{POL}$, and that would increase the value of resources consumed in reaching the abatement target. The total cost of achieving the target would increase by $M_Y^{POL} - M_x^{ADJ}$.

We have implicitly assumed that a firm's revenue is independent of the amount of pollution abatement. In that case, with revenue constant, profit maximization reduces to cost minimization. However, we would expect to see the firm's output fall when it increased its pollution abatement effort, and the change in output would usually result in a change in revenue. We can rescue the argument of the previous two paragraphs simply by setting M_x^{ADJ} equal to X's profit when it dumps x + 1 units of SO₂ minus its profit when it dumps x tons. (We do *not* include the effect on profit of the purchase or sale of pollution permits.) Similarly, M_Y^{POL} will now denote Y's profit when it uses *q* permits minus its profit

2. Marketable Pollution Rights

when it uses q-1 permits, before taking into consideration the purchase or sale of permits. The equilibrium condition $M_x^{ADJ} < P < M_Y^{POL}$ then follows.

With the modified definitions of M_x^{ADJ} and M_Y^{POL} , to take into consideration revenue effects, we can demonstrate that the equilibrium with marketable pollution rights leads to the maximum industry profit, *given* the target overall abatement level: Any change in the SO₂ adjustment pattern away from equilibrium would shift the adjustment burden from some firm *X* with $M_x^{ADJ} < P$ to some firm *Y* with $P < M_Y^{POL}$, and that would reduce total industry profit by $M_Y^{POL} - M_x^{ADJ}$.

It is important for you to see that our proof that the equilibrium leads to maximum industry profit is independent of how the pollution rights are initially allocated. As long as these rights can be bought and sold in a competitive market, the resulting equilibrium maximizes total profit. We know that to be the case because our proof did not require the specification of the initial rights assignment. Therefore, it is valid for any initial distribution of rights that sum to a given total.

When the market for pollution rights reaches an equilibrium, total industry profit is maximized, subject to the constraint that total pollution in the industry has to fall to the target level specified by the regulatory authority. Moreover, this holds true for any assignment of rights summing to the given target.

Why do we use profit as a measure of the net benefit that consumers derive from a firm's activities? Revenue is a rough measure of the benefit that consumers receive from the firm X's output of goods and services. The cost of production is equal to the market value of inputs used, and that in turn is a rough measure of the value to consumers of other goods that could have been produced with the resources that were employed instead in firm X. The difference measures the net value to consumers of the firm's activities. Therefore, we want to distribute pollution abatement costs in a way that maximizes the total profit over all firms subject to the total amount of pollution not exceeding the specified amount.

Example 2.2: Pollution and profit

Firms *X* and *Y* each released 100 tons of SO_2 last year, and they are required to bring that down to 80 tons each this year. As long as the total SO_2 output is 160, the target pollution abatement will be achieved. Table 3.4 gives individual firm profit figures for different levels of SO_2 output. Higher pollution abatement levels involve higher costs and hence lower profit.

There are nine different ways that the two firms can combine to reduce the amount of SO₂ released to 160 tons, and Table 3.5 gives total profit for each combination. We let *x* (respectively, *y*) denote the SO₂ output of firm *X* (respectively, firm *Y*).

We see that total profit is maximized when firm *X* releases 90 tons of sulphur dioxide and *Y* releases 70 tons. If the government gives each firm the right to

Firm SO ₂ output	Firm X profit	Firm Y profit
100	515	450
95	510	430
90	500	425
85	480	420
80	470	415
75	460	410
70	440	400
65	420	350
60	390	300

	Та	ble	3.4
--	----	-----	-----

release 80 tons, firm *Y* can sell 10 tons of its entitlement to firm *X* for a fee of *P* dollars. On one hand, because the profit realized by firm *X* is \$470 when x = 80 and \$500 when x = 90, it would be to its advantage to buy the rights to 10 tons for any price less than \$30. On the other hand, *Y*'s profit is \$415 when y = 80 and \$400 when x = 70, so it would be prepared to sell the right to dump 10 tons if *P* exceeds 15. At any price between \$15 and \$30 both would profit from the transaction. From the standpoint of consumer welfare, the potential for trading in pollution rights results in the target pollution abatement being achieved in a way that maximizes consumer welfare. Why wouldn't the two firms strike a different deal? Because any other pollution abatement assignment α would result in less total profit, and hence there is a price *P* that would divide the total profit from x = 90 and y = 70 in a way that gives each firm more profit than α .

Carbon permits allocated on a global scale would go a long way toward reducing the carbon dioxide (CO₂) emissions that are suspected of causing global climate change. The permits should be allocated by auction, rather than by giving a larger share of the permits to firms that did proportionately more polluting in the past. The auction would provide more incentive to invest in research to develop new production technologies that reduce CO₂ emissions. Heavy polluters have to pay more to continue operating with the old technology when auctions are used (Cramton and Kerr, 1999, 2002).

Trade in pollution rights allows the two firms to maximize their total profit *and to share that total in a way that gives each more profit than if the government had insisted on them sharing the burden of pollution abatement equally*. Because total profit is maximized, the cost to consumers of pollution abatement is minimized. (If it helps you appreciate this point, you can assume that each firm's revenue is unchanged when sulphur dioxide output is reduced. In that case, maximizing total profit is equivalent to minimizing total cost.)

Finally, we consider a model with many firms and assume that it is feasible for each to reduce its pollution by-product by any real number amount between zero and the

Table 3.5

x	у	Total profit
100	60	815
95	65	860
90	70	900
85	75	890
80	80	885
75	85	880
70	90	865
65	95	850
60	100	840

previous period's level. Let *S* be the target level of pollution for the current period. (Of course, *S* is below last period's level.) Let σ be the assignment of pollution levels to the individual firms that maximizes total profit. Because σ maximizes total profit, that total can be distributed among the *n* firms in a way that leaves each with more profit than if each had been required to set its pollution level to *S*/*n*. Allowing some firms to sell some part of their entitlement to dump *S*/*n* tons of pollutant to other firms allows them to achieve this distribution of the total profit.

Example 2.3: Two firms, each with quadratic profit functions

We let *x* denote the amount of SO₂ released by firm *X* as a by-product of its production process. The higher is *X*'s output of goods and services, the larger is *x*. We let f(x) denote the profit realized by *X* when the SO₂ output is *x*. Specifically

$$f(x) = 190x - 5x^2.$$

Using the formula for maximizing a quadratic, we see that *f* is maximized when $x = 190/(2 \times 5) = 19$. The reason why profit declines as *x* increases when x > 19 is that higher levels of SO₂ result from higher sales of the firm's product, and because marginal cost of production is increasing and marginal revenue is nonincreasing, there is a point when profit falls as output increases—and consequently, profit falls when SO₂ increases.

We let g(y) be firm Y's profit when y tons of SO₂ are emitted by firm Y. Specifically

$$g(y) = 110y - 5y^2$$
.

Y's profit is maximized when $y = 110/(2 \times 5) = 11$.

If each firm can maximize profit without constraint we will have x = 19 and y = 11, in which case a total of 30 tons of SO₂ will be released into the air. But suppose that the government wants to restrict the total emissions to 20 tons. To incorporate this constraint we can set y = 20 - x. Now, maximize total profit, f + g, with y = 20 - x. We want to maximize

$$190x - 5x^{2} + 110(20 - x) - 5(20 - x)^{2} = 280x - 10x^{2} + 200.$$

This is maximized when $x = 280/(2 \times 10) = 14$. We have x = 14 and y = 6. (Remember, y = 20 - x.) Confirm that f(14) = 1680 and g(6) = 480. However, if the government required the firms to share equally the burden of reduced SO₂ emissions, each firm would have to reduce its pollution to 10 tons per year (for a total of 20). In that case f(10) = 1400 and g(10) = 600 are the respective profit figures. Table 3.6 summarizes.

	Та	bl	e	3.	6
--	----	----	---	----	---

x	у	X's profit	Y's profit	Total profit
10	10	1400	600	2000
14	6	1680	480	2160

Suppose that instead of insisting that each firm reduce its SO₂ output to 10 tons, the government gave each firm the right to release 10 tons into the air and allowed each firm to sell all or part of that right. Then if Y sells firm X its right to dump 4 tons of SO₂ for a price between 120 and 280, each firm would have more profit than under the equal burden formula. (X would pay up to 1680 -1400, and Y would have to be paid at least 600 - 480.) And the total amount of pollution would still be 20. Conversely, if X had been given the right to dump 15 tons of SO₂ and Y had been given the right to dump 5 tons, the target of 20 would still be reached. But this time firm X would sell Y the right to dump 1 ton, resulting in the maximum total profit once again (subject to x + y = 20). The minimum that X would accept in payment is the difference between profit at x = 15 and profit at x = 14. The maximum that Y would be prepared to pay is the difference between profit at y = 6 and profit at y = 5. Marketable pollution rights give the two firms the chance to share the maximum total profit in a way that makes each better off than under a rigid assignment of individual firm pollution limits, however the rights are initially assigned.

We see that marketable pollution permits achieve the target SO_2 reduction. The government determines the number of permits, and a firm must surrender one permit for every ton of SO_2 released. The permits achieve that target in the way that is most beneficial to consumers—reducing SO_2 output requires resources to be diverted from the production of goods and services so that the electric utility can modify its production technology, and we have seen that, if revenue is unaffected, the lower cost firms will do the adjusting. The low-cost firms maximize profit by selling some of their permits, requiring them to further reduce their SO_2 output. The high-cost firms maximize profit by buying additional permits, allowing them to release more SO_2 than their initial allotment of permits allows. Marketable pollution permits give each firm the incentive to implement the production plan that would be assigned to it by a central planning authority *if* the planning authority were able to obtain reliable information about the firm's adjustment cost.

Links

This book's web site provides a general proof that consumer welfare is maximized at equilibrium when pollution allowances can be traded. The fact that the distribution of emissions levels across firms is independent of the way that

2. Marketable Pollution Rights

rights are initially allocated across firms (given total emissions) is a special case of the Coase theorem (Coase, 1960). Schmalensee, Joskow, Ellerman, Montero, and Bailey (1998) and Stavins (1998) provide good overviews of the success of the American pollution allowance program. Joskow, Schmalensee, and Bailey (1998) provide a deeper, more technical analysis. For an application of the pollution permit idea to carbon regulation see Cramton and Kerr (1999). For global perspectives see Schmalensee, Stoker, and Judson (1998) and Chichilnisky and Heal (1993, 1999).

Problem set

1. Show that for any total SO₂ output *S*, for any assignment *A* of emissions levels to the individual firms that totals *S*, at the assignment of emissions levels that maximizes total profit (subject to total emissions being *S*), the total profit can be shared in a way that gives each firm more total profit than under *A*.

For the remaining questions, x denotes the amount of SO₂ released by firm X as a by-product of its production process, with f(x) denoting the resulting profit of firm X. Similarly, let g(y) be firm Y's profit when y tons of SO₂ are emitted by firm Y.

- **2.** Let $f(x) = 1200x 10x^2$ and $g(y) = 4000y 20y^2$. Determine the profitmaximizing values of *x* and *y* respectively when there is no limit on pollution and each firm maximizes its profit independently of the other. Now, suppose that the government limits each firm's pollution output to 50 but allows either firm to sell some or all of its pollution allowance to the other. Determine the equilibrium values of *x* and *y*.
- **3.** Let $f(x) = 144x 4x^2$ and $g(y) = 120y 5y^2$.
 - **A.** Assuming that the firms are not regulated in any way, find the profitmaximizing levels of *x* and *y* for firms *X* and *Y*, respectively. Determine the profit realized by each firm.
 - **B.** The regulatory authority requires $x \le 15$ and $y \le 6$ but allows any firm to sell some or all of its right to pollute to the other firm. Determine the resulting equilibrium values of *x* and *y*, and the profit realized by each firm under two conditions: (i) assuming that pollution rights are *not* tradeable, and (ii) assuming that pollution rights *are* tradeable, but in this case compute each firm's profit at the new values of *x* and *y* before taking into consideration the money that changes hands as a result of the exchange of pollution rights.
 - **C.** Which firm sells pollution rights and which firm buys them? How many rights are exchanged?
 - **D.** Let *P* denote the price of a right to dump one ton of SO₂. Find the range in which the equilibrium value of *P* must fall when the constraints $x \le 15$ and $y \le 6$ are imposed.

152 Hidden Action

- 4. The firms are the same as in question 3: $f(x) = 144x 4x^2$ and $g(y) = 120y 5y^2$.
 - **A.** The regulatory authority requires firm *X* to reduce its output of SO_2 by 6 tons and *Y* to reduce SO_2 output by 3 tons. It allows any firm to sell some or all of its right to pollute to the other firm. Determine the resulting equilibrium values of *x* and *y* and the profit realized by each firm before adding or subtracting money from the sale or purchase of pollution rights.
 - **B.** Answer questions 3C and 3D for the situation of 4A (with the inequalities of part D appropriately modified).
- **5.** Let $f(x) = 300x 10x^2$ and $g(y) = 120y 5y^2$.
 - **A.** Assuming that the firms are not regulated in any way, find the profitmaximizing levels of *x* and *y* for firms *X* and *Y*, respectively. Determine the profit realized by each firm.
 - **B.** The regulatory authority requires $x \le 12$ and $y \le 12$ but allows any firm to sell some or all of its right to pollute to the other firm. Determine the resulting equilibrium values of *x* and *y* and the profit realized by each firm.
 - **C.** Which firm sells pollution rights and which firm buys them? How many rights are exchanged?
 - **D.** Let *P* denote the price of a right to dump 1 ton of SO₂. Find the range in which the equilibrium value of *P* must fall when the constraints $x \le 12$ and $y \le 12$ are imposed.
- 6. The firms are the same as in question 5: $f(x) = 300x 10x^2$ and $g(y) = 120y 5y^2$.
 - **A.** The regulatory authority requires firm *X* to reduce its output of SO_2 by 2 tons and *Y* to reduce SO_2 output by 1 ton. It allows any firm to sell some or all of its right to pollute to the other firm. Determine the resulting equilibrium values of *x* and *y* and the profit realized by each firm before adding or subtracting money from the sale or purchase of pollution rights.
 - **B.** Answer questions 5C and 5D for the situation of 6A (with the inequalities of part D appropriately modified).

3 INCENTIVE REGULATION OF THE TELECOMMUNICATIONS INDUSTRY

Incentive regulation allows the regulated party a choice from a menu that is governed by the regulatory authority. We use the telecommunications industry to illustrate. *Rate of return regulation* has been used for decades to curtail the market power of the suppliers of telephone services. It allows the regulated firm

3. Incentive Regulation of the Telecommunications Industry

to set prices high enough for its revenue to cover all costs of production and to provide a reasonable return on capital as well. But the return on capital cannot exceed the limit set by the regulatory authority. The implicit restraint on the firm's price reduces the consumer welfare losses that arise when a firm with substantial market power sets a price well above marginal cost. However, the firm that is governed by rate of return regulation has little incentive to innovate or reduce cost if it is a monopolist, because the resulting increase in profit would result in a rate of return above the limit, which in turn would force a reduction in price.

The failure to take advantage of productivity gains, or cost reductions in general, can result in consumer welfare losses that swamp any gains from lower prices. Allowing the firm to operate as an unregulated monopolist would seem to eliminate that problem because a reduction in cost of a dollar results in an increase in profit of a dollar. However, the discipline of competition is an important factor in driving a firm to reduce cost or to innovate. Monopolies tend to be sluggish, in spite of the fact that a dollar in the pocket of a shareholder of a monopoly is no less welcome than a dollar in the pocket of an owner of a firm operating under intense competition.

Price cap regulation eliminates the excessively high prices associated with monopoly power without eliminating the incentive to reduce cost or improve product quality. The regulated firm is required to reduce prices annually by a fraction x determined by the regulatory authority. This fraction, called the productivity offset, is an estimate of the industry's future productivity growth. If the value of input required per unit of output falls by x% then the price can decrease by that same x%, without causing revenue to fall short of cost. This gives the firm a strong incentive to innovate, in order to realize the productivity gain that will keep it from insolvency. Moreover, once that goal is reached, any additional dollar of profit-from further cost reductions or product improvementsis retained by the firm. Hence there is a strong incentive to innovate and cut costs under price cap regulation, which was imposed on British Telecom in 1984 and AT&T in the United States in 1989. The drawback is that the regulatory authority cannot predict future productivity increases with certainty. If they impose too stringent a price reduction on the firm it may be plunged into insolvency. The result is job loss and perhaps a disruption in supply.

The dilemma can be solved by giving the firm a choice between a price cap and rate of return regulation. If a firm cannot achieve a satisfactory rate of return on capital under a price cap, it will choose rate of return regulation because it not only allows the firm to raise prices to a level sufficient to cover costs—and hence avoid insolvency—but also a modest rate of return is allowed. The firm that would *not* have its rate of return driven below an acceptable level under price cap regulation will choose a price cap: If *r* is the maximum return allowed under rate of return regulation, and the firm can obtain a higher return under price cap regulation, it will obviously choose the price cap. The superior consumer benefits of price cap regulation will be realized in most cases but not at the cost of killing off the firms that would go bankrupt under the price cap. Here is a simple model that shows that adding the option of returning to rate of return regulation delivers higher consumer welfare than mandatory price cap regulation: Let F(x) be the probability that the firm's actual productivity gain is less than the productivity offset x imposed by the regulatory authority. If a price cap is mandated, and the firm is rendered insolvent, the level of consumer welfare will be A. However, if the firm remains healthy under the price cap, the benefit to consumers will be $B \times (1 + x)$. Quantity B is greater than A, and B(1 + x) is higher when the productivity offset x is higher. If the firm that would be insolvent under a price cap chooses to be governed by rate of return then the consumer benefit level is B, which is less than B(1 + x) but greater than A.

The probability that the firm would be insolvent under price cap is F(x), the probability that its actual productivity increase is smaller than the mandated price reduction *x*. Therefore, 1 - F(x) is the probability that the firm would be solvent under a price cap. The expected consumer benefit under a *mandatory* price cap is

$$V(x) = F(x)A + [1 - F(x)]B(1 + x).$$

The expected consumer benefit if the firm has the option of choosing rate of return regulation when it would otherwise go broke is

$$W(x) = F(x)B + [1 - F(x)]B(1 + x).$$

If price cap is mandatory, the regulatory authority chooses *x* to maximize V(x). Let x_M denote the solution. If price cap is optional, the authority chooses *x* to maximize W(x), and we let x_O be the solution. $W(x_O)$ is larger than $V(x_M)$. In other words, consumer welfare is higher when a price cap is optional. That follows from the fact that $W(x_O) \ge W(x_M) > V(x_M)$. We have $W(x_O) \ge W(x_M)$ because x_O maximizes *W*. And $W(x_M) > V(x_M)$ follows from the fact that B > A.

Example 3.1: F is the uniform distribution

We suppose that *x* is uniformly distributed on the interval 0 to β , with $\beta > 1$. Therefore, $F(x) = x/\beta$. (Review Section 6.5 of Chapter 2.) Consequently

$$V(x) = \frac{x}{\beta} \times A + \left(1 - \frac{x}{\beta}\right) B(1+x)$$

= $\frac{1}{\beta} [Ax + B(\beta - x)(1+x)] = B + \frac{1}{\beta} [(A + B\beta - B)x - Bx^2].$

Because *B* and β are constant, *V*(*x*) and *Ax* + *B*(β - *x*)(1 + *x*) will be maximized at a common value of *x*. The formula for maximizing a quadratic (Section 1 of Chapter 2) yields the solution value

$$x_M = \frac{A}{2B} + \frac{\beta - 1}{2}.$$

4. The Savings and Loan Debacle

Now we maximize

$$W(x) = \frac{x}{\beta} \times B + \left(1 - \frac{x}{\beta}\right) B(1+x)$$
$$= \frac{B}{\beta} [x + (\beta - x)(1+x)] = B + \frac{B}{\beta} [\beta x - x^2]$$

W(*x*) and $\beta x - x^2$ are maximized at a common value of *x* because *B* and β are constant. The solution is

$$x_0 = \frac{\beta}{2}.$$

When a firm that is in danger of going broke can choose rate of return, the regulatory authority can impose a more stringent (higher) productivity offset, resulting in lower prices set by firms operating under price cap. That is reflected in this example, because $x_O > x_M$. That is a consequence of the fact that A < B and thus A/2B < 1/2. Therefore, $x_M < 1/2 + (\beta - 1)/2 = \beta/2 = x_O$.

Source

Much of this section is based on Sappington and Weisman (1996).

Links

Leibenstein (1966, 1976) discusses the effect that the discipline of competition has on innovation.

4 THE SAVINGS AND LOAN DEBACLE

A savings and loan firm (*S&L*, or *thrift*), like a bank, takes in depositors' money, paying interest on those deposits, and then lends their money for a fee. Until the early 1980s its profit came mainly from the difference between the interest rates on lending and borrowing. Loans by an S&L were essentially limited to residential mortgages until the Depository Institutions Act of 1982 eased restrictions. Maximization of general consumer welfare requires monitoring of borrowers to ensure that the funds are devoted to the installation of capital equipment with the highest rate of return to society. And it is certainly in the interest of depositors as a whole to monitor their creditors to ensure that the funds will yield the maximum monetary return. However, no *individual* has an incentive to do the monitoring.

Deposit insurance eliminates the lender's (i.e., depositor's) incentive to comparison shop. Deposit insurance means that even if the institution holding your money fails, the balance in your account will be covered by the insurer. Consequently, depositors have no incentive to shop for a bank or S&L that will be careful with their money, thereby diminishing the borrower's incentive to avoid excessive risk. So why not eliminate deposit insurance? Because it is key to preventing bank runs. (There was an epidemic of them after the 1929 stock market crash.) If I anticipate that many of my bank's depositors are going to withdraw their money, and my deposit is not insured, then it is in my interest to try to get to the bank first to withdraw my deposit, before the bank's cash reserves are exhausted. This is true whether or not I think that the others are foolish for wanting to withdraw their money in the first place. Thus, deposit insurance provides a clear social benefit—stability of the banking system. But there is a cost: Even solvent banks or S&Ls will undertake more risky loans if their depositors do not penalize them for doing so—by withdrawing their money. And, as we discuss, S&Ls that are insolvent but that are allowed to continue operating have a very strong incentive to assume risks that significantly diminish social welfare.

Prior to 1980 the thrift industry was a sleepy one, protected and coddled by Congress and state governments. A thrift's primary-almost exclusive-source of income was long-term home mortgages. The loan paid a fixed interest rate for a thirty-year period-sometimes for a shorter period-determined when the mortgage was obtained. The thrift's deposit liabilities were primarily savings accounts that could be withdrawn at any time. This made it vulnerable to an interest rate squeeze: If rates increased significantly, competition for deposits forced the institution to raise the interest paid on deposits, while the bulk of its income came from low-interest mortgage loans many years from maturity. Some breathing room was provided by the fact that competition with other lenders was limited by a law that restricted a thrift's lending ability to a 100-mile radius. In fact, post-World War II economic growth, especially in housing construction, kept the industry fairly healthy. S&L failures were rare. When interest rates increased in the 1960s and the thrifts were squeezed, Congress responded by placing a ceiling on the rate that an S&L could pay on deposits. This eliminated price competition within the industry-as long as the equilibrium rate was above the ceiling-and the firms then competed by offering gifts to anyone who would open a deposit. When Congress limited the value of those gifts, the thrifts competed by staying open longer. The interest rate ceiling protected the industry from the interest rate squeeze for the rest of the 1960s and most of the 1970s.

By 1980 30% of the nation's thrifts reported losses as a result of sharply rising interest rates during the period 1979 to 1982. The worst year was 1982, in which 80% reported losses. Congress responded by deregulating the industry, allowing a thrift to make a wide variety of new loans. An S&L could now offer variable rate mortgages, make car loans, and issue credit cards, among other new opportunities. They were also allowed to have a higher fraction of their loans in the business sector. The interest rate ceiling was phased out, thrifts were allowed to pay interest on checking deposits, and the amount of deposit insurance was increased to \$100,000 per account. All of this would have rescued the industry if deregulation had not been accompanied by incentives to assume excessive risk. Accounting standards were relaxed. For instance, an asset purchased with depositors' money could be kept on the books at its original value for several years after a drop in that value, and a thrift could record \$11,000 as current income if a borrower seeking \$100,000 for a project was given \$110,000 on the understanding that the extra \$11,000 was to be used to pay the first year's interest. S&L deregulation also meant a reduction in monitoring by the government board charged with overseeing the industry. (The term deregulation

4. The Savings and Loan Debacle

refers not to the elimination of all kinds of oversight but to the substitution of regulation by consumers for regulation by a government agency. As we have noted, consumers—that is, depositors—had no incentive to monitor the S&Ls.)

Diversification gave a thrift new opportunities for increasing its income but also new opportunities for risk taking. The wave of failures in the 1980s included a disproportionate number of S&Ls with heavy investments in land loans, direct equity (i.e., the purchase of stocks), and commercial mortgages. Fraud also played a role but in a minority of cases: False statements were made to the regulatory authority, inappropriate loans were made to relatives and business partners of the thrift's officers, a borrower's assets were dishonestly valued to justify large loans, and sometimes excessive amounts of money were spent on the offices and other amenities of the thrift's chief executives.

The *Federal Savings and Loan Insurance Corporation (FSLIC)* is the federal program that guaranteed customers' deposits. If a thrift failed then FSLIC covered any part of a deposit that couldn't be collected from the failed institution. The wave of S&L failures beginning in the 1970s led to a crisis in which more than \$30 billion of deposits had to be redeemed in this way. Depositors who had to be bailed out had placed their money in S&Ls, which had used the money to purchase assets that subsequently fell in value. In fact, these assets collectively fell by more than \$30 billion in market value. This was an enormous waste in resources. For instance, if a thrift used \$10 million of depositors' money to finance the construction of an apartment building for which few tenants could be found, then the market value of the building would be far less than \$10 million. If the money had been invested more wisely, the value of consumer goods and services would have increased not decreased.

The initial S&L failures have their explanation primarily in the drop in oil prices, which had serious implications for real estate values and business activity in the "oil patch," particularly Oklahoma and Texas; a slump in real estate generally; and a rise in interest rates that left many S&Ls locked into long-term mortgages yielding low rates of return while paying high interest rates to current depositors. In this section we do not examine the onset of the crisis. Rather, we ask, "Given the original conflagration, why was gasoline poured on the flames instead of water?"

In 1981 almost 4000 thrifts were insured by FSLIC. Seventy percent of U.S. thrifts reported losses that year, and the whole industry's net worth was negative—the market value of assets fell short of the dollar deposit liabilities. Here was a clear warning sign. Yet in 1986 the President's Council of Economic Advisors was still trying to get the attention of the president, the Congress, and the country, calling for reform and warning of the potential bill that would be presented to taxpayers.

Here is the key to understanding how we managed to pour gasoline on the flames: *Zombie* institutions—thrifts that were insolvent and should have been pronounced dead—were allowed to *gamble for resurrection*. They took in more money from depositors and sunk it into risky investments in desperation. The risky investment would likely turn sour, but in the unlikely event that it succeeded it would restore the company to financial health. Let's look at this from

the standpoint of incentives. Why did depositors entrust their wealth to zombie institutions? Why did the regulatory agency overseeing the thrift industry (the Federal Home Loan Bank Board) permit zombie thrifts to continue gambling? And finally, why did the owners of the S&Ls want their firms involved in wildcat schemes?

First, why didn't depositors do a better job of monitoring the thrifts that borrowed from them? Because the federal deposit insurance program removed the incentive for depositors to do comparison shopping. Lenders still had a strong incentive to look for the highest interest on their deposit, but they had little reason to care about financial insolvency or imprudent thrift mangers. If the deposit were lost as a result of the thrift's insolvency then U.S. taxpayers, through the federal government, would replace the money. The Canadian banking system did not have a formal deposit insurance scheme until 1967. The stability of the Canadian system before 1967 can be partly attributed to the incentive for monitoring by lenders and to market discipline on the part of the individual bank. There is little incentive for lenders to monitor U.S. banks-in addition to the family of savings and loan institutions. The crisis was confined mainly to the thrift industry because banks were subject to more stringent regulation. However, the value of outstanding loans to Latin America by nine giant U.S. banks was almost double the capital of those banks, and repayment of the loans was problematic. The U.S. government and Federal Reserve indirectly rescued the banks by assisting Mexico and other Latin American countries.

Between 1890 and 1966 only twelve Canadian chartered banks failed, and in only six of those failures did depositors lose any money. The stability can be traced to the monitoring incentive, as well as to portfolio and geographical diversification of Canadian branch banks. (Nationwide branch banking is severely limited by regulation in the United States; Carr, Mathewson, and Quigley, 1995.) Why was the regulation of the thrift industry much more permissive than that of the banking industry? In particular, why did the regulatory agency (FHLBB) not put a stop to gambling for resurrection in the thrift industry? Because Congress generally favored regulatory forbearance. Why would a federal regulatory agency be sensitive to the mood of Congress? Because congress can restrict the powers of a regulatory agency. Also, many who serve on the regulatory board look forward to lucrative careers in Washington when they leave the agency counseling firms on how to approach Congress,

for example. So, even an independent agency is wary about defying Congress. Congress can cut the agency's budget as well as its powers. In fact it refused to increase the fund that FSLIC used to redeem the deposit liabilities of failed thrifts, even though the fund was not large enough to cover the deposits held in zombie S&Ls. The regulators faced a dilemma: If they shut down the zombie S&Ls there would not be enough money to rescue the stranded depositors. If they allowed the zombies to continue operating the crisis would deepen. They chose the latter. Finally, the 1982 Depository Institutions Act changed the accounting rules to allow ailing S&Ls to hide their insolvency, making them appear healthy. Before we consider why Congress wanted a permissive regulatory climate, let's see why the owners of a thrift would be in favor of gambling for resurrection in the first place.

4. The Savings and Loan Debacle

Values	Depositors	Owners	FSLIC	Society
Initial outlay	1000	40	0	1040
High return	1050	46	0	1096
Low return	1050	42	0	1092
Average return	1050	44	0	1094
Rate of return	5%	10%		5.2%

Table 3.7. Prudent investment (PI)

Consider a prudent investment (*PI*). It requires an initial capital outlay of \$1040 and it pays off exactly one year after the project is undertaken. *PI* will provide a return of \$1096 to the S&L, net of labor and materials costs, with probability $\frac{1}{2}$. And with probability $\frac{1}{2}$ the net return is \$1092. If the S&L invests in the project the owners will put up \$40 of their own money, and the remaining \$1000 will be money entrusted to them by depositors. We assume that the market rate of interest is 5%, so depositors will be paid \$1050 before the owners can claim their profit. The net return for the owners then is 46 = 1096 - 1050 with probability $\frac{1}{2}$, and 42 = 1092 - 1050 with probability $\frac{1}{2}$. The expected (i.e., average) return to the shareholders is $44 = \frac{1}{2} \times 46 + \frac{1}{2} \times 42$. Because they put up \$40 to begin with, their return on capital is 10%. This is summarized in Table 3.7, which shows the return to a thrift from *PI*, whether it is undertaken by a zombie or a solvent firm.

The Society column is just the total of the other columns and shows what happens to the economy as a whole. The FSLIC column gives the amount that the insurer has to pay to depositors. Of course, the prudent investment does not require any outlay by FSLIC. The owners might find *PI*'s 10% return attractive, but we have yet to compare *PI* with *WS*, the wildcat scheme, which is represented by Table 3.8. Table 3.8 gives the point of view of a zombie.

WS requires an initial outlay of \$1040, of which \$1000 is funded by deposits the same starting position as *PI*. With *WS* the high and low returns also occur with equal probability, but the investment is much riskier than *PI*. *WS* returns only \$150 (gross) with probability $\frac{1}{2}$, although it will yield \$1450 with probability $\frac{1}{2}$. The low return of \$150 does not come close to allowing the S&L to discharge

Values	Depositors	Owners	FSLIC	Society
Initial outlay High return Low return Average return Rate of return	1000 1050 1050 1050 5%	40 400 0 200 400%	0 0 -900 -450	1040 1450 150 800 -23%

Table 3.8. Wildcat scheme (WS) undertaken by a zombie

Values	Depositors	Owners	FSLIC	Society
Initial outlay	1000	40	0	1040
High return	1050	400	0	1450
Low return	1050	-900	0	150
Average return	1050	-250	0	800
Rate of return	5%	-725%		-23%

Table 3.9. WS when undertaken by a solvent thrift

its deposit liabilities. If disaster occurs, the owners must turn over all of the \$150 recovered to the depositors. This is far short of their initial deposit, so the deposit insurance kicks in and pays the remaining \$900. The owners lose all of their invested capital, but *they do not have to tap into their private wealth to pay off depositors*; the insurance fund makes up the difference. Therefore, the average return to the owners from *WS* is not

$$\frac{1}{2} \times 400 + \frac{1}{2} \times -900 = -250$$

which would be a negative rate of return of more than 700%. With FSLIC covering the shortfall when the investment turns sour, the average return to the owners is

$$\frac{1}{2} \times 400 + \frac{1}{2} \times 0 = 200$$

a *positive rate of return of* 400%. It is clear that the owners of a zombie firm will prefer *WS*, although the return to society is negative (-23%) with *WS*, and it is a respectable + 5.2% with *PI*.

The selling of naked call options on bonds is a good example of a wildcat scheme. When *A* sells a naked call option to buyer *B*, *B* has the right to purchase bonds from *A* at any time in the future, at a fixed price determined when the call option is sold. It is a naked call option if *A* doesn't actually own any bonds! This was the only "asset" of an S&L that failed after only a year in business (Milgrom and Roberts, 1992, p. 174). The *WS* is valuable to the zombie S&L only because it has no assets that can be used to honor its deposit liabilities in case the project fails. Calculation of the rate of return for the solvent S&L is quite different because it would have to reduce its asset holdings by enough to pay its depositors. Table 3.9 shows how drastically that affects the owners' rate of return. Given a choice between *PI* and *WS*, the zombie chooses *WS* but the solvent bank chooses *PI*.

Return to our examination of the zombie firm. The deposits that are used for either of

these schemes would be fresh deposits, brought in to allow the S&L to undertake new investments. If the firm has outstanding deposit liabilities that it is unable to honor and it is in danger of being shut down by the FHLBB, then the WS scheme offers the zombie S&L a last chance for financial health. In the unlikely event that the risky investments pay off, there will be plenty for everyone—depositors and owners. If they fail to pay off, the owners do not lose because the institution

4. The Savings and Loan Debacle

is already insolvent, which means that they could not recoup any of the wealth they invested in their firm, even without *WS*. Gambling for resurrection is comparable to a basketball team deliberately fouling when it is behind by three points with twenty seconds remaining in the game. If the opposing team fails to make either of its free throws, then the other team has a chance of taking the ball down court and sinking a three-point basket. The strategy rarely works, but it gives the trailing team *some* chance of staying alive by sending the game into overtime. There is a high probability that the strategy will fail, but losing by four or five points is no worse than losing by three points. There is no chance of winning without a desperation move and some slight chance with it. This logic made things tougher on responsible thrifts. The firms that were gambling heavily offered higher interest rates on deposits to attract new funds to finance the wildcat schemes. Competition forced the responsible firms to pay higher interest rates too, making them more vulnerable.

If depositors had cared how an S&L was managed, many would have accepted lower interest rates to have their money stored in a safer place. As it was, the higher interest rates even influenced the size of the national debt. This consideration apart, gambling for resurrection constitutes a significant welfare loss for consumers because valuable resources are employed in ways that yield a much lower return to society than they are capable of providing. If only one firm gambled for resurrection it would be difficult to claim that the decision was bad from an overall social welfare perspective. It might turn out very badly, but it might turn out very well. But when more than a thousand S&Ls undertake this sort of plunge, we can say that the outcome will be harmful for sure. Consider the case of 1000 firms each adopting WS, which returns 1450 with probability 1/2 but only 150 with probability $\frac{1}{2}$. The law of large numbers tells us that, with very high probability, close to 500 of the firms will see the investment turn sour. The total value of these 1000 investments will then be $500 \times 150 + 500 \times 1450 = 800,000$. Each firm began with \$1040 worth of assets, for a total of \$1,040,000. The borrowers collectively turned that into \$800,000 worth of assets, a very bad deal for society.

Now, why would members of Congress want a milder regulatory climate? We have to assume that they failed to understand the impact on the efficacy of markets when the incentive for comparison shopping is diminished. And Congress itself would have had more incentive to work at understanding the banking industry if it they had not been playing a version of the prisoner's dilemma game. To simplify, we suppose that a member of Congress simply has to choose between stringent regulation of the thrift industry and mild regulation. Consider the implications of these two strategies for the legislator's own constituency. With *WS*, U.S. taxpayers have to shell out \$900 when the scheme fails. The scheme will fail half the time, so if there is a large number of gambling S&Ls in the legislator's state the actual number of failures per investment will be close to the average. Therefore, we can assume that U.S. taxpayers have to contribute \$450 per *WS*. But only one-fiftieth of that will come out of the pockets of the legislator's none state, and when it

162 Hidden Action

fails 98% of the costs are passed on to other states. This argument may explain the temptation that induced some members of Congress to intervene in the regulatory process on behalf on local thrifts, especially when it is coupled with the intense lobbying for regulatory forbearance by the thrift industry. However, it does not fully explain the creation of a milder regulatory climate via legislation. When it comes to the framing of legislation, we must think in terms of group decision making rather than the independent individual choice that can lead to the prisoner's dilemma. (But don't lose sight of the fact the legislation results from *individual* voting behavior.)

New regulations were introduced at the end of the 1980s. One effect was to increase the deposit insurance premiums paid by individual thrifts. (These premiums are used to build up the fund that is tapped when an S&L fails and depositors have to be bailed out.) However, thrifts that take bigger risks are still not charged higher premiums. The life insurance counterpart would be to charge smokers the same premium as nonsmokers or to charge drivers who have speeding tickets and accidents on their records the same premium for car insurance as people with clean records. The careful person would be subsidizing the careless. More significantly, society would lose an opportunity to give risky decision makers incentive to modify their behavior.

There is an important hidden characteristic element to the thrift debacle. The 1982 Depository Institutions Act broadened the scope of activities available to an S&L. At the same time the thrift regulators lowered the capital-asset requirements on individual thrifts. The new regulatory climate attracted entrepreneurs who saw an opportunity to raise easy money to finance their personal get-rich-quick schemes. This is the *adverse selection* phenomenon: The incentives are such that characteristics that are least beneficial to society are selected.

In addition to exacerbating the adverse selection problem, the new regulatory environment made it easier to profit through fraud. In some cases, an *S* &L that was managed by its largest shareholder would make a loan to a friend of the manager on terms guaranteed to result in a loss to the thrift. But the borrower would make a secret payment to the manager, resulting in a net gain for both at the expense of the other owners, of course. This is referred to as *looting*, to distinguish it from gambling for resurrection, which at least offered some hope of restoring the health of the S&L.

Source

The introduction to this section is based on White (1991). Although they were not the first to highlight the critical role of gambling for resurrection, Romer and Weingast (1991) take the analysis further than others in tracing the problem back to Congress. Part of this section is based on their article.

Links

For additional discussion of the S&L debacle, see Kane (1989), Demirgüç-Kunt and Kane (2002), Milgrom and Roberts (1992, pp. 170–6), Chapter 11 in Mishkin (1992) on the crisis in banking regulation, and Litan (1991), a comment on Romer and Weingast (1991). Dewatripont and Tirole (1994, p. 95) discuss the

4. The Savings and Loan Debacle

regulators' dilemma: allow the crisis to deepen or shut down the zombie S&Ls at a time when the insurance fund was insufficient to meet all the deposit liabilities. See Mishkin (1992, p. 260) on the scandal surrounding Charles H. Keating Jr. and Lincoln Savings & Loan for a case of adverse selection. Akerlof and Romer (1994) discuss looting. Dewatripont and Tirole (1994, p. 94) touch on the indirect rescue of U.S. banks by the U.S. government and the Federal Reserve when they assisted Latin American countries. Shoven, Smart, and Waldfogel (1992) show that the increases in interest rates caused by zombie S&Ls attracting new deposits even increased the national debt.

Problem set

- **1.** How would a private insurance carrier respond to a client that always took extreme risks and frequently submitted large claims?
- **2.** Rework Tables 3.7–3.9 when the high return occurs with probability 0.25 and the low return occurs with probability 0.75. Which investment would a solvent S&L choose and which would be chosen by a zombie?
- **3.** Rework Tables 3.7–3.9 when the high return occurs with probability 0.75 and the low return occurs with probability 0.25. Which investment would a solvent S&L choose and which would be chosen by a zombie?

Questions 4 and 5 each pertain to a pair of investments, *X* and *Y*. Each investment requires a \$1000 capital outlay, \$100 of which must be funded by the owners of the S&L, with the rest coming from the cash entrusted to the S&L by depositors. An interest rate of 10% is paid on deposits. For each investment, prepare a table (similar to the ones in this section) and fill in the cells. Determine which of the pair of investments would be selected by a solvent S&L and which would be selected by a zombie firm.

4. The investments *X* and *Y* are given by Tables 3.10 and 3.11 respectively.

Table 3.10. Investment X

Return	Probability	Payoff
Low	0.4	500
High	0.6	1500

Table 3	3.11.	nvestment	Y
---------	--------------	-----------	---

Return	Probability	Payoff
Low	0.2	1100
High	0.8	1200

5. The investments *X* and *Y* are given by Tables 3.12 and 3.13 respectively.

Table 3.12. Return on investment X

Probability	Payoff
1.0	1200

Table 3.13. Investment Y

Return	Probability	Payoff
Low	0.5	600
High	0.5	2000

5 PERSONAL BANKRUPTCY

Personal bankruptcy filings in the United States have increased fourfold in the past twenty years. At present about 5% of consumer loans will not be repaid. The default of some borrowers raises the cost to those who repay their loans. (If half of all borrowers defaulted then lenders would have to double the interest rate charged to get the same return on loans as they would if there were no default.)

To the extent that default is a consequence of a loss of income beyond the control of the borrower—due to ill health or unemployment, for example—we can think of the higher interest charge as an insurance premium. Moreover, the availability of such insurance—via the right to file for bankruptcy—enhances individual welfare, just as automobile or health insurance does. And it's financed in the same way. Those who do not make a claim pay a tiny amount of money— the insurance premium or the increase in the interest rate—that's pooled and used to pay a large sum to those who do have a claim. In the case of a drastic loss of income, the claim payment is the discharge of the debt. We choose to buy car insurance because we're better off giving up the small annual fee in return for the guarantee of receiving a large sum in case of a serious loss.

If insurance against being unable to pay one's debts is a good thing, why isn't it provided by the private sector? Because if an individual's income were guaranteed by insurance there would be a severe moral hazard problem: One would have very little incentive to work effectively. An insurance contract that paid a claim only on the condition that the individual supplied appropriate effort on the job could not be enforced because there is no way for a third party—a judge—to verify the policyholder's effort level. If the contract couldn't be enforced it wouldn't be offered. If default only occurred after a loss of income due to events beyond the control of the borrower then the higher interest charge is the insurance premium, and the availability of a bankruptcy procedure enhances individual welfare. However, a large fraction of bankruptcy filings are made by individuals who have not suffered a severe financial setback. These individuals file simply because the financial benefits of doing so exceed the financial costs. How can that be?

The cost of filing for bankruptcy is the \$400 filing fee and the increased difficulty of borrowing in the future. To compute the benefit of filing we need to examine the U.S. bankruptcy law. It is a federal law, and one can file under Chapter 7 or Chapter 13. Chapter 7 leaves future

income untouched but requires the individuals to turn over their assets to their creditors—up to the value of the outstanding debts. Chapter 13 leaves assets untouched but requires the individuals to submit a plan to commit a share of future income to repay debts. However, individual states are allowed to impose asset exemptions for Chapter 7 filings. Some states exempt the entire value of one's house. Most states have some level of exemption on retirement accounts and the cash value of life insurance, in addition to the homestead exemption.

Suppose state *X* has an unlimited homestead exemption and a borrower can cash in other assets and put them into housing just before filing for bankruptcy.

6. Mandatory Retirement

Suppose also that courts do not check to see whether an individual is in dire financial straits. Then an individual in state X with \$200,000 worth of stocks and bonds, a \$300,000 house, and \$500,000 of debt can sell his assets for \$500,000, purchase a new \$500,000 house, and then file for bankruptcy. The entire debt will be discharged, and the house will not be touched. Seventy percent of all bankruptcies are filed under Chapter 7.

If the individuals who purchased automobile collision insurance took advantage of the fact that any damage was covered by the insurer and drove carelessly in parking lots, there would be a great many more dented fenders to be repaired. The overall increase in claims would increase everyone's premium. That's why the deductible is part of the insurance contract. Otherwise, individuals would

Since 1998 twenty American steel companies have declared bankruptcy. Bethlehem Steel, the second-largest U.S. steel firm, did so because it was able to receive additional bank financing and other benefits (*The Economist*, October 20, 2001, p. 62). devote far less than the efficient amount of effort to preventive care. Similarly, the fact that bankruptcy filing can be beneficial for someone who has not suffered a financial setback means that more than the socially optimal amount of "bankruptcy insurance" is supplied. Note that these *strategic* bankruptcy filings increase the default rate on loans and result in an additional increase in interest rates. One

might expect to see lenders offering lower interest rates to borrowers who waived their right to file for bankruptcy, but that waiver could not be enforced because it is contrary to the bankruptcy act. (Chapter 7 has recently been modified to make bankruptcy less attactive.)

Source

This section is based on White (1999) and Fay, Hurst, and White (2002).

6 MANDATORY RETIREMENT

Mandatory retirement is the practice of an employer preventing employees from working beyond a specified age. Employees must retire at that time, no matter how able they are to work or how eager they are to continue working. For the first three-quarters of the twentieth century, U.S. firms typically required workers to retire at age sixty-five. The 1978 Age Discrimination in Employment Act outlawed compulsory retirement before seventy. The 1987 amendment to the act eliminated the practice for most U.S. employers, regardless of the worker's age. (Coverage was extended to college and university professors in 1994.) It is doubtful that involuntary retirement at age sixty-five or seventy is discriminatory. A worker obviously cares about the entire profile of lifetime earnings, and if all workers coming on stream are treated in the same way—as far as retirement is concerned—where is the discrimination? However, we won't debate the issue. Our main purpose is to reveal an economic rationale for mandatory retirement in a society in which there are hidden action and hidden characteristic problems on the job.

6.1 Posting a bond

Hidden action problems arise whenever work is performed in a team and it is difficult or impossible to identify the contribution made by a particular member of the team, as is the case with modern manufacturing processes. In the long run, malingerers can be detected in a number of ways. However, firing malingerers when they are identified is not by itself enough to discourage shirking if workers can switch jobs with impunity. But by accepting employment in a firm that pays its workers less than the competitive wage (i.e., the value of the marginal product) in the early years and more than the competitive wage in later years the worker is in effect posting a bond. The bond is forfeited if the worker is caught persistently shirking, because he or she will then be fired. And if a worker is fired, he or she won't be around to collect the deferred pay. The boss must monitor occasionally for the threat of bond forfeiture to have force, but the existence of the threat substantially reduces monitoring costs.

DEFINITION: Compensation

The worker's compensation is his or her annual income plus other benefits such as the employer's contributions to the employee's health insurance plan.

In an economy that did not solve this hidden action problem, workers in general would perform poorly, total output would be low, and everyone's utility would be far below what it would have been if everyone had contributed more effort and had more consumer goods and services in return. How do we know? After all, an increase in individual effort involves a cost—lower leisure consumption—in addition to the benefit of increased consumption of other goods. But when workers in general have no disincentive to shirk, the cost to an individual of reducing effort is zero. But the cost to the society—reduced output of goods and services—is positive and large. When social cost pricing is not used, outcomes are typically inefficient.

Even when there is a single worker, such as a hired hand on a farm, it is impossible to determine the extent of the worker's contribution by observing output if that output is affected by random events (weather, insects, etc.) in addition to the worker's effort. Over a long period of time the law of large numbers can be used by the employer to determine the worker's average effort from average output. In other words, the worker's actions do not remain hidden in the long run, and shirking is penalized by forfeiture of the "bond." Posting a bond in the form of deferred compensation also brings the labor market closer to the efficient level of on-the-job training. An otherwise profitable investment in worker training will be unprofitable if workers leave the firm after the new skills have been acquired. This problem will be mitigated if the worker posts a bond with the firm that pays for the training.

Inefficiency can still result if effort levels are observable but not verifiable. It may be quite evident to a manager that a worker is shirking, even though the
6. Mandatory Retirement

manager is unable to prove this with objective evidence that would convince a judge or jury. In that case, it will not be possible to employ a contract that directly penalizes a worker for shirking. The contract could not be enforced because the employer could not prove in court that shirking did in fact occur. An example of an observable but unverifiable shirking is discourteous behavior by a waiter to a restaurant customer.

There is also a hidden characteristic element to the employer-employee relationship. Even if there were no shirking there would be more talented and less talented workers. Again, team production makes it impossible to identify less talented workers in the short run. These workers may know who they are but they would not voluntarily identify themselves and accept less pay. However, if compensation is below the competitive level in the early years and above the competitive level in later years a less talented worker would not accept a contract designed for a talented worker. Such a contract would be beneficial only if the worker collected the late-career high pay, but the worker would be dismissed or kept on at lower pay when it became clear that he or she were not a high-quality worker. The compensation profile can be used to sort less talented workers from talented workers even though the former attempt to conceal their identity. The compensation profile induces the less talented workers to *self-select*. This also promotes efficiency.

The fact that pay is low at the beginning of the career, when young people want to start a family and buy a home, might prevent the deferred compensation formula from persisting in equilibrium were it not for the possibility of borrowing against future income by taking out a home mortgage. Lenders will know that compensation increases over time and take that into account when reviewing the loan application.

We have argued that the standard compensation profile, paying below competitive levels early and above competitive levels later in one's tenure in the firm, has an economic rationale. However, this compensation profile would be unprofitable for employers if workers were able to collect the high late-career pay indefinitely into old age, hence the mandated cut-off age.

Now, let's illustrate with a simple model of labor supply.

6.2 The formal argument

There are two goods, leisure consumption *X* and a composite commodity *Y*, which is total expenditure on all goods other than *X*. Let *x* and *y* denote the respective amounts consumed of the two goods. The consumer has utility function U(x, y) = B(x) + y. It is assumed that MB_X , the marginal utility of *X*, is positive but diminishing. That is, $MB_X(x)$ is positive for all $x \ge 0$, but x' > x'' implies $MB_X(x') < MB_X(x'')$. The production of *Y* is represented by a production function f(E), with labor as the input, and *E* as the total labor employed (over all workers). Thus, *E* is the number of years worth of labor used in production. It is assumed that MP(E), the marginal product of labor, is positive for all $E \ge 0$, but beyond some value of labor input it is diminishing in the sense E' > E'' implies MP(E') < MP(E'') for values of E' and E'' beyond some threshold. (When discussing retirement it is convenient to measure time in years.)

We put the spotlight on a particular individual *J*, so we let L_J denote *J*'s labor supply. An individual can't control the amount of labor supplied by others, so we take that as given and denote it by L_O . Therefore, $E = L_J + L_O$ and thus $f(E) = f(L_J + L_O)$. Because we are treating L_O as a constant, we can view *f* as a function of L_J , which means that *MP* is also a function of L_J . In fact, we simplify by writing *MP*(*L*), which is the increase in output when individual *J* works one more year, given that *J* worked for *L* years and that the total amount of additional labor employed is L_O . We begin by showing that efficiency requires $MB_X(x) = MP(L)$.

If $MB_X(x) < MP(L)$ we can have individual *J* supply an additional unit of labor, resulting in the production of MP(L) additional units of *Y*, which we give to that same consumer. This will increase utility, but the net change in *J*'s utility must reflect the loss in one unit of leisure consumption. The reduction in the utility derived from leisure is $-MB_X(x)$, and hence the net change in utility is

$$-MB_X(x) + \Delta y = -MB_X(x) + MP(L),$$

which is positive when $MB_X(x) < MP(L)$. We have increased one individual's utility without affecting the utility of anyone else. The worker's extra consumption of *Y* was generated by increasing that worker's time on the job. No one else's consumption changed, and no one else's labor supply changed.

Suppose, now that $MB_X(x) > MP(L)$. Then we can increase individual *J*'s utility without affecting anyone else by increasing *J*'s leisure consumption by one unit and letting *J*'s consumption of *Y* fall by the resulting drop in output, which is MP(L) because an increase in leisure of one hour reduces labor input by one hour. Again, we made one person better off without harming anyone else. Therefore, efficiency is incompatible with $MB_X(x) < MP(L)$ and also with $MB_X(x) > MP(L)$. It follows that efficiency requires $MB_X(x) = MP(L)$ for an arbitrary individual *J*. Assume that ninety is the time endowment. That is, the individual does not anticipate living longer than ninety years. Then once we specify *J*'s leisure consumption *x* we have determined *J*'s labor supply *L*. It's 90 – *x*. Consequently, $MB_X(x) = MP(L)$ can be written $MB_X(x) = MP(90 - x)$. Let x^* be the solution of this equation. We can say that x^* is the efficient leisure consumption for *J*, and $L^* = 90 - x^*$ is *J*'s efficient retirement date. (Different consumers would have different *B* functions, and hence different efficient levels of *X*, even with the same production function.)

If in every period the worker's compensation equals the worker's marginal product in that period then that worker will choose the efficient retirement date.

(Here is the calculus derivation of x^* : The utility function of individual *J* is U(x, y) = B(x) + y. Let L_0 denote total amount of labor contributed by everyone but *J*. If the outcome is efficient it must maximize *U* given the labor supply and the consumption plan of every other individual. Therefore, we can derive a necessary condition for efficiency by maximizing

$$V(x) = B(x) + f(L + L_0) - y_0$$



Figure 3.1

where y_O is the total *Y* consumption of everyone but individual *J*. Use the chain rule and the fact that dL/dx = -1 to take the derivative of *V* with respect to *x*. The first derivative is $B'(x) - f'(L_O + L)$, and when we set this equal to zero we get equality between the marginal utility of *X* and the marginal product of *J*'s labor, and we let x^* denote the solution of that equation. We know that the first derivative will equal zero at the maximum because we can assume that x = 0 won't maximize the individual's utility nor will L = 0. That is, there will not be a corner solution.)

The efficient labor supply is $90 - x^*$, which is represented as L^* in Figure 3.1. Because MB_X falls as leisure consumption increases, when the marginal utility of leisure is plotted as a function of L (which is on the horizontal axis), it increases as L increases. MP(L), the marginal product of labor, increases early in the career, as the individual learns on the job, and then declines after a point, as age takes its toll.

If there were no hidden information problems, a compensation profile equal to the marginal product of labor schedule would induce the individual to retire at the efficient date L^* . Consider: When the compensation *C* equals *MP*(*L*), the marginal product of labor, at each date *L*, if the worker were to retire at date $L < L^*$ then $x > x^*$. An increase in *L* of one year would cause the worker's utility to increase by

$$-MB_X(x) + C.$$

(Remember, *J*'s utility function is B(x) + y, and $\Delta y = C$ if time on the job were to increase by one year.) $C - MB_X(x)$ is positive when $x > x^*$ because C = MP(L), $MB_X(x^*) = MP(L^*)$, and $MB_X(x)$ decreases as *x* increases and MP(L) increases as *L* decreases. Therefore, if *J* were free to choose, *J* would not retire

before L^* . However, if $L > L^* J$ can increase utility by reducing L by one year. The change in utility is $MB_X(x) - C$, which is positive when $x < x^*$. J would not retire later than L^* if the decision were J's to make. We have shown that if at each point in time the worker's compensation is equal to the marginal product of labor, then the utility-maximizing consumer will choose the efficient retirement date L^* .

The compensation profile represented by MP(L), the marginal product of labor, in Figure 3.1 is rarely observed. Much more typical is the upward sloping compensation profile represented by the curve *UCP*. We can find an upward sloping *UCP* such that the consumer is indifferent between *UCP* and retirement at L^* on the one hand, and on the other hand always having a compensation package equal in value to the current marginal product of labor and retirement at L^* . Just find some *UCP* with the same present value as the marginal product compensation schedule. (See the next section.) If the individual faces a different interest rate as a borrower than as a lender, then the present value calculation is somewhat misleading. But there is *some UCP* that gives exactly the same utility as the profile *MP*(*L*). The proof of that claim follows from the fact that if *UCP* is sufficiently low the individual will prefer *MP*(*L*) and if *UCP* is sufficiently high the individual will prefer *UCP*. There must be some intermediate upward sloping compensation profile to which the individual is indifferent, and this is represented in Figure 3.1.

At L^* the actual compensation (located on UCP) is above $MB_X(x^*)$ and the individual will want to keep working at the current rate of pay. An upward sloping compensation profile is in society's interest, because it helps solve hidden information problems, leaving everyone with more utility. Consequently, mandatory retirement is in society's interest because the upward sloping profile will not be offered by profit-maximizing firms if workers continue on the job beyond L^* . Because UCP and the marginal product compensation profile MP(L) have the same present value when each is truncated at L^* , the firm will prefer the marginal product schedule to the UCP schedule if the worker chooses the retirement date. The worker will choose to retire at L^A with compensation schedule UCP. Between L^* and L^A the value of compensation is above the marginal product of labor and the firm loses the compensation minus MP(L) on each unit of additional labor employed.

The overall outcome could be very unprofitable with UCP and no mandated retirement date. Therefore, the equilibrium will not include firms offering an upward sloping compensation profile without specifying the retirement date. If we look at labor supply only, we see that the equilibrium could include firms that offer the marginal product compensation profile with the retirement date chosen by the worker and contracts that offered an upward sloping compensation profile with retirement mandated at L^* . Firms that employ the latter will be more profitable because they will have fewer hidden information problems. These firms will be able to set lower prices and drive the other firms out of the market. Therefore, when we look at labor demand as well as supply, we see that the equilibrium will feature only firms that offer an upward sloping compensation profile with mandatory retirement at L^* .

6. Mandatory Retirement

Have we seen a change in U.S. compensation profiles since the end of mandatory retirement? No, because there are other reasons for requiring the employee to post a bond. Moreover, workers often choose to retire before seventy and even before sixty-five. Decades of economic growth have made that possible.

6.3 The intertemporal budget constraint

This section shows why there are many compensation profiles that provide the same level of utility to a given worker. Initially, assume that there are only two periods: period 0 (the present) and period 1 (which is one year from now). Let C_0 be the number of dollars available for consumption now and let C_1 be the number of dollars available for consumption one year from now. We assume an interest rate of *r* that is the same for lenders and borrowers. We express *r* as a decimal fraction. (If the interest rate is 7%, then r = 0.07.) To specify the budget constraint we need to know current income, which we denote by I_0 , and income one year from now I_1 .

To derive the intertemporal budget constraint put yourself in the position of the consumer one year from now, and ask simply, "How much money can I spend on goods and services in period 1?" If the consumer saved in period zero there will be

Savings + Interest on Savings + Period 1 Income.

Saving is, by definition, equal to the amount of income not spent on consumption. Therefore, saving equals $I_0 - C_0$. Interest earned on saving is the amount saved multiplied by the interest rate, which is $(I_0 - C_0) \times r$ in this case. Therefore, the amount that a saver can spend on consumption in period 1 is

$$I_1 + I_0 - C_0 + (I_0 - C_0) \times r = I_1 + (I_0 - C_0)(1 + r).$$

Therefore, a saver is constrained by the following equation in period 1:

$$C_1 = I_1 + (I_0 - C_0)(1 + r).$$

What about someone who borrows initially? How much money can someone who borrowed in period 0 spend in period 1? The answer is clearly

Period 1 Income - Amount of the Loan - Interest on the Loan.

The principle has to be repaid in period 1 in a two-period model, and so does the interest on the loan. It is easy to determine the amount borrowed; it will be equal to the amount spent on consumption in period zero in excess of period zero income. That is, borrowing $= C_0 - I_0$. The interest charge is the interest rate times the amount of the loan, or $(C_0 - I_0) \times r$. Therefore, the amount that a borrower can spend on consumption in period 1 is

$$I_1 - (C_0 - I_0)(1 + r) = I_1 + (I_0 - C_0)(1 + r).$$

Therefore, borrowers and savers are governed by the same intertemporal budget constraint:

$$C_1 = I_1 + (I_0 - C_0)(1 + r).$$
^[1]

Of course, if the individual neither lends nor borrows in period 0 we will have $C_0 = I_0$ and hence $C_1 = I_1$, which also satisfies [1]. Therefore, [1] is *the* two-period intertemporal budget constraint.

Suppose the individual will live T + 1 periods. We claim that the individual's consumption opportunities are governed by

$$C_T = I_T + (1+r)(I_{T-1} - C_{T-1}) + (1+r)^2(I_{T-2} - C_{T-2}) + (1+r)^3(I_{T-3} - C_{T-3}) + \dots + (1+r)^{T-1}(I_1 - C_1) + (1+r)^T(I_0 - C_0).$$
[2]

To prove that [2] is the correct representation of the constraint that the market places on the individual's lifetime consumption plan ($C_0, C_1, \ldots, C_{T-1}, C_T$) we suppose that we have already proved it for T = t. We then demonstrate that that supposition implies the claim for T = t + 1. Because we have already established the claim for T = 1, we will then have proved by induction that [2] holds for any finite number of years.

If the right-hand side of [2] is the available purchasing power in period t, given the previous consumption levels $C_0, C_1, \ldots, C_{t-2}, C_{t-1}$, then the amount of purchasing power R_t left over after C_t is spent in period t is the right-hand side of [2] minus C_t . That is,

$$R_{t} = I_{t} - C_{t} + (1+r)(I_{t-1} - C_{t-1}) + (1+r)^{2}(I_{t-2} - C_{t-2}) + (1+r)^{3}(I_{t-3} - C_{t-3}) + \dots + (1+r)^{t-1}(I_{1} - C_{1}) + (1+r)^{t}(I_{0} - C_{0}).$$

If R_t is positive it will add to the individual's purchasing power in the next period. In other words, saving is carried over to the next period, with interest of course. If R_t is negative, debt is carried forward to the next period and will have to be paid back, with interest. In either case, the purchasing power available in the next period is $I_{t+1} + (1 + r)R_t$. When we set $C_{t+1} = I_{t+1} + (1 + r)R_t$ we get [2] for T = t + 1. Therefore, [2] is the intertemporal budget constraint for any lifetime T + 1, for any value of T.

If we divide both sides of [2] by $(1 + r)^T$ and gather the consumption terms to the left of the equality sign we get

$$C_{0} + \frac{C_{1}}{1+r} + \frac{C_{2}}{(1+r)^{2}} + \frac{C_{3}}{(1+r)^{3}} + \dots + \frac{C_{T-1}}{(1+r)^{T-1}} + \frac{C_{T}}{(1+r)^{T}}$$

= $I_{0} + \frac{I_{1}}{(1+r)} + \frac{I_{2}}{(1+r)^{2}} + \frac{I_{3}}{(1+r)^{3}} + \dots + \frac{I_{T-1}}{(1+r)^{T-1}} + \frac{I_{T}}{(1+r)^{T}}.$ [3]

We refer to [3] as the *present value* form of the intertemporal budget constraint. The right-hand side is the present value of the income stream (I_0 , I_1 , ..., I_{T-1} , I_T), and the left-hand side is the present value of the consumption stream (C_0 , C_1 , ..., C_{T-1} , C_T). Clearly, there are many different income streams that will produce the same number on the right-hand side of [3]. All such streams provide the consumer with the same consumption opportunities. If two income streams *A* and *B* have the same present value, then a consumption



Figure 3.2

stream will be affordable with *A* if and only if it is affordable with *B*. Suppose *A* delivers high levels of income in the early years and relatively low levels later on. If it's the other way around with *B* but they have the same present value, then by borrowing and lending the consumer can finance a particular consumption stream with *A* if and only the consumer can finance that consumption stream with *B*. Consequently, a particular consumer will wind up with the same utility with either income stream.

Source

The economic rationale for mandatory retirement is based on Lazear (1979).

Links

See Carmichael (1989) for more on this problem. The mandatory retirement story doesn't fit U.S. data perfectly. See Stern and Todd (1992). For example, pension funds should be included in the model because they also play the role of bonds posted by the employees. See Lazear (1992). Since mandatory retirement was outlawed in 1978 for U.S. workers under the age of seventy, the increase in the average retirement age has been slight (Costa, 1998, p. 24).

Problem set

All of the questions refer to the Figure 3.2.

1. What is the efficient retirement age if *H* is the value of the worker's marginal product as a function of time *K* is the marginal utility of leisure, and *J* is the compensation profile?

- **2.** What is the efficient retirement age if *H* is the value of the worker's marginal product as a function of time *J* is the marginal utility of leisure, and *K* is the compensation profile?
- **3.** What is the efficient retirement age if *K* is the value of the worker's marginal product as a function of time *J* is the marginal utility of leisure, and *H* is the compensation profile?
- **4.** What retirement age would the worker choose if *H* is the value of the worker's marginal product as a function of time *J* is the marginal utility of leisure, and *K* is the compensation profile?
- **5.** What retirement age would the worker choose if *H* is the worker's compensation profile, *J* is the value of the worker's marginal product as a function of time, and *K* is the marginal utility of leisure?

7 TENURE AND THE PERFORMANCE OF PROFESSORS

This section briefly considers a hidden action problem in which the agent who is carrying out a task for the principal is a university professor. The spotlight is on the professor's hiring, promotion, and sometimes dismissal. The promotion regimen employed in colleges and universities in Canada, the United States, and many other countries is an example of the *up-or-out* policy, which is also used in most firms that are organized as partnerships. After a probationary period of six or seven years, the employee is either given permanent employment or is released. But why up or out? If workers are found to be of low quality, why not offer them a lower wage? Why terminate employer giving the worker a false low rating to cut labor costs. If other firms could observe the worker's quality, this wouldn't work. But because information is hidden it would be a serious possibility and a serious problem: If the university administration were to systematically underrate professors, they might respond by working less, because the return to hard work is reduced.

In the case of colleges and universities, a professor's performance is reviewed after six years by members of his or her own department. Outside evaluations of the candidate's research are obtained. Evidence of teaching effectiveness is also examined, but the candidate's contributions to scholarship receive almost all of the weight in the top research universities. If the decision is negative, the teacher must leave the university. Even if he or she offers to stay on at a big cut in pay, the teacher will not be retained after a negative tenure decision. If the decision is favorable, the professor is granted lifetime tenure in the department. This means that the professor can never be fired for incompetence—only for moral turpitude. The only other way that a university can dismiss a tenured professor is to close down the entire department. Assuming no serious moral lapses, a tenured professor has a lifetime job; his or her position in the department is terminated only by death—the professor's death (or retirement) or the department's, whichever comes first.

7. Tenure and the Performance of Professors

What benefit could come from a policy that prevented an employer from firing a worker for incompetence? Given the roster of faculty members in a particular university department, it would seem to be in society's interest to allow the employer to fire low-quality workers at any time and replace them with higher quality faculty. The university would decide the relative weight to place on teaching and research and then rank the members of a department. The poor performers could then be identified and released. However, if this policy were adopted by universities, the departments themselves would make different hiring decisions in the first place. The established department members would be reluctant to hire the best young people available. These high-quality young people might outperform the original department members by such a wide margin that the university would want to reduce the pay of the incumbents or even fire them. Therefore, there would be a strong incentive for departments to hire low-quality newcomers. The net result of allowing the university to fire low-quality professors at any time and replace them with higher quality faculty would be departments with very low-quality workers! The overall result of abolishing tenure could well be lower quality colleges and universities. But why let the department members themselves hire their new colleagues? Because the members of a particular department are the only people in the university community capable of judging the candidates for an opening in that department.

Compare the university system with professional sports: In both cases performance declines significantly with age. The public has an interest in seeing that workers are replaced when the quality of their performance falls below that of newcomers waiting in the wings who not only have the benefit of youth but also the most up-to-date tools. In athletics, management *can* replace one worker with a superior one—it is relatively easy for management to evaluate new talent. There is no efficiency argument for awarding tenure to professional athletes. The case for academic tenure does not apply here. In academe the weak performers can be identified by the administration with the passage of sufficient time—a decade or two, say. Why not allow the university to replace them at that point? We come back to the hiring decision. Professors will have a strong incentive to hire weak newcomers if there is a possibility that a strong candidate could eventually replace the incumbent.

How about comparing university hiring with the way it's done in the legal profession? The senior partners in a law firm do not seem to be reluctant to hire the best young people. But this is a field in which newcomers bring revenue to the firm in addition to talent. The better the lawyers the more their clients will be billed for their services. The new lawyers create room for themselves without displacing the incumbents. A university department, on the other hand, has a limited number of positions. In a law firm, a bad hire will diminish the firm's revenue and hence the income of those making the hiring decision. In a university, a bad hire may eventually diminish the quality of the students admitted, but if the original department members get to keep their jobs by making bad hires they will realize a net gain. They may notice the deterioration in student quality, but that is more than compensated for by the increased probability of holding onto their jobs.

Granting lifetime job security to professors may be necessary to ensure that they do not have an incentive to hire weak newcomers. But that should not prevent the university from periodically examining the performance of a faculty member to ensure that the individual's *own* rate of productivity is maintained or from having departments compete for their shares of the pay raise pot.

Although lifetime tenure is desirable, why base teachers' promotion decisions even partly on their research output? If the creation of new knowledge is *not* an important part of teachers' jobs, as in the case of preuniversity education, then the instructor is just passing on the discoveries of others. New hires can be evaluated by the administration and the case for tenure vanishes. At the university level, research is an important part of the professor's job—not just because it adds to the stockpile of knowledge. Professors who are not sufficiently interested in their subjects to go beyond what is already known will probably be less than inspiring in the classroom. Moreover, if the instructors are not engaged in intensive research it is very hard to for them to pick up the new tools that allow them to pass breakthroughs on to students.

In general, writing a paper for publication in a leading scholarly journal requires skills that are closely related to the talents needed for effective teaching at the university level: intelligence, thorough knowledge of one's field, intellectual discipline, creativity, and an interest in the subject. Less talented scholars take a lot longer to prepare an article that's suitable for the high-quality journals. Therefore, it is less costly for the high-quality workers to signal their qualityin this instance, a signal is publication of an article in a high-prestige journal. Basing hiring, firing, and promotion decisions heavily on the individual's publication record has a rationale as a partial solution to a principal-agent problem, which in this case has both hidden characteristic and hidden action elements. The university could not take job applicants' word that they are diligent scholars with a keen interest in the discipline, a determination to work long hours learning more about the subject, and the intelligence to keep up with the other scholars in the field. Even if the university and society in general had no interest in scholarly research there would be a signaling rationale for using publication records in employment decisions. (Of course, if research had no social value it would receive much less funding and there would be less of it.) If you wish to test the proposition that publication has its uses apart from the scientific value of the output, you need to go well beyond estimating the correlation between teaching ability and success in publishing that you observe at your own college. You want to compare the present situation with what you would expect to find if universities were unable to use a fairly objective quality signal such as publication.

Source

This function of the up-or-out contract to solve the problem of the principal (employer) falsifying information was pointed out by Kahn and Huberman (1988). The economic argument for awarding tenure to professors is based on Carmichael (1988).

8. Pay and Performance in U.S. Public Schools

Links

Carmichael (2001) explains why professors' unions are highly undesirable if professors also have tenure.

8 PAY AND PERFORMANCE IN U.S. PUBLIC SCHOOLS

American public schools have a dismal reputation. American *universities* are the envy of the world. Students from around the world come to the United States for their postsecondary education. The situation for grades 1 through 12 is remarkably different. In 1995, American seventh- and eighth-grade students ranked twenty-third in mathematics and twelfth in science out of the forty countries involved in the third International Math and Science Study. (Ireland ranked seventeenth in math, and Canada ranked eighteenth.) Moreover, the standing of American students is lower at the higher grade levels. High school seniors ranked below those of every country except Lithuania, Cyprus, and South Africa. In addition to these data we have a steady stream of media reports of egregious conditions in public schools.

Between two-thirds and 75% of the world's top research universities are located in the United States (Rosovsky, 1990, Chapter 2). In 1997, 29% of the PhDs awarded by American universities were earned by noncitizens, and 43% of the degrees in mathematics and computer science went to noncitizens (Ehrenberg, 2000, p. 4). Contrast those data with the reputation of U.S. schools in the grade range 1 through 12: 67% of low-income parents say that they would be inclined to take their children out of the public school system if the alternatives were not so costly (Moe, 2001).

This section argues that the way that teachers are paid in the public school system creates serious moral hazard and averse selection problems that exacerbate the performance problem—and may be the main obstacle to correction. (There is strong evidence that increasing expenditure on education will not lead to improved student performance.) We begin with the hidden action element.

In the vast majority of school districts in the United States, teachers' pay depends on the number of years of college attained, the number of graduate courses taken, and especially the number of years they have been employed as teachers. The quality of the colleges that the teachers attended is irrelevant in determining their salary, as is the nature of the courses taken.

The teacher's performance in the classroom is *not* part of the salary formula. The largest teacher's union, the NEA (National Education Association), has always resisted—very successfully—the idea that good performance be rewarded and bad performance be punished. This creates a severe moral hazard problem: Two things that have a profound effect on the performance of workers in other sectors of the economy—the carrot and the stick—are not employed in the public school system. Instead, teacher pay is determined by factors that have little bearing on the quality of teaching.

Most professionals—physicians, professors, public school teachers, lawyers, and so forth—claim to do the best job that they can as a matter of pride and

ethics, whether or not that is reflected in their pay. There is strong evidence to the contrary. For instance, physicians recommend more expensive treatments in cities with more doctors per capita. (See Section 9.1 of this chapter.)

In addition to the moral hazard problem that is created when pay is independent of performance, there is a serious adverse selection problem. Highly motivated, hard working, and talented individuals are discouraged from entering a field that does not increase the person's pay when those qualities are manifest in superior performance. (The educational quality of U.S. teachers has been steadily declining over the past quarter century.) The problem is compounded in the public school system because science and math teachers are paid no more than others, although they receive an 8% bonus in the private sector. Many school districts are forced to hire unqualified math and science teachers.

The core of the NEA's attack on pay-for-performance is that good performance by teachers is very hard to measure. However, it can't be any harder to measure the quality of teaching by American college and university instructors, although merit pay is a crucial part of their salary formula. In fact, 90% of large public and private sector organizations attempt to measure the quality of an individual's work and adjust pay accordingly, in spite of the measurement problems. Inequities are inevitable, but there is far more harm done by the inefficiency of a pay schedule that includes no incentive for good work.

The performance of American students on international tests, the argument that moral hazard and adverse selection problems are built into the pay formula of American teachers, and the groundswell of parental dissatisfaction with the performance of public schools have prompted the NEA to propose an additional program of certification. Mastery of certain skills and/or knowledge would be certified and a teacher's pay would increase with the number of certificates presented. However, less that three-tenths of 1% of North American firms use certification in that way. The NEA proposal is not an improvement on the present system but just more of the same. It is noteworthy that only 1% of private school teachers have the certification demanded by public school boards.

How do we know the system would respond to a change in the incentive environment? One type of evidence comes from a comparison of the performance of public schools that face serious competition with schools that do not. The city of Boston has seventy school districts accessible from the city center within half an hour, whereas Miami has but a single district. Some public schools face competition from relatively inexpensive Catholic schools, and many do not. Some school districts employ the voucher system or have a charter schools program, both of which provide stiff competition for the local public school. Public schools that face competition perform better than those that do not, after adjusting for factors such as the level of parental education and income that would otherwise cloud the results. Competition provides disincentive to the public schools authorities to stand pat, for fear of losing enrollment and then government revenue. Presumably, one of the consequences is the provision of better incentives for teachers.

Concluding note on the voucher system and charter schools: The voucher system gives parents the right to transfer the amount of money that would have

been given to a public school for the child's education to the private school of their choice. Any difference between the private school fee and value of the voucher comes out of the parents' pockets. The system greatly expands the family's range of choice, especially if the difference between the private fee and the value of the voucher is small. The NEA claims that vouchers will drain the public school system of the best students and consequently further disadvantage those left behind. That has not happened. In Michigan and Milwaukee the poor and minority students remaining in the public system have made impressive gains—presumably as a result of the schools responding to competition. And the fraction of poor students in the public system has not changed. Charter schools are largely publicly funded but have considerable autonomy and find it much easier to respond to parental concerns.

Source

The lack of a meaningful link between teacher pay and teacher performance has been thoroughly studied by Dale Ballou and Michael Podgursky. See for example, Ballou and Podgursky (1997, 2001). Data on the standing of U.S. students in international tests is taken from Hanushek (2002) and Woessman (2001). The effect of competition on public schools has been intensively researched by Caroline Hoxby (2001a, 2001b, 2002).

Links

See Hanushek (2002) for a review of the evidence revealing that increasing expenditure on education will not lead to improved student performance. Lazear (2003) points out that the educational quality of U.S. teachers has been steadily declining over the past quarter century.

9 MORAL HAZARD AND INSURANCE

The term *moral hazard* was first used in the insurance industry to refer to the fact that individuals with insurance coverage have diminished incentive to devote effort to preventive care. Preventive care reduces the probability of the kind of accident that is covered by insurance. This is a concern for insurance companies because diminished preventive care results in a larger number of accidents and hence more claims paid by the insurer. It is a concern for society as a whole because, although insurance coverage increases individual welfare, it also induces individuals to devote less than the *efficient* amount of effort to preventive care.

Effort is costly to the individual, and efficiency calculations always require benefits to be weighed against costs. Why would expected utility-maximizing individual decisions not lead to an efficient outcome? After all, this book does not consider the effort that people devote to vacuuming their carpets to determine whether an efficient outcome results from self-regarding individual decisions. That's because when I vacuum my carpet there are no direct effects on the welfare of anyone else. Admittedly, there is an indirect effect on your welfare. The electrical energy that I used might have been used instead by you to prepare your dinner. But I pay a price for the electricity that I use, and that price reflects the value of electricity to you.

Now, return to the case of accident insurance. I could lower the probability of an accident by increasing the effort that I devote to prevention, but the cost to me of the extra effort is not accompanied by an increased benefit. If I do have an accident, the loss will be financed by the other policyholders. Their premiums provide the money with which the insurance company pays my claim. But every policyholder has diminished incentive to invest in prevention, and that increases the total number of accidents and the total value of claims paid. That in turn results in a higher insurance premium. No *individual* can reduce his or her premium by investing in prevention, however. It is easy to construct examples in which everyone would have been better off if each devoted more effort to preventive care, although no one has an incentive to do so. That is precisely what we do in Sections 9.3 and 9.4.

DEFINITION: Moral hazard with insurance coverage

Moral hazard refers to the fact that insurance coverage drives a wedge between the net benefit to the individual and the net benefit to the society when the individual acts to reduce risk. The former falls far short of the latter. Contracts that condition a claim payment on the individual's actions cannot be enforced because the amount of effort devoted to preventive care could not be verified in court.

You may feel that an individual has sufficient incentive to invest in preventive care even with insurance coverage when there is also the potential for personal injury or even loss of life—burglary, fire, automobile insurance, and so forth. Don't jump to the conclusion that a person would employ every available device for minimizing the chance of accident and injury, independent of any financial incentive. You probably drive a car that is not as safe as a more expensive car that you might have purchased instead—perhaps with the aid of a car loan. You chose a less expensive car because, even after factoring in the probability of an accident and injury, you have higher expected utility with that vehicle and a larger basket of other goods and services. It is obviously not in our interest to spend all our money, or all our time, on preventive care. If we did, each household would want to live next to a hospital, and no one would ever take a vacation because the money saved on vacations could be devoted to increased fire protection for the home. Why not hire a night watchman for your home to reduce the probability that you will die in your bed in a fire?

When drivers of police cars and rescue vehicles are monitored by means of devices similar to the "black box" (flight data recorder) installed on commercial aircraft, the frequency of accidents goes down dramatically, giving us additional evidence that individuals left on their own do not devote maximum effort to preventive care—not even when life and limb are at stake. (Monitoring is effective

even when it takes the relatively primitive form of a loud noise going off inside the vehicle when speed is excessive.)

We first look at some examples of moral hazard and then offer a formal model (in Section 9.2). Sections 9.3 and 9.4 calculate the equilibrium level of prevention resulting from individual choice and then the efficient level of preventive care. We see that the latter is substantially higher.

9.1 Overview

The most striking example of moral hazard is the case of an individual who commits suicide so that his family can collect the life insurance benefits. This possibility is in fact eliminated by the insurance contract, which releases the insurance company from its obligation to pay when death is the result of suicide. (That provision usually lapses a year or two after the insurance is purchased. Why?) It is costly for the company to determine if the insured did commit suicide, but the costs are typically small relative to potential claim.

In many cases the costs of verifying moral hazard are too high for it to be part of the contractual relationship: Homeowners insurance pays the cost of replacing objects stolen when your home is robbed. The most severe loss is sometimes the utility destroyed when an article with extremely high sentimental value but low market value is taken. Why can't the policyholder be compensated financially for the loss in sentimental value? Because there would be no way for the insurance company to verify that it would take \$25,000 to compensate for the loss of great grandmother's button collection. It would be extremely costly and in most cases impossible—to determine if an object really was treasured by the policyholder.

Why can't you buy insurance to protect against a loss of home equity should the market value of your house fall below the price you paid for it? Because of the extreme moral hazard. The insurance would all but eliminate the incentive to keep your house in good repair. It would also diminish the incentive to work hard to get a good price when selling it. (Writing the insurance contract so that a claim is paid only when the home owner has maintained the house well and fought to get the best price wouldn't work. Why?) However, basing the coverage on the average value of houses in the surrounding neighborhood will restore appropriate incentives: If the average value falls by 10% then you can claim 10% of the original purchase price of your house when you sell it. If the owner has actually increased the value of the house through maintenance and renovation then the owner will realize the fruits of that effort because the claim is based on the neighborhood average selling price. This type of equity insurance is only available in a few areas in the United States at present, but it will likely become commonplace.

"Christopher and Laurie: I had to go out for an hour. The key is under the mat. Make yourself at home." Would you leave this note on your front door if you weren't covered by burglary insurance? Some homeowners with fire insurance will burn leaves in the driveway but would not do so if they were not insured against fire. I once went to the emergency ward of an Ottawa hospital to get a prescription for our son's medication. We had forgotten the medicine at home in Toronto. The hospital visit was covered by insurance. We're not that careless in Virginia because our present health insurance provider would not reimburse us for that kind of "emergency," and rightfully so. My hospital trip could have been avoided by taking a tiny amount of preventive care-checking to make sure we had the medicine before leaving home. Sensible insurance coverage would not allow me to pass the costs of that visit on to the rest of the community. I want to make it clear that I'm not proud of this example, and also that the Canadian health insurance system produces better overall outcomes than the American system. In Canada, health care insurance is provided to everyone by the government and funded through income taxation.

Health insurance is fraught with moral hazard. I don't mean to say that people who have health insurance allow their health to deteriorate. But given that one has a health problem, there are often a number of ways of successfully treating it. If the alternative methods impose different burdens on the community's resources and these social costs are not reflected in the *private* costs incurred by the individual making the decision then the private decisions will not contribute to efficiency.

What exacerbates moral hazard in health care is that the key decisions are usually made by a third party—the physician. Doctors know that the patient will pass the costs of health care onto the insurance carrier *and* that patients typically have almost no expertise in determining the appropriate treatment of their condition. Economists use the term *induced demand* to refer to a treatment prescribed by a physician that does not benefit the patient but which augments the physician's income. It is difficult to interpret the data on inducement. If patients receive more medical treatment in areas with a

high ratio of doctors per capita, is this because doctors have fewer patients and regulate their incomes by prescribing unnecessary procedures or is it the case that communities with a high demand for medical care attract more physicians per capita?

There tends to be a higher frequency of baby deliveries by Caesarian section in communities that have more obstetricians relative to the number of women of child-bearing age (Gruber and Owings, 1996). In the 1980s escalating heath care costs due in part to the moral hazard and perceived induced demand—motivated insurance companies to play a small role in the selection of the method of treatment. Until then, the typical scenario was that the physician would recommend a course of treatment, the patient would approve, and the insurance company would pay whatever costs were incurred. Patients still

had little incentive to shop for the least expensive provider of a specific treatment or to elect a simple procedure when a more complicated one has been urged by the doctor. This remains true today, and it can lead to the doctor overprescribing medical care. According to one medical study, 20% of the heart pacemaker implants in the United States were not endorsed in retrospect as the most appropriate treatment, and 36% of the implants were recommended on the basis of an extremely optimistic forecast of expected benefits (Hsiao, 1988).

Medical practitioners in Western Europe often rely on drug therapy to treat heart disease and Americans are more likely to recommend surgery. Of course,

surgery is more expensive. Americans spend far more per capita on health care than Canadians but have about the same health status. (Canadians have a slightly higher life expectancy in fact. And virtually all Canadians are covered by some form of comprehensive health insurance, whereas 15% of Americans are not covered at all.) And because the direct cost to the recipient of medical services is typically very low, hospitals have very little incentive to compete on the price dimension. They tend to appeal to consumers by publicizing the acquisition of high-cost, high-tech equipment, even when it has little overall effect on the community's health status. The equipment does have a big impact on health care costs, of course.

Insurance companies try to mitigate moral hazard by requiring the insured party to pay a small fraction of the loss. Very often the patient will have to pay 20% of the health care costs—the *copayment*—while the insurance company pays 80%. This lightens the patient's financial burden considerably—compared to someone without any health care insurance—but at the same time imposes a charge on the patient that is proportional to the social cost of medical care. This makes it costly for individuals to incur expenses that add little to their utility. If medical care is free to individuals then they have an incentive to consume any health care service as long as it adds *something* to their utility, regardless of the cost to society.

An influential study by the RAND corporation tracked more than six thousand individuals. Some of them received free medical care and the others were charged significant copayments. The two groups were equally healthy after five years, in spite of the fact that the group receiving free care incurred 30% more treatment costs. (See Dranove, 2000, pp. 30–1, for a discussion of this research.) Deductibles also diminish the gap between private and social costs for small losses. The deductible clause makes the insured party liable for any expenses under the deductible limit, which is usually around \$200 for automobile collision coverage. The individual is protected against big losses, which is really what one needs, but for small losses, which are often the ones that are most easily avoidable, the individual suffering the loss is the one who pays. This means that private costs are equal to social costs for small losses—that is, for losses below the deductible limit. In addition, the

deductible is used by the insurance carrier as a screening device. If two policies are offered, one with a low premium and a high deductible and one with a high premium and a low deductible, the people who know they are good drivers will choose the former. This menu also provides some incentive for motorists to improve their driving habits. The high-deductible, low-premium policy could provide drivers with more expected utility *if* they drive safely.

Automobile insurance companies use experience rating to encourage careful driving. Drivers pay higher premiums if they have speeding tickets or accidents on their records. Some companies won't accept business at any price from drivers with very poor records. This gives individuals direct financial incentive to take preventive care. Health insurance companies use experience rating to determine the amount of premium paid by firms that purchase group policies. Firms in industries in which the incidence of AIDS is unusually high sometimes cannot purchase health insurance at all. If insurance companies were able to sort us into risk categories with perfect precision and charge higher premiums to individuals in higher risk groups they would do so. Any firm that didn't do this would not be very profitable if the other firms did sort. And if other firms did not sort, then a typical firm could increase its profit by sorting according to risk.

Up to a point, sorting by risk is socially beneficial because it reduces the moral hazard problem. But if it is taken too far then each risk category cantains relatively few individuals, and the law of large numbers will not apply. That diminishes the social benefit of insurance. In a large pool of insured individuals the number of accidents varies little from year to year. (If twenty classmates each tossed a coin 1000 times, there is a very high probability that very close to 10,000 heads would be recorded.) Therefore, the premium can be more or less constant and still generate just enough revenue for the insurance carrier to pay off on claims. In a small pool, the number of accidents would vary considerably from year to year, requiring significant changes in the premium from one year to the next. The individual is less insulated against risk. If everyone belongs to a small pool then everyone's expected utility could be increased by aggregating many of the pools.

However, if insurance companies did not sort into risk categories they would face a serious adverse selection problem. Consider health insurance: If all policyholders paid a common premium the most healthy of them might find that their expected utility was higher without insurance (or with a small amount of coverage). Then the remaining policyholders would have a higher probability of submitting a claim, and the premium would have to rise to cover the value of claims paid. In that case, healthy policyholders who benefitted from insurance under the lower premium might find that *their* expected utility is now higher without insurance. When they opt out, the riskiness of the remaining group increases yet again, resulting another increase in claims. And so on. This unraveling is prevented by *group insurance* coverage, which requires a participating firm to enroll all of its employees.

Experience rating and risk sorting for health care insurance can go well beyond a due consideration of incentives. On one hand, it is in society's interest to make individuals who choose to smoke pay higher health insurance premiums. On the other hand, an individual with a genetic predisposition to breast cancer should be treated as a victim of bad luck, rather than as someone who has made unwise choices. "Genetic testing may become the most potent argument for state-financed universal health care" (*The Economist*, October 19, 2000, cited in Wheelan, 2002, p. 90). Society should insure risks over the individual's lifetime, but at the same time charge higher premiums to individuals who are in a higher risk category *because of behavior over which the individual has control*.

If one were able to get more than 100% fire insurance coverage, the moral hazard problem would be particularly acute: If the building were completely destroyed by fire then the value of the insurance claim would exceed the market value of the building. The owners would have a strong financial incentive to torch their own buildings. This would definitely affect the probability of a loss

and would have a big impact on the size of fire insurance premiums. Insurance companies will not give more than 100% coverage.

Very often the market system provides its own solution to a hidden action problem. We have discussed the example of taxi fares (Section 2 of Chapter 1). Health maintenance organizations (HMOs) came into prominence in the 1980s and 1990s in response to rapidly rising health care costs. HMOs provide comprehensive medical care to the individual in return for a fixed annual fee. The HMO monitors costs-and hence claims-by giving the physician a strong financial incentive to keep the patient in good health, in part by heading off problems before they require expensive treatment by specialists. The pay of a physician under contract to an HMO has two components. First, the HMO pays the doctor a fixed monthly fee for each patient registered with that doctor. Second, there is an adjustment based on the frequency with which the doctor's patients visit specialists or hospitals: The doctor is given a monthly allowance of F dollars. The HMO reduces that allowance by C dollars for every such visit by a patient. At the end of the month the doctor is paid a bonus equal to Fminus all these deductions. This may be a negative number, in which case the physician pays that amount to the HMO. This discourages the physician from making too many referrals to specialists. But it also discourages the doctor from delaying a vital referral-the illness could become more severe and require more expensive treatment.

Are patients getting lower quality care under HMOs? The evidence is mixed. Has the HMO system (and other managed care programs) had a mitigating effect on U.S. health care costs? There was indeed a drop in the rate of growth of health care expenses in the early 1990s, so that it was roughly the same as the rate of growth of the U.S. gross domestic product. However, by 2000 the differential was again positive—and widening (Reinhardt, Hussey, and Anderson, 2004).

The United States is the only major industrial country without publicly provided universal health care insurance. Nevertheless, administrative costs account for 24% of total U.S. health spending. Astonishingly, administrative expenses for health insurance in the private sector are 150% higher per dollar of coverage than for public programs (Reinhardt, Hussey, and Anderson, 2004). Sometimes the government can nudge consumers toward the efficient effort supply. A law requiring insurance companies to give a premium discount if a silent alarm is installed can enhance social welfare. Consider burglary insurance. A homeowner with insurance is less inclined to check that the windows are locked before leaving the house and certainly less likely to install an expensive security device unless the insurance contract provides some inducement. Even the *type* of security device has efficiency implications. Some provide protection for others, and some shift criminal

activity *to* others. On one hand, if bars are placed on the windows, burglars will pass up that house and move on to another one. On the other hand, a silent burglar alarm, which rings in the police station, may discourage thieves from attempting to rob any house in the neighborhood because they will not know which houses have silent alarms. If a thief breaks into two houses per week in a particular neighborhood, then even if 1% of the houses have silent alarms the probability is 0.65 that he will be caught before the year is out. That is, the

probability is only 0.35 that all the houses he enters will not have a silent alarm. If 5% of the houses have a silent alarm then the probability of not entering a house with an alarm in 104 burglaries is 0.005.

These calculations show that invisible security devices provide a substantial spillover benefit to individuals in addition to the one installing the device. Suppose that the homeowner has a choice between visible security that costs \$200 and an invisible device that costs \$400 but provides a total of \$4000 of benefit to everyone, including the individual who installs it. If the visible device merely shifts criminal activity to others then it provides a net social benefit of zero but has a positive social cost. The invisible device adds \$3600 to social benefit, net of cost. The individual has a strong incentive to purchase the cheaper, crimeshifting technology, and thus society has an interest in promoting the invisible, crime-reducing technology.

In fact it has been estimated that a \$400 investment in the silent car alarm Lojack results in an average \$4000 reduction in losses due to automobile theft. One alternative is the much less expensive Club, which attaches to the steering wheel and primarily shifts crime to others because it is visible to the thief. Automobile theft in Boston, Massachusetts, has fallen by 50% since the enactment of a state law requiring insurance companies to provide a 25% discount to any policyholder with Lojack. (Why haven't insurance companies introduced the discount on their own? If an insurance provider has only a small fraction of the business in a neighborhood then only a small fraction of the claims saved by a silent alarm would have been paid by that company. But surely a number of firms have a large enough share of the business for the discount to precipitate an increase in profit.)

Many insurance companies have stopped covering Hollywood films. The insurance usually takes the form of underwriting the loans used to finance the movie's production. Such coverage diminishes the incentive to control spending on the projects (*The Economist*, March 31, 2001, p. 71). We conclude with a moral hazard story from a very different industry. In the 1940s American movie producers began giving major stars a share in the profits from their movies. This gave the stars an incentive to avoid the silly temper tantrums that cause production delays and escalate the costs of the movie. When profit sharing became a common practice some movie producers began disguising the profit earned by the most lucrative movies—for

instance by charging them with some of the fixed costs, such as set construction, from other projects. Contracts that offered the performers a cut of the profits were less rewarding as a result, and many responded by holding out for a percent of the *gross*—that is, they demanded a cut of the picture's revenue instead of its profit.

9.2 The formal model

We investigate the conditions for efficiency when there is moral hazard by means of a simple model. There are only two commodities, *W*, wealth, and *L*, leisure. Preferences are quasi-linear, and thus

$$U(w, \ell) = B(w) + \ell,$$

where B(w) is the utility of w dollars of wealth, and ℓ is the amount of leisure consumed. We assume that the individual is risk averse, so that the marginal utility of wealth is positive but diminishes as wealth increases. The individual is endowed with T units of leisure but if he or she devotes e units of effort to preventing accidents then $\ell = T - e$. Because we want to highlight effort supply we express utility as

$$U = B(w) + T - e.$$

The effort supply *e* is determined by the individual and is not a random variable. Of course uncertainty does affect the individual's wealth, which is either partially destroyed with probability $\pi(e)$ or remains intact with probability $1 - \pi(e)$. Note that the probability that the individual suffers a loss in wealth is a function of his or her effort supply *e*. We assume that $\pi(e'') < \pi(e')$ for all e'' > e'. In other words, an increase in effort reduces the probability of an accident.

Let *a* represent the value of an individual's wealth when there is an accident but no insurance has been purchased, and let *z* represent the value of the same individual's wealth when there is no accident and no insurance. Of course, a < z. The actual wealth will be different from both *a* and *z* if insurance is purchased. Let *x* denote the individual's actual wealth when he or she suffers an accident, taking into account any insurance benefits that may be paid. Let *y* denote wealth when there is no accident but the individual has paid an insurance premium. The individual's expected utility (EU) is

$$EU = \pi(e)B(x) + [1 - \pi(e)]B(y) + T - e.$$

The individual will choose e, x, and y to maximize EU. The values of x and y are subject to the individual's market opportunity equation, but the market places no restrictions on the choice of effort level e. Section 9.3 simplifies the calculations by assuming that e can be set equal to zero or one but nothing in between. (The individual either does or does not devote effort to preventive care.) In the final two sections (9.4 and 9.5) the individual chooses from a continuum of effort supply levels.

DEFINITION: The individual variables and parameters

Without insurance, an individual's wealth is a if he or she has an accident and z otherwise. With insurance, the individual's wealth is x if he or she has an accident and y otherwise. The individual's EU is

$$EU = \pi(e)B(x) + [1 - \pi(e)]B(y) + T - e,$$

where *e* is the amount of effort that the individual devotes to preventive care, and π (*e*) is the probability of an accident as a function of *e*. *T* is the maximum possible effort supply.

The next two sections show that the individual will set e = 0 if he or she is insured, and that the resulting outcome is inefficient. Devoting zero effort to preventive care is a consequence of the fact that we are abstracting from the

188 Hidden Action

possibility of personal injury—by carelessly operating a chain saw, say. In our model, an accident merely reduces individual wealth. There are lots of situations in which an injury *can* occur with positive probability and the individual devotes some effort to preventive care as a result. However, our analysis of the extreme case can be applied here to demonstrate that the level of preventive care chosen by individuals will not result in an efficient outcome. *Without* insurance, the individual *will* invest in preventive care, even if only wealth is at risk, but that will yield less expected utility than complete insurance and zero investment in preventive care.

Section 7.4 of Chapter 2 proves that individuals will demand and obtain complete insurance (x = y) under competitive conditions. This theorem is exploited in the next two sections (9.3 and 9.4), although the value of x will depend on the amount of preventive care supplied by individuals. In employing the complete insurance theorem in this way we greatly simplify the calculations. However, doing so requires us to ignore the fact that insurance companies will provide less than complete insurance to partially offset the diminished incentive to take care. This oversight is corrected in the last section, 9.5.

9.3 The binary choice model of moral hazard

This subsection assumes that you are familiar with the economics of insurance without moral hazard—that is, the material in Sections 7.1, 7.2, and 7.4 of Chapter 2. To obtain quick insight, we initially suppose that there are only two possible effort supply levels: Either e = 1, which means that the individual devotes effort to prevention, or e = 0, which means that no effort is made. Assume a competitive insurance market and hence that the individual purchasing insurance faces fair odds. Competition forces insurance companies to offer the contract on the fair odds line that maximizes EU, as we showed in Section 7.4 of Chapter 2. Therefore, x = y by the complete insurance theorem. The fair odds line is

$$\pi(e)x + [1 - \pi(e)]y = \pi(e)a + [1 - \pi(e)]z.$$

When we set x = y we get $x = y = \pi(e)a + [1 - \pi(e)]z$. Set

$$w(e) = \pi(e)a + [1 - \pi(e)]z.$$

Then w(e) is the individual's wealth, whether there is an accident or not, when insurance is purchased under fair odds. Note that $\pi(e)B[w(e)] + [1 - \pi(e)]B[w(e)] = B[w(e)]$. Therefore, the individual's utility after purchasing insurance is

$$\mu(e) = B[w(e)] + T - e.$$

We have implicitly assumed that individuals are identical. This is unrealistic, but it simplifies our calculations without generating any misleading conclusions. All that remains is to calculate individual effort supply and to determine if we have an efficient supply of effort at the market equilibrium. If the individual sets e = 0 we have

$$\mu(0) = B(w) + T.$$

We have used *w* instead of w(0) as the argument of *B* because the effort supply of the *other* policyholders will determine their probability of an accident, which in turn influences the number of dollars in claims that have to be paid and hence the level of wealth *w* that is available with insurance. Now, before working out the individual's utility when e = 1 we introduce one more assumption: There is a large number of policyholders, and hence a change in one individual's probability of an accident does not appreciably affect the premium charged because it does not appreciably affect the value of claims paid per capita. Hence, if our individual sets e = 1 his or her wealth with insurance will still be *w*. Therefore, if e = 1 we have

$$\mu(1) = B(w) + T - 1.$$

Clearly, B(w) + T is larger than B(w) + T - 1, so the individual will set e = 0. Then everyone will set e = 0, and hence $w = w(0) = \pi(0)a + [1 - \pi(0)]z$.

No individual has an incentive to devote effort to prevention, and thus each individual's utility will be $B(\pi(0)a + [1 - \pi(0)]z) + T$ at the market equilibrium.

Because the individual is risk averse, this utility level will be significantly higher than $\pi(1)B(a) + [1 - \pi(1)]B(z) + T - 1$, the utility without insurance. (Note that without insurance the individual typically *does* have incentive to set e = 1.)

Will the market equilibrium be efficient? *If* everyone were to set e = 1, then with complete insurance the individual utility level would be $B(\pi(1)a + [1 - \pi(1)]z) + T - 1$. For many real-world applications (and for the example to follow) we would have

$$B(\pi(1)a + [1 - \pi(1)]z) + T - 1 > B(\pi(0)a + [1 - \pi(0)]z) + T,$$

in which case the market equilibrium is not efficient.

Example 9.1: An inefficient effort supply at equilibrium

Let $B(w) = 2.4\sqrt{w}$ and T = 1. Therefore, $U = 2.4\sqrt{w} + 1 - e$. If e = 1 suppose that the probability of an accident is $\frac{1}{3}$, but if e = 0 the probability of an accident is $\frac{1}{2}$. That is, $\pi(1) = \frac{1}{3}$ and $\pi(0) = \frac{1}{2}$. Finally, a = 30 (individual wealth is 30 if there is an accident but no insurance), and z = 72 (individual wealth is 72 if there is no accident and no insurance). There are *n* identical individuals, where *n* is a large number. Now, $w(0) = \frac{1}{2} \times 30 + \frac{1}{2} \times 72 = 51$. Therefore at the market equilibrium (where everyone sets e = 0) individual utility is $2.4\sqrt{51} + 1 - 0 = 18.14$. *If* everyone were to set e = 1 then individual wealth would be $w(1) = \frac{1}{3} \times 30 + \frac{2}{3} \times 72 = 58$, and individual utility would be $2.4\sqrt{58} + 1 - 1 = 18.28$. Because each individual's utility is higher when each sets e = 1, the market equilibrium is inefficient.

We conclude by demonstrating that if there is no insurance the individual will set e = 1. He or she will maximize $\pi(e) \times B(30) + [1 - \pi(e)] \times B(72) + 1 - e$. Because the only choice is e = 1 or e = 0 we just have to compare $\frac{1}{3} \times 2.4\sqrt{30} + \frac{2}{3} \times 2.4\sqrt{72} + 0$ with $\frac{1}{2} \times 2.4\sqrt{30} + \frac{1}{2} \times 2.4\sqrt{72} + 1$. The former is 17.96, and the latter is 17.75. Therefore, when there is no insurance the individual chooses e = 1, the efficient amount of preventive care. *That does not mean that we get an efficient outcome without insurance*. Note that individual utility is 18.14 at the market equilibrium, which is higher than 17.96, the utility without insurance. Therefore, everyone is better off with insurance—because individual risk is diminished—even though the individual has no incentive to invest in preventive care when insured and when there is no insurance the individual supplies maximum effort.

$\partial 9.4$ A continuum of effort supply levels

Again, we assume that you are familiar with the material in Sections 7.1, 7.2, and 7.4 of Chapter 2. We employ the model of Section 9.2 but this time with effort continuously variable between zero and one. That is, *e* can be any fraction between zero and one inclusive. The individual's EU is

$$EU = \pi(e)B(x) + [1 - \pi(e)]B(y) + T - e.$$

The derivative of π with respect to *e* is negative because preventive care reduces the probability of an accident.

As in the previous subsection, we exploit the fact that the competitive insurance market results in the individual receiving the contract on the fair odds line that maximizes EU, given that the individuals are identical and everyone chooses the same level of *e*. Therefore, x = y by the complete insurance theorem, and hence the individual's wealth is

$$w(e) = \pi(e)a + [1 - \pi(e)]z$$

whether there is an accident or not. (This is explained at the beginning of the previous subsection.) Because $\pi(e)B[w(e)] + [1 - \pi(e)]B[w(e)] = B[w(e)]$, the individual's utility after purchasing insurance is

$$\mu(e) = B[w(e)] + T - e.$$

To determine the choice of effort supply by the individual at equilibrium, we set wealth equal to w, independent of one individual's effort supply, because a change in the probability of an accident by a single individual will not have an appreciable effect on the terms on which insurance can be offered to the market. Therefore, the individual will choose e to maximize B(w) + T - e. Obviously, this is achieved by e = 0. Everyone is in the same position, so each individual sets e = 0. Therefore, at the market equilibrium individual wealth is $w(0) = \pi (0)a + [1 - \pi (0)]z$, whether or not the individual has an accident. Then each individual's utility is B[w(0)] + T. This will be inefficient if there is an effort level e such that

$$B[w(e)] + T - e > B[w(0)] + T.$$

∂ Example 9.2: Calculating the efficient effort supply

 $B(w) = 4 \ln(w+3)$ and T = 1. Therefore, $U = 4 \ln(w+3) + 1 - e$. Note that we have B' > 0 and B'' < 0 for all w. Assume that a = 24, z = 96, and $\pi(e) = \frac{1}{2} - \frac{1}{4}e$, with $0 \le e \le 1$. At equilibrium, e = 0 and thus expected wealth is $\frac{1}{2} \times 24 + \frac{1}{2} \times 96 = 60$. Individual utility at equilibrium is $4 \ln 63 + 1 = 17.57$.

Now, let's calculate the efficient level of preventive care: The end of Section 7.2 of Chapter 2 shows why setting x = y maximizes EU on the individual fair odds line. We are assuming that the probability of an accident is the same for everyone, and that *a* and *z* are the same for everyone. Given *e*, the probability $\pi(e)$ of an accident is determined. Given the probability of an accident, individual EU is maximized by complete insurance. Therefore, the individual's wealth will be

$$w(e) = 24\pi(e) + 96[1 - \pi(e)] = 24(\frac{1}{2} - \frac{1}{4}e) + 96(\frac{1}{2} + \frac{1}{4}e) = 60 + 18e$$

whether there is an accident or not. Hence, individual utility is

$$G(e) = 4\ln(60 + 18e + 3) + 1 - e$$

Now, choose *e* to maximize *G*(*e*). We have $G'(e) = (4 \times 18)/(63 + 18e) - 1$. Then $G'(e) = 72(63 + 18e)^{-1} - 1$, and thus the second derivative is $G''(e) = -72(63 + 18e)^{-2} \times 18$, which is negative for all *e*. Therefore, if G'(e) = 0 gives us a value of *e* between 0 and 1 it will maximize G subject to $0 \le e \le 1$.

The statement 72/(63 + 18e) - 1 = 0 implies 72 = 63 + 18e, the solution of which is e = 1/2. Therefore, the equilibrium, with e = 0, is not efficient. To verify this we compute individual utility when e = 0 and utility when e = 1/2. (Recall that w(e) = 60 + 18e.)

$$G(0) = 4\ln[w(0) + 3] + 1 - 0 = 4\ln(63) + 1 = 17.57.$$

$$G(1/2) = 4\ln[w(1/2) + 3] + 1 - 1/2 = 4\ln(72) + 1/2 = 17.61.$$

Individual EU is 17.57 at equilibrium but would be 17.61 if everyone could be induced to set $e = \frac{1}{2}$. The market equilibrium is not efficient.

If insurance were not available at all then the individual would choose e to maximize

$$(1/2 - 1/4e)4\ln 27 + (1/2 + 1/4e)4\ln 99 + 1 - e.$$

The first derivative of this function is $-\ln 27 + \ln 99 - 1 = +0.299$. The individual will increase *e* until it reaches the upper bound of one. In other words, if insurance is not available then the individual will set *e* = 1, in which case the probability of an accident is $\frac{1}{4}$ and EU is $\frac{1}{4} \times 4 \ln 27 + \frac{3}{4} \times 4 \ln 99 + 1 - 1 = 17.08$. Note that individual EU is higher when insurance is purchased in a competitive market, even though the individual then has a strong incentive not to devote effort to preventive care.

$\partial 9.5$ Incomplete insurance

With fair odds the individual will devote no effort to prevention. We can expect insurance companies to modify the contract to give individuals some incentive to devote effort to prevention. Let *c* denote the net insurance coverage. That is, x = a + c. If *p* is the cost of insurance per dollar of net coverage, then *pc* is the policy premium and thus y = z - pc.

The consumer will choose *e* and *c* to maximize EU, which is

$$V(e, c) = \pi(e)B(a + c) + [1 - \pi(e)]B(z - pc) + T - e,$$

subject to the constraints $0 \le c \le z/p$ and $0 \le e \le T$. We will simply assume that the solution value of *c* lies strictly between 0 and z/p.

Let's rewrite EU as $V(e, c) = \pi(e)B(x) + [1 - \pi(e)]B(y) + T - e$, with x = a + c and y = z - pc. We will use the chain rule, and the fact that dx/dc = 1 and dy/dc = -p. Then

$$\frac{\partial V}{\partial c} = \pi(e)B'(x) + [1 - \pi(e)]B'(y) \times -p$$

and

$$\frac{\partial V}{\partial e} = \pi'(e)B(x) - \pi'(e)B(y) - 1 = -\pi'(e)[B(y) - B(x)] - 1.$$

When $p = \pi/(1 - \pi)$ the individual faces fair odds. Let's check: We have x = a + cand y = z - pc. The first equation implies c = x - a, and when we substitute x - a for c in the second equation we get y = z - p(x - a). This can be expressed as px + y = pa + z. And if $p = \pi/(1 - \pi)$ we can multiply both sides of the equation by $1 - \pi$ yielding $\pi x + (1 - \pi)y = \pi a + (1 - \pi)z$. In summary, if $p = \pi/(1 - \pi)$ then the individual's market opportunity line embodies fair odds.

We know that with fair odds a risk-averse individual will maximize EU at the point where x = y. In that case $\partial V/\partial e = -1$. In other words, the individual can always increase expected utility by reducing *e*. Therefore, the individual will set e = 0, as we discovered in the two previous subsections.

With moral hazard, fair odds will not likely be offered, even under competitive conditions. However, if *p* is close to fair—that is, close to the ratio of probabilities—then *x* will be close to *y* and $-\pi'(0)[B(y) - B(x)]$ will be positive but small. (Recall that π' is negative at all effort levels. We have y > x because without complete insurance the individual's wealth will be smaller when he or she suffers an accident, even with insurance.) Then $\partial V/\partial e$ will be negative when e = 0. If we assume diminishing returns to effort supply (or just nonincreasing returns), then $\partial V/\partial e$ will be negative for all *e*. In that case, the individual will still set e = 0 to maximize expected utility.

We have discovered that, even without fair odds, the individual will not devote any effort to prevention if the odds are close to fair. What is the efficient effort supply? Again, we maximize per capita EU. Efficiency implies fair odds. (Review the first two paragraphs of Section 7.5 of Chapter 2.) Fair odds and risk aversion imply complete insurance, and hence x = y. Then the individual will have wealth $w(e) = \pi (e)a + [1 - \pi (e)]z$, whether or not there is an accident. Note that $dw/de = -\pi'(e)(z - a)$. We maximize

$$G(e) = B[w(e)] + T - e.$$

The first derivative is $G'(e) = B'[w(e)] \times -\pi'(e)(z-a) - 1$. Because $B', -\pi'$, and (z-a) are all positive, we would expect $-B'[w(0)]\pi(0)[z-a]$ to be large and positive. That is, G'(0) is positive in all but rare cases, and the efficient effort supply is positive (where G'(e) = 0).

The difference between the maximization of *V* and the maximization of *G* is that in the former case we are modeling individual decision making, and we have to be careful not to give the individual control over the effort supply of others. Consequently, we don't have *x* and *y* change when the individual changes *e*. However, when we maximize *G* we are not modeling individual decisions but rather determining the highest level of per capita expected utility that the economy is capable of producing. Therefore we are free to change everyone's effort supply simultaneously.

Example 9.3: The efficient effort supply when odds are not fair

 $U = 4 \ln(w + 3) + 1 - e$ as in Example 9.2, with a = 24, z = 96, and $\pi(e) = \frac{1}{2} - \frac{1}{4}e$. $(0 \le e \le 1.)$ In Example 9.2 we calculated the efficient effort supply by maximizing $G(e) = 4 \ln(63 + 18e) + 1 - e$. We concluded that G(e) is maximized at $e = \frac{1}{2}$. Will the individual set $e = \frac{1}{2}$ when the odds are not fair? Note that $\pi(\frac{1}{2}) = \frac{3}{8}$ and $1 - \pi(\frac{1}{2}) = \frac{5}{8}$. Fair odds would require p = (3/8)/(5/8) = 0.6. Suppose, instead, that p = 0.8. What effort will the individual supply? Because $B(w) = 4 \ln(w + 3)$, $\pi'(e) = -\frac{1}{4}$ for all e, and $\pi(\frac{1}{2}) = \frac{3}{8}$ we have

$$\frac{\partial V(\frac{1}{2}, c)}{\partial c} = \frac{3}{8}B'(x) + \frac{5}{8}B'(y) \times -0.8 = 0.375B'(x) - 0.5B'(y)$$
$$= 0.375 \times \frac{4}{27 + c} - 0.5 \times \frac{4}{99 - 0.8c}.$$
$$\frac{\partial V(\frac{1}{2}, c)}{\partial e} = \frac{1}{4}[B(y) - B(x)] - 1 = \frac{1}{4}[4\ln(y+3) - 4\ln(x+3)] - 1$$
$$= \ln(99 - 0.8c) - \ln(27 + c) - 1.$$

If we set $\partial V/\partial c = 0$ we get [1.5/(27 + c)] - [2/(99 - 0.8c)] = 0, the solution of which is c = 29.5. When we substitute c = 29.5 into $\partial V/\partial e$ we get $\partial V/\partial e = -0.71$. That is, at $e = \frac{1}{2}$ we have $\partial V/\partial e < 0$, which means that the individual would not set $e = \frac{1}{2}$ but would reduce effort supply.

Do we have an equilibrium with e = 0 when the odds are not fair? With e = 0 and fair odds we would have p = 0.5/0.5 = 1. Suppose that p actually equals 1.2. Now we have

$$\frac{\partial V(0, c)}{\partial c} = 0.5 \times \frac{4}{27 + c} - 1.2 \times 0.5 \times \frac{4}{99 - 1.2c}$$
$$= \frac{2}{27 + c} - \frac{2.4}{99 - 1.2c}.$$
$$\frac{\partial V(0, c)}{\partial e} = \frac{1}{2} [4 \ln(y + 3) - 4 \ln(x + 3)] - 1$$
$$= 2 \ln(99 - 1.2c) - 2 \ln(27 + c) - 1.$$

Setting $\partial V/\partial c = 0$ yields c = 27.75, and thus x = 51.75 and y = 62.7. When we substitute c = 27.75 into $\partial V/\partial e$ we get $\partial V/\partial e = -0.64$. Therefore, the individual has no incentive to raise e above 0, even if the odds are not fair, although the efficient effort supply is e = 1/2.

Source

The claim that the frequency of accidents goes way down when the drivers of police cars and rescue vehicles are electronically monitored is based on Nalebuff and Ayres (2003, p. 107–8). Data on the net social benefit of the silent alarm Lojack is taken from Ayres and Levitt (1998). The notion that home equity insurance would be viable if the claim were based on average house prices originated with Shiller and Weiss (1994). The sketch of HMOs is based on Dutta (2000, pp. 304–5.) Reinhardt, Hussey, and Anderson (2004) track the rate of change of health care expenditure as a fraction of GDP.

Links

See Zeckhauser (1970) for an important early contribution to the study of deductibles and similar devices. Lazear (1992) contains additional examples of market-generated solutions to hidden action problems. Dranove (2000) provides a thorough (and nontechnical) economic analysis of the U.S. health care industry. See Diamond (1992) for a thorough discussion of the social significance of the difference between risks that individuals can modify with their behavior and those that they cannot. See Dranove (2000) for a review of the evidence on the quality of care under HMOs. See Chapter 5 (especially pages 131 and 132) of Kotlikoff and Burns (2004) to see why the U.S. government's attempt to control Medicare and Medicaid costs by enrolling participants in HMOs has not worked.

Problem set

The first two questions assume that T = 1 and each individual can set e = 0 or e = 1 but not any intermediate value. If e = 0 then the probability of an accident is $\frac{1}{2}$ but if e = 1 the probability of an accident is $\frac{1}{4}$.

- **1.** $B(w) = 2.5\sqrt{w}$, a = 30, and z = 72 for each individual. Find the competitive equilibrium and determine whether it is efficient.
- **2.** Each individual's utility-of-wealth function is $B(w) = 5 \ln(w + 1)$. If there is an accident then the individual's wealth will be 40 but if there is no accident wealth will be 120.
 - **A.** If insurance is not available determine the value of *e* chosen by the individual, the individual's wealth if there is an accident, wealth if there is no accident, expected wealth, and EU.
 - **B.** If insurance is available in a competitive market determine the value of *e* chosen by the individual, the individual's wealth if there is an accident, wealth if there is no accident, expected wealth, and EU.
 - **C.** What is the efficient level of effort? Determine the resulting wealth if there is an accident, wealth if there is no accident, expected wealth, and EU utility.

The remaining questions assume that T = 1 and e can assume any value between 0 and 1, inclusive.

3. If an individual devotes *e* units of effort to preventive care then the probability of an accident is 1 - e. Each individual has the expected utility function

$$\pi (0.2\sqrt{x}) + (1-\pi)(0.2\sqrt{y}) + 1 - e$$

where π is the probability of an accident, *x* represents wealth if there is an accident, and *y* represents wealth if there is no accident. If there is no insurance then *x* = 50 and *y* = 150.

- **A.** Assuming that the values of *x* and *y* are determined in a competitive insurance market, show how *x* depends on *e* (that is, display *x* as a function of *e*).
- **B.** Assuming that everyone can be made to employ the same level of preventive care, find the value of *e* that maximizes per capita EU.
- **C.** Explain briefly why the competitive equilibrium is not efficient. (A verbal argument will suffice, but a calculus-based explanation is also perfectly satisfactory.)
- **4.** Each individual's utility-of-wealth function is $B(w) = 2.5\sqrt{w}$. If there is an accident then the individual's wealth will be 36 but if there is no accident then wealth will be 72. The probability of an accident is $\frac{1}{2} \frac{e}{6}$. Answer questions A, B, and C of question 3 for this model.
- **5.** Let $B(w) = 10 \ln(w + 1)$, and $\pi(e) = \frac{1}{2} \frac{1}{4}e$. Show that the individual will set x = y and e = 0 at the competitive equilibrium.
- **6.** Let $B(w) = \beta \ln(w+1)$, $\pi(e) = \frac{1}{2} \frac{1}{4}e$, and $U(x, y, \ell) = \pi B(x) + (1 \pi) B(y) + \alpha \ell$. Find a condition on the positive parameters α , β , a, and z that implies inefficiency of the competitive equilibrium of the insurance market. The last four questions do not assume fair odds.
- **7.** Prove that the individual will set x < y if $p > \pi/(1 \pi)$.
- **8.** If $B(w) = 4 \ln(w+3)$, a = 24, z = 96, and $\pi(e) = \frac{1}{2} \frac{1}{4}e$, for what values of *p* will the *individual* set e > 0?
- **9.** Prove that if $p < \pi/(1 \pi)$ then the policy will not generate enough premium revenue to pay all of the claims submitted.
- **10.** For Example 9.3, show that the individual will set e = 1 if p is sufficiently high.

Corporate Governance

1.	A Brief Tour of Several Countries		197
2.	Partnerships		198
	2.1	The Model	199
	2.2	A two-person partnership	202
	2.3	Reputation and repeated interaction Problem set	204 206
3.	The	Owner-Employee Relationship	207 212
4.	The Owner-Manager Relationship in Practice		212
	4.1	How managers are disciplined	213
	4.2	Examples of managerial shirking	222
	4.3	The Enron debacle	227
	4.4	Why shareholders allow managerial shirking	228
5.	Agency Theory		231
	5.1	A diagrammatic introduction	232
	5.2	The basic agency model	237
	5.3	Risk-neutral managers	239
	5.4	Risk-averse managers and binary effort supply	244
	5.5	Risk-averse managers and a continuum of effort	
		levels	249
		Problem set	252

1. A Brief Tour of Several Countries

This chapter investigates incentives in firms. We explore the hidden action problems of a modern corporation. Section 1 compares firms in several leading industrialized countries. Section 2 examines the relationship between two senior executives who share the firm's profits and is followed by a brief look at the relationship between the owner and employees in an owner-managed firm (Section 3). The rest of the chapter is devoted to the hidden action problem confronting a widely dispersed group of shareholders whose objective is to have the firm that they jointly own maximize the value of their shares. Can they rely on the board of directors to provide the appropriate incentives to the company's management team, even though it is extremely costly for the shareholders to monitor the management and the board members themselves?

1 A BRIEF TOUR OF SEVERAL COUNTRIES

We are primarily concerned with the attempt of a firm's owners to obtain a satisfactory return on the capital they supply to the firm. The owners provide financing through the purchase of shares in the firm and also when the firm uses retained earnings for replacement of, or addition to, the capital equipment. Firms also borrow financial capital, and in many industrialized countries bank loans are a much more important source of finance than in the United States. All of the firm's suppliers of finance wish to ensure that management runs the firm in a way that brings them a high return. We refer to this as the *agency problem*.

DEFINITION: The modern corporation's agency problem

The firm's owners and creditors seek a high return on their money but the daily decisions that determine that rate of return are made by the firm's management team, and the managers may be assumed to have their own welfare at heart.

One striking difference between the pattern of ownership across countries lies in the role of the financial sector. U.S. banks were prohibited from holding equity in corporations until the repeal of the Glass-Steagall Act in 1999. In the United States only 5% of shares are held by banks and other financial institutions, but in France, Germany, Japan, and the United Kingdom the fraction is closer to 30%. In fact, a Japanese corporation has a long-term relationship with a particular bank, called its *main bank*. The main bank is expected to play a significant role in monitoring the firm with which it is associated. In practice, the bank lets the firm have its way, except in times of crisis. The United States and United Kingdom are quite similar in that a firm's owners are expected to do the monitoring through their representatives, the board of directors. Apart from those in Canada, Britain, and the United States most corporations are private their shares are not traded on a public exchange—and even the ownership of some public firms is highly concentrated, often with family ownership and control.

The fraction of shares held by *individuals* is much higher in the United States than in other countries. And the level of top executive pay is much higher in the United States, as is the fraction of executive pay that is received in the form of bonuses. (It may not be a coincidence that 43% of the total investment in research and development—R&D—by the leading industrial countries comes from the United States. See Baumol, 2002, for data on R&D.)

German firms have a supervisory board of directors, half of whom are elected by shareholders and half of whom are elected by employees. Daily operations are directed by a management board, appointed by the supervisory board. Firms in Canada, the United States, and Great Britain have a single board of directors with outside members, elected by shareholders, and inside members, the firm's top executives. The CEO (chief executive officer) is the head of the management team and is often the chairman of the board. Because the outside directors are typically nominated by the incumbent management they usually remain loyal to those chief executives. French law allows a firm to choose between the Anglo-Saxon corporate form and the German form. Japanese law makes it relatively easy for shareholders to nominate and elect directors, but the board is large and unwieldy, and in practice the owners have less influence over management than in the United States. In part that is because Japanese management is expected to give higher priority to stable employment for the firm's workers than dividends for the owners.

Allen and Gale (2000) conclude that the agency problem is substantial regardless of the form of corporate governance or corporate finance. "Managers seem to get their way most of the time" (Vives, 2000, p. 2). However, Carlin and Mayer (2000) find considerable evidence that corporate governance and finance can be a significant factor for a *country*'s economic performance.

In the next section we begin our in-depth study of the agency problem by examining a simple two-person production team. By cooperating with each other, the pair can take advantage of a production of technology that is richer than the one available to an individual working alone. But as soon as two or more individuals are involved, incentives come into play. Is one worker motivated to consider the effect that his or her actions have on the welfare of the other members of the team?

Source

Allen and Gale (2000) and Vives (2000).

2 PARTNERSHIPS

In this section we discover why few firms are organized as partnerships. The key workers in a partnership are also the firm's residual claimants, who share the profits. Sometimes one partner puts more capital into the firm than the others and receives a larger share of the profits, but the essential point is that each partner shares in the income created by the effort of the other partners. By the

2. Partnerships

same token, the partner receives only a fraction of the income generated by his or her own effort and thus each partner contributes less than the efficient amount of effort. The owner-employee relationship (Section 3) results in each person

In the 1960s the Cuban government used state stores to provide citizens with equal rations of food and clothing. Housing was almost free, and every family was provided with a free vacation at the beach. Absenteeism on the job soared and productivity and product quality declined precipitously. The government reluctantly implemented a complicated and comprehensive system of monitoring (Kohler, 1977).

receiving a leisure-income package that he or she prefers to the one obtained in a partnership. Why, then, are partnerships are observed at all? We provide the answer at the end of the next section.

Why does production take place in *teams* in the first place? Because there is more output per worker in a team than when individuals work independently. That is, one-person firms generate far less output per unit of labor input than do multiperson firms. We examine these issues by means of a simple framework.

2.1 The Model

There are two consumer goods, leisure and income. Income is a composite commodity—the total number of dollars available for expenditure on goods other than leisure. Let x be the amount of leisure consumed and let y be the income level. U(x, y) is the individual's utility function. Let T denote the length of a period, in hours. (T = 168 if the time period is a week.) The income generated by a production team depends on the amount of effort expended by each of the team members. If e is the amount of effort contributed by an individual, then x = T - e. Note that x is not just T less the number of hours "worked." An individual may show up for work but not put in much effort, consuming leisure on the job. This will affect the amount of output and income generated by the firm and also the individual's utility through its effect on leisure. Therefore, we need to keep track of effort, not hours on the job. (The quality of effort is just as important as the quantity, but we simplify the analysis by focusing on only one dimension of the principal-agent problem, the incentive to work rather than shirk.)

When individuals work on their own (in one-person firms) the equation $y = \alpha e$ represents the production technology of a single firm. It expresses the income available when e units of effort are expended. The positive constant α is the income generated per unit of effort. Because x = T - e we have $y = \alpha(T - x)$ for a one-person firm. The individual chooses the bundle (x, y) to maximize U subject to the production constraint $y = \alpha(T - x)$, or $\alpha x + y = \alpha T$. The chosen point $C^{\alpha} = (x^{\alpha}, y^{\alpha})$, illustrated in Figure 4.1, is a point of tangency of the indifference curve and the production line L_{α} . In economic terms the marginal rate of substitution at C^{α} equals α , which is the opportunity cost of leisure—that is, the amount of income sacrificed per hour of leisure consumed. (Of course, individuals can observe their own effort levels.)

Now consider a two-person firm. Let β denote the income generated per unit of effort when two individuals cooperate in production. Let e_i denote the effort





expended by *i*, x_i the leisure consumed by *i*, and y_i the income of individual *i* (*i* = 1, 2). Then $y_1 + y_2 = \beta(e_1 + e_2)$. Of course $\beta > \alpha$.

We abstract from a lot of real-world phenomena to focus on the role of incentives. For one thing, we assume that the individuals in our firm have identical preferences, and U(x, y) will again represent the individual's preference scheme. And we study only two-person firms, although the generalization to *n* persons is straightforward. If the individuals have identical consumption then we will have $x_1 = x_2$ and $y_1 = y_2$, and hence $e_1 = e_2$. Then $y_1 + y_2 = \beta(e_1 + e_2)$ implies $2y_i = \beta(2e_i)$ and thus $y_i = \beta e_i$. This allows us to contrast the two-person firm with the one-person firm (Figure 4.1). Because $\beta > \alpha$ the *per capita* production line $y_i = \beta e_i = \beta(T - x_i)$, denoted L_β , lies above its one-person counterpart L_α . Therefore, the two-person firm can provide a higher level of per capita income than the one-person firm.

DEFINITION: Ingredients of the partnership model

Let *x*, *e*, and *y* denote, respectively, individual leisure consumption, effort, and income. The time endowment is *T* and thus e = T - x. Individual utility *U* is a function of *x* and *y*. The per capita production function is $y = \beta e$.

2. Partnerships

Let $C^{\beta} = (x^{\beta}, y^{\beta})$ denote the utility-maximizing bundle available with a twoperson firm assuming that the team members consume the same bundle. That is, C^{β} maximizes U(x, y) subject to $y = \beta(T - x)$. The marginal rate of substitution at C^{β} equals β , which is the opportunity cost of leisure per person in a twoperson team. Note that we can also say that C^{β} maximizes U(x, y) subject to $y \le \beta(T - x)$: If $y < \beta(T - x)$ we can increase utility by increasing *both* x and ywithout violating $y \le \beta(T - x)$.

We can call the outcome that gives each person the bundle C^{β} fair precisely because both people have identical bundles and identical preferences. Because the outcome is also efficient it is a reasonable standard by which to measure the performance of a particular contractual arrangement.

For any number of identical workers, the outcome that gives C^{β} to each person is fair and efficient.

Proof

As we have said, the outcome is fair by definition. To prove efficiency, let (x_1, y_1) and (x_2, y_2) be two bundles that give one person more utility than C^{β} and the other person at least as much. Say,

$$U(x_1, y_1) > U(x^{\beta}, y^{\beta})$$
 and $U(x_2, y_2) \ge U(x^{\beta}, y^{\beta})$.

Now, C^{β} maximizes U subject to $y_i \leq \beta(T - x_i)$ so anything that gives higher utility than C^{β} must violate the inequality $y_i \leq \beta(T - x_i)$. Therefore $y_1 > \beta(T - x_1)$. If we actually had $y_2 < \beta(T - x_2)$ then we could increase both y_2 and x_2 to satisfy $y_2 = \beta(T - x_2)$, and that would result in an increase in utility for person 2. This new utility level would be higher than $U(x^{\beta}, y^{\beta})$ because we already have $U(x_2, y_2) \geq U(x^{\beta}, y^{\beta})$. In that case we have contradicted the fact that (x^{β}, y^{β}) maximizes U subject to $y = \beta(T - x)$. Therefore we have

$$y_1 > \beta(T - x_1)$$
 and $y_2 \ge \beta(T - x_2)$,

and hence $y_1 + y_2 > \beta(T - x_1 + T - x_2) = \beta(e_1 + e_2)$. Then the new outcome that assigns (x_i, y_i) to each *i* is not feasible: The total income allocated exceeds the total available income $\beta(e_1 + e_2)$ generated by the effort that would be supplied. Similarly, $U(x_1, y_1) \ge U(x^{\beta}, y^{\beta})$ and $U(x_2, y_2) > U(x^{\beta}, y^{\beta})$ cannot hold for any feasible pair (x_1, y_1) and (x_2, y_2) . Therefore, the outcome assigning C^{β} to each is efficient—there is no feasible outcome that gives both persons at least as much utility as C^{β} and one person strictly more. (Note that this argument easily extends to a team of more than two individuals.)

Can C^{β} in fact be realized? C^{β} is feasible, but only when the two persons cooperate. But then one person's income depends on the total income of the team, which in turn depends on the amount of effort contributed by *both* persons. Will there be incentive for each to contribute the required amount of effort? Consider the partnership case.

202 Corporate Governance

2.2 A two-person partnership

The rules of partnership are simple. The partners share equally in the income that is created by their joint effort, and any losses are absorbed equally by the individual partners. Professional service industries employ the partnership method of team organization more than any other contractual form. Partnership is the typical form for accounting firms, law firms, and medical clinics. Apart from professional services, however, large firms are rarely organized as partnerships. (Investment banking has largely converted from partnerships to the standard corporate owner-employee form.) Why are partnerships widely employed in the professional service industries but rarely in evidence elsewhere? To answer that question we need to focus on the amount of effort contributed by a utilitymaximizing partner.

We assume n = 2 until further notice. If effort is unobservable, how can a team member determine the effort supplied by the other partner at equilibrium? By working out the utility-maximizing responses to the incentives governing the partner's behavior. Also, the individual can determine the total effort supplied by others simply by observing total output *y*, inferring the total effort supplied, and then subtracting the person's own effort.

We put the spotlight on partner 1. Person 1 chooses (x_1, y_1) to maximize $U(x_1, y_1)$ subject to the sharing rule $y_1 = \frac{1}{2}\beta(e_1 + e_2)$ that determines a partner's income. Partner 1 cannot control e_2 , so we take it as fixed, at *c*. That is, when partner 1 changes her effort supply we assume that partner 2's effort level does not change. But y_1 changes as a result of the change in e_1 . We will be at equilibrium if each partner's effort level maximizes his or her own utility given the other partner's effort level. We have $e_1 = T - x_1$ so partner 1 will endeavor to maximize $U(x_1, y_1)$ subject to $y_1 = \frac{1}{2}\beta(T - x_1 + c)$. The constraint can be expressed as $\frac{1}{2}\beta x_1 + y_1 = \frac{1}{2}\beta(T + c)$, which is $L_{0.5\beta}$ in Figure 4.2. This is a budget line. The individual opportunity cost of leisure is $1/2\beta$ under the partnership sharing rule. It is the ratio of the "prices" of the two goods. If person *i* consumes one more hour of leisure the *firm* loses one hour of effort and thus β dollars of income, but individual *i* loses only $\frac{1}{2}\beta$ dollars of income because the β dollars of income generated would have been shared with the other person. Therefore, utility maximization requires equality between the marginal rate of substitution (MRS) and the opportunity cost $\frac{1}{2\beta}$. Compare this with our derivation of C^{β} : In that case the constraint line was $y_1 = \beta e_1$, or $y_1 = \beta (T - x_1)$, or $\beta x_1 + y_1 = \beta T$. Because at C^{β} the social (or team) opportunity cost of leisure is β , and the marginal rate of substitution equals β .

Because the partners are identical, we can drop the subscript. The variables will pertain to a single individual. Let $C^P = (x^P, y^P)$ be the individual's choice at equilibrium under the partnership arrangement. The partners will have the same consumption at equilibrium because they have the same preferences and are confronted with the same incentives. We know that $C^P \neq C^\beta$ because the MRS is $1/2\beta$ at C^P and double that at C^β . We also know that the utility of C^β exceeds the utility of C^P because C^β maximizes utility subject to $y \leq \beta e$, and C^P is one of the bundles that satisfies $y \leq \beta e$. To prove the latter claim note that $y^P + y^P \leq \beta (e^P + e^P)$ because the partnership outcome is feasible. (Recall that the partners make the same choices at equilibrium, and


Figure 4.2

whenever we fail to employ a subscript it is implicit that the variable applies to each individual.) Therefore, $y^P \leq \beta e^P$. So the partnership outcome gives each person less utility than C^{β} , which is feasible. Not only that, C^P provides more leisure but less income than C^{β} . The increased consumption of leisure at C^P is due to the fact that the opportunity cost of leisure to the individual in a partnership is half of the opportunity cost to the firm: Half of the β dollars lost to the firm when *i* consumes another unit of leisure would have been given to the other partner.

To prove that leisure consumption is higher under C^P than under C^β , try placing C^P above C^β on L_β . Now draw the indifference curve through C^P . It has an MRS of $1/2\beta$ at C^P so the curve is flatter than L_β at C^P . The indifference curve through C^P is flatter than L_β at C^P , and it gets flatter as we move to the right. The MRS at C^β is equal to β , and thus the indifference curve through C^β is tangent to L_β at C^β , and it gets steeper as we move to the left. Consequently, the two indifference curves would have to intersect, which is impossible.

Here is a formal proof that leisure consumption is higher at C^P than at C^{β} : Suppose to the contrary that $x^{\beta} > x^P$. Then C^P is above C^{β} on L_{β} . Therefore $L_{0.5\beta}$, the line representing the partnership income formula, which is the individual's partner's budget constraint, cuts L_{β} above C^{β} . Then there will be a point *S* on $L_{0.5\beta}$ that is northeast of C^{β} . (Mentally shift $L_{0.5\beta}$ up in Figure 4.2, until it cuts L_{β} above C^{β} .) But then the value of *x* at *S* exceeds x^{β} , and the value of *y* at *S* exceeds y^{β} . It follows that $U(S) > U(C^{\beta})$. We have $U(C^P) \ge U(S)$ because C^P maximizes *U* on the line $L_{0.5\beta}$. Therefore, $U(C^P) > U(C^{\beta})$, contradicting the fact that C^{β} maximizes *U* on L_{β} . Therefore, we have to drop the supposition that $x^{\beta} \ge x^P$. (Why can we rule out $x^{\beta} = x^P$?)

Individual utility is higher at the fair and efficient outcome C^{β} than at the partnership equilibrium C^{P} , and each partner supplies less effort than at the fair and efficient outcome.

We used the theory of consumer choice to derive C^{P} . We did so by determining the individual's demand for each commodity. But commodity 2 is income. Why doesn't the individual have an infinite demand for income? Because at the margin the individual has to pay for each dollar of income with an increased supply of effort and hence a sacrifice of leisure.

Example 2.1: A specific case with two partners

Let U(x, y) = xy, $\beta = 2$, and T = 24. (We'll ignore the subscript at first.) To find C^{β} we maximize U subject to $y = 2 \times e = 2(24 - x)$. Replace y in the utility function with 48 - 2x and then maximize $x[48 - 2x] = 48x - 2x^2$. This quadratic is maximized at $x^{\beta} = 48/4 = 12$. Then $y^{\beta} = 48 - 24 = 24$. We have $C^{\beta} = (12, 24)$.

To find C^P , the partnership equilibrium, we solve

maximize $x_1 y_1$ subject to $y_1 = \frac{1}{2} \times 2(24 - x_1 + e_2)$.

Substitute $24 - x_1 + e_2$ for y_1 in the utility function, and then maximize $x_1(24 - x_1 + e_2)$, treating e_2 as a constant. Then we maximize $x_1(24 - x_1 + e_2) = (24 + e_2)x_1 - x_1^2$. This quadratic is maximized when $x_1 = (24 + e_2)/2$. Setting $e_2 = e_1 = 24 - x_1$ and substituting $24 - x_1$ for e_2 yields $x_1 = 12 + 12 - \frac{1}{2}x_1$ and hence $x_1 = 16$. Therefore, $y_1 = y_2 = 16$. At the partnership equilibrium, $x^P = 16$, $y^P = 16$, and $e^P = 8$ for each person.

Let's compare utility levels: At the fair and efficient allocation C^{β} each person's utility equals $x^{\beta} \times y^{\beta} = 12 \times 24 = 288$. At the partnership equilibrium C^{P} utility equals $x^{P} \times y^{P} = 16 \times 16 = 256$ for each, which is about 11% less than the utility at C^{β} .

The partnership outcome is not efficient; there is another *feasible* outcome that would give each more utility. Section 3 shows that a different contractual arrangement can give each team member the incentive to supply the amount of effort required by the fair and efficient outcome C^{β} . First, we consider whether incentives are different in a long-term partnership relationship.

2.3 Reputation and repeated interaction

Subsection 7.3 of Chapter 1 showed that an efficient level of cooperation can be sustained in a relationship that is repeated indefinitely. We now apply this reasoning to partnerships. (This section is self-contained, but it might be wise to read the concluding section of Chapter 1—particularly 7.3.) The infinite horizon assumption is a good way to model a long-term business relationship in which the finite lifetime of a business has no bearing on individual decision making in the early stages or even the intermediate term.

The two partners interact repeatedly for an infinite number of periods, 1, 2, ..., *t*, We represent the individual's preferences by discounting the sum of the single period utilities. If δ is the discount factor and u_t is the partner's period *t* utility, then the individual maximizes

$$u_1 + \delta u_2 + \delta^2 u_3 + \cdots + \delta^{t-1} u_t + \cdots$$

where $0 < \delta < 1$. Because $\delta < 1$, the discount factor will be close to zero if *t* is very large and hence periods that are very remote will get almost no weight in the decision. Recall that the infinite sum $a + a\delta + a\delta^2 + \cdots + a\delta^t + \cdots$ equals $a/(1 - \delta)$ when $0 < \delta < 1$ (Section 7.2 of Chapter 1).

As in the static partnership model we simplify the analysis by assuming that the partners have identical utility functions. In the static (one-shot) case we found that there is an efficient individual effort supply e^{β} that is larger than the effort supply e^{P} at the partnership equilibrium. The individual partner maximizes utility by choosing $e^{P} < e^{\beta}$ because the individual opportunity cost of leisure consumption is only half the opportunity cost to the two-person team. The effort supply e^{P} by each partner results in each consuming $C^{P} = (x^{P}, y^{P})$ where $x^P = 24 - e^P$ and $y^P = \frac{1}{2}\beta(e^P + e^P)$. The efficient effort supply e^β supports the commodity bundle C^{β} . Although $U(C^{\beta}) > U(C^{P})$, the consumption plan C^{β} does not emerge at equilibrium because e^{β} is not a best response by one partner to the supply of e^{β} by the other. However, when the partnership relationship is repeated period after period, one partner has an opportunity to punish the other for deviating from the efficient effort supply. The punishment takes place in future periods of course and, unless the discount factor is very low, the one-period gain to a deviating partner will not be large enough to offset the infinite number of rounds of future punishment.

Specifically, if the discount factor is sufficiently large then we have a Nash equilibrium when each partner supplies e^{β} in the first period, and e^{β} in any period *t* provided that the other supplied e^{β} in the previous t - 1 periods, *and* a partner threatens to supply e^{P} every period following any period *t* in which the other partner failed to supply e^{β} .

Example 2.2: Infinitely repeated version of Example 2.1

Let U(x, y) = xy, $\beta = 2$, and T = 24. We derived $C^{\beta} = (12, 24)$ with $e^{\beta} = 12$ in the one-shot case (Example 2.1). At the partnership equilibrium, $e^{P} = 8$ with $C^{P} = (16, 16)$. Note that $U(C^{\beta}) = 288$ and $U(C^{P}) = 256$. One Nash equilibrium for the infinitely repeated partnership has each partner supply 12 hours of effort in the first period and every subsequent period as long as the other partner supplied 12 hours in each previous period, but will supply 8 hours of effort in period *t* and every subsequent period if the other partner did not supply 12 hours of effort in period t - 1.

Suppose that partner 1 deviates from e = 12 in period *t*. What's the highest one-period utility that a partner can achieve when the other partner supplies 12 hours of effort? The answer is obtained by maximizing U(x, y) when $y = \frac{1}{2} \times 2 \times (24 - x + 12)$. That is, we maximize $x \times (24 - x + 12) = 36x - x^2$. Using the formula for maximizing a quadratic (Section 1 of Chapter 2) or calculus yields x = 36/2 = 18. Then partner 1 will supply 6 hours of effort. Each individual's income will be $\frac{1}{2} \times 2(6 + 12) = 18$ and partner 1's period *t* utility will be U(18, 18) = 324. (One can also solve for *x* by setting the *MRS*, which is $\frac{y}{x}$, equal to partner 1's opportunity cost of leisure, which is $\frac{1}{2} \times 2$, and then using the budget constraint y = 24 - x + 12.)

By deviating in period t, partner 1 gets an increase in utility of at most 324 - 288 = 36. But he is then punished by partner 2 in period t + 1 and every subsequent period. Partner 2 supplies 8 hours of effort in period t + 1 and beyond. We already know what partner 1's best one-shot response is because we have a unique Nash equilibrium of the one-shot game when each supplies 8 hours of effort. Therefore, each will supply 8 hours of effort from period t + 1 on. Hence the deviating partner will get a utility of at most 256 in each of those periods. Had he not deviated, utility would have been 288 each period. Therefore, by deviating partner 1 gets a utility bonus of at most 36 in period t but suffers a utility penalty of at least 288 - 256 = 32 in every period after the tth. Discounting to period t, we find that deviating will not be profitable if

$$36-32\delta-32\delta^2-32\delta^3-\cdots\leq 0.$$

and this simplifies to $36 \le 32\delta/(1-\delta)$. Therefore, deviating cannot benefit either player if $36 - 36\delta \le 32\delta$, or $\delta \ge 36/68 = 0.53$. If the discount factor is 0.53 or higher then the trigger strategy specified in the first paragraph is a Nash equilibrium.

The efficient outcome can be sustained if the partnership lasts many periods and the partners are not too impatient. However, it is just *one* of the Nash equilibria in the infinitely repeated partnership. There are many other equilibria. At the other extreme, if both announce their intentions to supply in each period of the repeated game the amount of effort that emerges in the one-shot Nash equilibrium whatever the other does, then we have a Nash equilibrium of the repeated game. (This is true whatever the discount factor.)

Source

Alchian and Demsetz (1972). The discussion of repeated interaction in a partnership is based on Radner (1991).

Links

See Milgrom and Roberts (1992, pp. 522–3) on the conversion of investment banks from partnerships. See Kandel and Lazear (1992) on the role of peer pressure in partnerships. See Aoki (2000) for a perspective on the computer industry in the Silicon Valley. Levin and Tadelis (2002) provide an in-depth examination of a partnership extending over time. Williams and Radner (1995) show how the introduction of uncertainty *improves* the prospects for risk-neutral partners achieving an efficient outcome.

Problem set

1. Find the fair and efficient outcome and the partnership equilibrium for Example 2.1 by using the theory of consumer choice. This means that you have to set the MRS equal to the price ratio for the appropriate budget line. For the utility function U = xy, the MRS at generic bundle (x, y) is y/x.

3. The Owner-Employee Relationship

- **2.** Find the partnership equilibrium in a two-person firm with the following features: T = 24, and each individual has the utility function $U = 5 \ln(x + 1) + y$. Each dollar of income generated by the firm requires a total of two hours of effort per day as input.
- **3.** Determine the equilibrium of a two-partner firm in which each partner has the utility function $U(x, y) = 8\sqrt{x} + y$, T = 24, and the output/input ratio is two. Show that the outcome is inefficient.
- **4.** Determine the equilibrium of a two-partner firm in which each partner has the utility function $U(x, y) = 16\sqrt{x} + y$, T = 24, and the output/input ratio is two. Show that the outcome is inefficient.
- **5.** A firm has *four* partners and each has the utility function $U(x, y) = \sqrt{x} \times y$, with T = 24. An individual's MRS at the bundle (x, y) is y/2x. The firm's profit, before deducting the partners' pay, is \$50 multiplied by the total effort supplied. Prove that the partnership equilibrium is $C^P = (16, 400)$. That is, prove that at the equilibrium, each partner has x = 16 and y = 400.
- **6.** Consider a model of team production in which total income is four times the total amount of effort supplied. There are two individuals on the team and each individual *i* has the utility function $U(x, y) = x^2 y$ and T = 24.
 - **A.** Determine the commodity bundle that maximizes person 1's utility subject to the production technology constraint and the requirement that the partners wind up with identical utility levels.
 - **B.** Determine the partnership equilibrium. Make sure you identify the amount of each good consumed by each person.
- **7.** Using the parameters of Example 2.2 show that there is a Nash equilibrium of the infinitely repeated partnership in which each partner supplies 10 hours of effort each period.
- **8.** Translate the condition on δ into a condition on the interest rate, guaranteeing that the Nash equilibrium of Example 2.2 is in fact a Nash equilibrium of the infinitely repeated partnership. Do the same for the equilibrium of question 7.

3 THE OWNER-EMPLOYEE RELATIONSHIP

We continue with the simple case of a two-person production team without repetition. The previous section showed that the partnership sharing rule resulted in an inefficient outcome. Under that formula, the individual's opportunity cost of leisure consumption is half of the opportunity cost to the team. Therefore, each partner overconsumes leisure—in the sense that the resulting outcome is inefficient. We are about to see that an efficient outcome can be reached if one of the team members is singled out as the owner, and then pays the other person—the worker—a high income, but only if the worker supplies a high level of effort. The owner then keeps every dollar of profit after paying the worker. That makes the owner the *residual claimant*, with an individual opportunity cost of

leisure consumption of β , which is the opportunity cost to the team. Because the individual and the social (i.e., team) opportunity costs of leisure consumption are identical, the owner is motivated to supply the amount of effort that leads to the fair and efficient outcome. The worker is also induced to deliver that effort supply because he or she doesn't get paid otherwise.

DEFINITION: Residual claimant

A member of a production team (typically a profit-seeking firm) is the residual claimant if that person has title to all of the revenue left over after all contractual obligations are met. In everyday language, the residual claimant gets the enterprise's profit.

Suppose that person 1 is the sole owner of the two-person firm of Section 2, and she hires individual 2 as the second member of the team. The two individuals work together as in a partnership but the reward scheme is quite different: Person 1 pays person 2 an income of y_2 with the residual income going to the owner person 1. That is $y_1 = \beta(e_1 + e_2) - y_2$. To determine e_1 and e_2 we need to be more explicit about the worker's contract. The owner agrees to pay the worker exactly y^{β} , no more and no less, provided that the worker, person 2, supplies *at least* e^{β} units of effort. Recall that $C^{\beta} = (x^{\beta}, y^{\beta})$ is the bundle that maximizes individual utility on the per capita production line $y = \beta e = \beta(T - x)$. If person 2's effort supply is less than e^{β} he is not paid at all.

Definition: The worker's contract

The worker's pay y_2 equals y^{β} if $e_2 \ge e^{\beta}$ and $y_2 = 0$ if $e_2 < e^{\beta}$.

Note that this contract requires monitoring by the owner to ensure that the threshold effort level e^{β} is reached. We assume initially that monitoring is costless. When we turn to the case of significant monitoring costs, at the end of this section, we will see that the higher these costs are, the more likely it is that the firm will be organized as a partnership.

Assuming that the worker's utility at (x^{β}, y^{β}) is at least as high as he can obtain by working elsewhere—or by staying home and consuming leisure—it is in the worker's interest to supply exactly e^{β} units of effort and receive the income y^{β} . If $e_2 < e^{\beta}$ the worker is dismissed. Even if he gets a new job, there will be costs associated with the transition, and consequently the worker's utility will fall below C^{β} .

Now consider the owner's situation. The owner wishes to maximize $U(x_1, y_1)$ subject to the two constraints $y_1 = \beta(e_1 + e^{\beta}) - y^{\beta}$ and $x_1 = T - e_1$. Recall that $y^{\beta} = \beta e^{\beta}$. Then person 1 will maximize $U(x_1, y_1)$ subject to $y_1 = \beta e_1 + \beta e^{\beta} - y^{\beta} = \beta e_1 = \beta(T - x_1)$. But $y_1 = \beta(T - x_1)$ is the equation of the line L_{β} in Figure 4.1, and we already know that C^{β} maximizes utility on that line. Therefore, it is in the

owner's interest to choose C^{β} , which means that she supplies e^{β} units of effort. We have designed a contract regime such that each person has an incentive to supply e^{β} units of effort and as a result each receives the bundle C^{β} , and this outcome is efficient. (Efficiency was proved in Subsection 2.1.)

Instead of a contract that pays the worker y^{β} if his effort is e^{β} or more and zero otherwise, the owner could simply offer a wage of β and let the worker choose his utility-maximizing basket (x_2 , y_2) subject to the budget constraint $\beta x + y = \beta T$. The line representing this budget constraint is L_{β} of Figure 4.2. Therefore, the offer of a wage of β leads the worker to choose C^{β} as in the case of the "contribute e^{β} or else" contract. Either contract is effective, although they have slightly different monitoring consequences. (What are the differences?)

The fair and efficient outcome will be implemented by a contract that makes one member of the team the residual claimant, whereas the other gets paid the fair and efficient income level, but only if he or she supplies the fair and efficient level of effort.

Why do the owner and employee receive the same level of utility, $U(x^{\beta}, y^{\beta})$, at equilibrium? The owner does not get a bonus for risk taking, because there

The reason why the benefits of technological progress are not all captured by suppliers of capital is that the owners of firms compete with each other for the skilled labor needed to implement the technological innovations. This drives up wages, passing on part of the fruits of progress to workers. is no risk in our simple model. Imagine a community in which many of the individuals have access to the technology that converts input into output at the rate of β dollars of income per unit of effort. If *all* firms pay their workers less than y^{β} then owners must get more than y^{β} . Workers can leave the firm and start their own businesses in which they receive $U(x^{\beta}, y^{\beta})$ as owner. The Silicon Valley of California is celebrated for this. However, if the workers do

better than the owners, the latter can be expected to sell their businesses and seek jobs as workers. This will increase the supply of workers and lower their pay.

Very few firms are organized as partnerships. That's because each partner receives only a fraction—(1/n)th if there are n partners—of the income generated by his or her own effort. Thus there is an incentive to undersupply effort, and hence the equilibrium leisure-income bundle consumed by each partner provides a lower level of utility than the ideal plan C^{β} , which can be realized by the owner-worker regime. This utility differential is greater the larger is the number of partners, because the individual's opportunity cost of leisure consumption falls as n increases. Then why do partnerships exist at all? The answer has to do with the costs of monitoring the worker to ensure that the threshold level of effort e^{β} has been supplied. In some enterprises the manager (or the manager's agent) need do little more than ascertain that the worker is on the job and at the appropriate workstation to determine that there is no shirking (i.e., that $e_i \ge e^{\beta}$). If bicycle wheels are being produced and a sample reveals wheels with missing spokes it is a relatively easy matter to determine



Figure 4.3

the source of the problem. The costs of monitoring are low in these cases. (A worker's output can be tracked electronically in some production processes.) Even when the per capita monitoring cost *m* is subtracted from C^{β} there is still substantially higher utility, $U(x^{\beta}, y^{\beta} - m)$, than is provided by the partnership outcome C^{P} .

However, monitoring costs are very high in firms that provide sophisticated consulting or diagnostic services. Consider a team of accountants, lawyers, or physicians. If one member of the team is to verify that another has done a good job for a client then the former would essentially have to retrace the latter's steps. In that case the per capita technological income-leisure trade-off line L_m for the owner-worker regime will be parallel to L_β but strictly below it, the vertical distance between the two providing a measure of the per capita monitoring cost m. The utility-maximizing bundle on L_m is C^m , assuming that the two individuals receive the same bundle (Figure 4.3). The partnership scheme does not require monitoring so L_β remains the appropriate trade-off line in that case. Figure 4.3 shows that $U(C^P)$ will exceed $U(C^m)$ if m is sufficiently large. However, given the per capita monitoring cost m, the per capita partnership utility level $U(C^P)$ falls as n increases. Therefore, large firms are much less likely to be organized as partnerships than small firms. (Do you tend to study in a team of two or three people or in a larger group?)

Example 3.1: A large number of workers

Let U(x, y) = xy, $\beta = 2$, and T = 24, as in Example 2.1, but this time we assume that there are *n* workers. With *n* team members the per capita production line is still y = 2e = 2(24 - x), and thus we still have $C^{\beta} = (12, 24)$. If person 1 is the residual claimant and pays each of the other n - 1 individuals \$24 provided that they supply 12 units of effort, then each worker will accept such a contract. (Suppose that the only alternative is staying home and consuming the bundle

3. The Owner-Employee Relationship

(24, 0), which delivers zero units of utility.) The owner then chooses *x* and *y* to maximize U = xy subject to

$$y = 2[e + 12(n - 1)] - (n - 1)24.$$

This equation reflects the fact that the total effort supply is e + 12 (n - 1), where e is the owner's effort, and the owner has to pay \$24 to each of the n - 1 workers, hence the subtraction of 24(n - 1) from the team revenue of 2[e + 12(n - 1)]. The equation y = 2[e + 12(n - 1)] - (n - 1) 24 clearly reduces to y = 2e, and we know that $C^{\beta} = (12, 24)$ maximizes U = xy on the line y = 2(24 - x). Therefore, the owner also supplies 12 units of effort. Each team member receives the fair and efficient bundle (12, 24).

Next we work out the partnership equilibrium: To find C^P we solve

maximize
$$x_1 y_1$$
 subject to $y_1 = \frac{1}{n} \times 2(24 - x_1 + c)$

where *c* is the sum of everyone's effort supply except the individual in question. That's not something that person 1 can control, so we treat *c* as a constant. Replace y_1 in the utility function with $(1/n) \times 2(24 - x_1 + c)$ and maximize

$$V(x_1) = x_1 \times \frac{1}{n} \times 2(24 - x_1 + c) = \frac{2}{n} \times (24 + c)x_1 - \frac{2}{n} \times x_1^2.$$

This quadratic is maximized at

$$\frac{x = (48 + 2c)/n}{4/n} = 12 + \frac{c}{2}.$$

(We can drop the subscript now.)

At equilibrium everyone supplies the same effort and hence consumes the same amount of leisure. Therefore, c = (n - 1)(24 - x). We now have

$$x = 12 + \frac{(n-1)(24-x)}{2},$$

the solution of which is x = 24n/(n+1). (Confirm that x = 16 when n = 2.) Because T = 24, effort supply is zero when $x_1 = 24$. Therefore, for large n an individual partner's leisure consumption is close to 24 and effort supply is close to zero. Specifically, e = 24/(n+1). Total effort supply is $n \times 24/(n+1)$. Therefore, the total profit to be shared by the partners is $2 \times [n/(n+1)] \times 24$. Individual income is thus 48/(n+1). What is individual utility at C^P where

$$x^{P} = \frac{24n}{n+1}$$
 and $y^{P} = \frac{48}{n+1}$?

 $U(C^P) = [24n/(n+1)] \times [48/(n+1)] = 12 \times 24 \times 4n/[(n+1)(n+1)]$. Recall that 12×24 is individual utility at the fair and efficient bundle (Example 2.1). Because n < n+1 we have

$$\frac{4n}{(n+1)(n+1)} < \frac{4}{n+1}$$

Therefore, $U(C^P)$ is less than $U(C^\beta) \times 4/(n+1)$. When *n* is large, $U(C^P)$ is a small fraction of $U(C^\beta)$. The differential between $U(C^P)$ and $U(C^\beta)$ increases with *n*. In fact, $U(C^P)$ is close to zero for large *n*.

Professional service firms do not require enormously large teams, in contrast to most manufacturing processes. Therefore, $U(C^P) > U(C^m)$ is plausible in the professional service field where the partnership format is employed. In fact, there will be some minimal monitoring in a partnership. Why would the partners in a medical clinic invest the time necessary to monitor each other even a little? The cost of malpractice can be very costly to the client, and hence to the partnership, and because the partners share losses as well as profits, each member is very vulnerable to the shirking of the others. This gives the partners a strong incentive to monitor each other, in contrast to the position of senior executives in a limited liability corporation. The personal assets of an owner cannot be tapped to pay the creditors or legal penalties of a limited liability corporation.

Source

Alchian and Demsetz (1972) provides the key insight for this section.

Problem set

- 1. Compute the equilibrium outcome for a firm that has ten workers, one of whom is the owner who manages the firm. The firm's net income (net of the cost of materials, etc.) is always five times the total amount of effort contributed. (The total effort includes the effort contributed by the owner.) Each individual has the utility function $U_i(x_i, y_i) = x_i^2 y_i$. Where x_i is the number of hours of leisure consumed by *i* per week, and y_i is *i*'s income per week. Assume that monitoring is costless.
- **2.** Suppose that the owner of the firm described in question 1 offered each worker a contract that paid exactly \$230 per week as long as 40 units of effort or more were contributed, and paid \$0 if less than 40 units of effort were observed. Show that this is *not* an equilibrium contract.
- **3.** Here is a more subtle version of question 2. Suppose that the owner of the firm described in question 1 offered each worker a contract that paid exactly \$200 per week as long as 40 units of effort or more were contributed, and paid \$0 if less than 40 units of effort were observed. Show that this is not consistent with *long-run* equilibrium.
- **4.** Let the monitoring cost per person *m* be a function *m*(*n*) of the number *n* of team members, with *m* increasing as *n* increases. How does this affect the analysis?

4 THE OWNER-MANAGER RELATIONSHIP IN PRACTICE

In the next section we derive the managerial contract that maximizes the owners' return. This section surveys the U.S. corporate landscape. A modern U.S. corporation has many shareholders. To take advantage of the economies of scale in production and advertising, the contemporary firm must be extremely large, beyond the capacity of all but a handful of individuals to finance on their own. Even if most firms *could* have single owners, risk aversion would motivate providers of capital to diversify their portfolios—in other words, to own a small

fraction of many different companies rather than large fractions of a few companies. Therefore, our starting point is the fact that a firm is owned by a large number of unrelated shareholders.

A firm's owners appoint a manager to run the company on their behalf. The owners do not make the daily decisions that determine how successful the firm will be. Those are made by the managerial team, which answers to a board of directors, who are supposed to represent the owners. The directors and the managers are agents of the owners, and we speak of an *agency problem* because the managers and the directors will, at least to some extent, act so as to enhance their own welfare, not that of the owners. The shareholders want the managers to make decisions that lead to the maximum value of shares on the stock exchange, and that requires high annual profits. However, the owners will not know when the highest possible profit is attained. If it were obvious how to maximize profit then the shareholders could issue the appropriate orders directly to the workers. As it is, the shareholders need to hire an agent.

The shareholders want the board to design a contract that provides every incentive for the manager to maximize shareholder value, even though the manager's immediate concern is his or her own material well-being, which will depend on factors that are not perfectly correlated with the firm's profit, such as the manager's future prospects. Even if a portion of the managers' wealth is held in the form of shares in the firms that they manage, the other shareholders cannot be sure that the managers will do all they can to maximize profits. The firm's long-run profitability will not be the only aspect of the managers' stewardship that affects their welfare. For example, the managers' prospects for future employment and income may be enhanced if they increase the size of their present enterprise, and this may induce them to increase the firm's output beyond the point at which profit is maximized. The managers may expose the firm to more risk than is in the best interest of the owners. They may even negotiate a merger that provides them with tens of millions of dollars in consulting fees but does nothing (or less than nothing) for the owners of the firm that spearheads the merger.

4.1 How managers are disciplined

The severity of the agency problem is mitigated in a variety of ways that impose discipline on the managers. The devices by which managers are regulated can be grouped into four categories:

Regulation by *shareholders*—through contracts that provide performance incentives and via direct oversight by the board of directors.

Regulation by the *capital market*: If the firm performs poorly, a financier can buy a controlling interest in the company and replace the incumbent management. This is called a *hostile takeover*. The motivation for the takeover springs from the fact that if the company subsequently does well, the market value of the financier's shares will increase. The implicit *threat* of a hostile takeover gives managers substantial incentive for doing their jobs well.

- Regulation by the *legal system*: In most countries the manager of a firm is legally responsible to the firm's owners. The degree of investor and creditor protection afforded by the legal system varies appreciably across countries, although it is usually substantial, at least in theory.
- Regulation by *product markets*: If competition is intense, profit margins will be small and a firm can be driven out of business if cost (and hence price) increases or product quality declines because of management shirking. The probability of bad decisions costing the top executives their jobs if the firm goes under will factor into managerial decision making.

We discuss each of these sources of discipline in turn, but most of our attention is given to regulation by shareholders and by capital markets.

The legal system

The U.S. legal system may place managers under more intense scrutiny than in any other country, but its role is still rather limited. First, American courts

Before the reunification of Germany, some East German automobile manufacturing plants were so poorly run that the value of the cars that they produced fell short of the value of the inputs used. A fertilizer plant in India operated for more than twelve years without producing a single ounce of fertilizer, yet twelve hundred employees came to work every day. The plant was funded by the government and managed by public officials (Wheelan, 2002, p. 27). do not intervene in a company's internal business decisions. Shareholders can sue the directors if the directors do a bad job of oversight. American courts are also willing to adjudicate allegations of self-dealing and challenges to the CEO's compensation package. However, practical obstacles stand in the way of significant judicial review of compensation packages. Nevertheless, relative to other countries, both the U.K. and the U.S. legal systems give investors and creditors substantial protection. One fairly simple technique by which owners of firms in other countries can purchase American or British legal protection is by listing their

stock on the exchanges of those two countries. Some firms in emerging markets and in the European Union have done this (Shleifer, 2000).

Alternatively, a firm operating in a country whose legal system affords investors little protection from mismanagement can adopt the legal environment of a country that is more protective of investors by being acquired by another firm that already operates in such an environment. This becoming quite common in Western Europe (Shleifer, 2000). Suppose for instance that firm *X* is controlled by a small group of owners who also manage the firm. Suppose also that the managers of *X* are diverting a significant fraction of the profits from the rest of *X*'s shareholders into their own bank accounts. If shareholder wealth is diverted at a cost—so that a dollar lost by the other shareholders results in a gain of less than a dollar to management—then a takeover firm can buy a controlling interest in *X* at a price that will leave the incumbent owner-management team better off and still leave a net gain for the company that takes over firm *X*.

Example 4.1: Transplanting the legal system

Iliad Corp. is managed by Homer, the founder of the company, who also owns 60% of the shares. The annual profit is \$1000 of which Homer is entitled to \$600. Not all of the \$400 to which the outside shareholders are entitled reaches them. They get only \$100, because Homer diverts \$300 of their share to himself. However, the diverted funds are used to buy luxury boxes at a baseball stadium. In term's of Homer's utility, these box seats are equivalent to only \$120 in cash. If Virgil purchased the firm for the equivalent of an annual payment of \$900, with \$760 of that going to Homer and \$140 going to the incumbent outside owners then everyone gains. Virgil pays \$900 for something worth \$1000. The outside owners get \$140 instead of \$100, and Homer gets \$760 instead of \$720.

Why can't the shareholders pay the incumbent manager \$150 to leave the \$300 in the firm? Even assuming that the owners could do the math (and that is problematic, with the manager controlling the flow of information), we would then have a situation in which the manager could repeatedly threaten to act adversely to the owners' interest, inviting them to bribe him not to do so.

Product markets

How important is competition in product markets in disciplining top executives? The manager will lose his or her job if the firm suffers losses year after year, and this gives the manager incentive to avoid losses. This is a long way from claiming that the manager will make every effort to maximize profit, but it will prevent extreme abuse. Moreover, the more intense is the firm's competitive environment the more efficient the management team has to be for the firm to stay afloat. Each of the world's major industrialized countries is host to firms that are among the world's leaders—even countries that do not appear to provide managers with strong motivation. This suggests that product market competition has substantial impact. But note that even if competition in the product market does prevent profit from declining, it doesn't stop the top executives from adopting strategies that transfer profit from the owners to management.

Performance bonuses

We now turn to the possibility of paying managers in ways that motivate them to look out for the owners' interests, even though the managers cannot be closely monitored. In the United States, not only are CEOs paid twice as much on average as their counterparts in other countries, but they also receive a much higher fraction of that pay in the form of performance bonuses—50% in the United States. Other countries are slowly approaching U.S. practice, however (Murphy, 1999).

It is important to understand that a performance bonus doesn't necessarily provide an incentive to perform. When the CEO or the board announces, "We had a great year so everyone gets a big bonus," it is past performance—or perhaps

good luck—that is being rewarded, and that is not necessarily an inducement to do well in the future. A contract that permanently ties senior executives' pay to profit is a much better incentive device.

The Walt Disney Corporation was run by family members for several years after the death of the founder, and the family members did a poor job. Profits were dismal, and the managers even used \$31 million of the owners' wealth to repurchase the shares of a financier attempting to buy a controlling interest in the company in hopes of being able to turn it around. (They paid \$31 million more than the shares were worth on the stock market.) When Michael J. Eisner was hired to run the company in 1984 he was given a bonus of 2% of all profits in excess of a 9% return on equity. Under Eisner's leadership, the return on equity soared to 25%. (It was well below 9% when he was hired.) Over a five-year period, Eisner received about \$10 million a year in performance bonuses, a tiny fraction of what he delivered to the company's owners (Milgrom and Roberts, 1992).

What kind of performance bonus would induce a manager to maximize profit? A substantial bonus that is paid only if the manager realizes maximum profit would provide the appropriate incentive. However, if the shareholders know how much profit the firm is capable of generating they can simply write a contract so the manager's continued employment is conditional on the firm reaching its potential. How can the owners induce profit maximization when they don't know what the maximum profit is?

Stock options are a partial answer. A stock option is a commitment by the shareholders to the manager, allowing the latter to purchase a specified number of shares in the company in a specified time interval and at a fixed price, usually the price of the stock at the time the option is granted. This gives the manager a strong incentive to take actions that lead to the largest increase in the value of the stock over that time interval, and this is usually accomplished by generating the maximum profit for the firm. Increases in the firm's profit will be

noticed by investors and will result in an increase in the demand for the firm's shares. That in turn cause the share price to increase.

Example 4.2. Stock options and managerial incentive

Suppose the share price of firm *X* is currently \$100 and firm *X*'s CEO is given the option to buy shares at that price. Then if the share price were to increase to \$250 in one year, the CEO can buy shares for \$100 each and sell them immediately for \$250 each.

Stock options can be abused, however, especially because managers are not usually required to hold the stock for any period of time. The chief executives can cause a temporary increase in the share price by temporarily overstating profit. They can exercise their stock options and sell their stock before the actual profit is discovered. When the truth is revealed, investor confidence can be undermined to an extend that the resulting collapse in the share price can bring many shareholders to ruin. Also, the board of directors sometimes weakens the incentive effect of stock options by *repricing* them after a fall in the share price.

(If the price was \$100 a share two years ago when the option was offered to the CEO, and the price is currently \$90, the option price may be lowered to \$90!) Repricing *can* benefit owners if it is done after a drop in the share price that is unrelated to the manager's performance. In that case, repricing would restore the incentive effect of stock options. (If the price decline was sufficiently large, it may be extremely unlikely that the stock price would rise above the exercise price of the option, no matter how hard the manager strives to increase profit.)

Too often bonuses and stock options are given to reward service in the past. But it is future performance that the shareholders want to inspire; it is too late to affect past performance. In fact, poor performance may be the result of unfavorable random events-changes in exchange rates and so forth-that are beyond the control of the manager. The manager may have been exceptionally industrious and creative. The aim is to reward effort, and effort is imperfectly correlated with performance and profit. Paradoxically, it may be smart to give a stock option to a manager after a period of low profits, giving the manager a strong incentive to work more effectively in the future. That won't work with cash bonuses of course. In fact, cash bonuses are usually tied to the office-with a chairman receiving more than a vice president. So, performance bonuses are a potentially useful tool in the shareholders' attempt to induce the manager to maximize profit, but they are not often used appropriately. Moreover, shareholders often view these sort of financial devices as bribes to get the managers to do something that they are paid salaries to do in the first place. Accordingly, shareholders sometimes oppose the use of stock options on ethical grounds and will sometimes sue the manager if the firm's board of directors agrees to this type of compensation. Nevertheless, performance bonuses of one kind or another are ubiquitous.

When the price of the company's shares is used to connect the manager's pay to managerial performance—whether or not a stock option is used—the manager can profit from an economy-wide increase in share price levels. This happened during the bull market of the 1990s. The median pay of the CEOs of the top 500 firms (the Standard and Poor's 500) increased by about 150% from 1992 to 1998 (Perry and Zenner, 2000). Tying the manager's performance to the differential between the firm's share price and a general stock market price index might be a more effective incentive device.

Why aren't incentive schemes that condition the manager's pay on the firm's performance more widespread? On average, an increase of \$1000 in the market value of a company's shares increases the CEO's compensation by only about \$3.25, most of which is attributable to stock ownership (Jensen and Murphy, 1990a, 1990b). Using more recent data (from the period 1980 to 1994) and taking the stock *option* component of pay into account, Hall and Liebman (1998) discover that pay is substantially more sensitive to performance. (See also Perry and Zenner, 2000.) Moreover, Haubrich (1994) demonstrates that when the manager's risk aversion is taken into account, contracts come much closer to the predictions of theory—the theory of Section 5, that is.

It may not be necessary to have close to maximum incentive to induce close to maximum CEO performance. It might be appropriate to divide the set of managers into good guys and bad guys. Even the good guys will disappoint the

owners if they are not rewarded for walking the extra mile—but they certainly won't ruin their principals by driving the firm into bankruptcy, even if that would add significantly to the agents' wealth. The good guys will deliver maximum performance if they are generously rewarded for doing so, even if they get much less than a dollar at the margin for every dollar of extra profit realized by the firm. The bad guys, however, are governed only by material incentives, and if they can increase their wealth from extremely high to obscene by exploiting a loophole in their contract, they will do so—even if that impoverishes many of the firm's owners.

Shleifer and Vishny (1988) suggest that the members of the *board of directors* be paid in the form of stock in the company rather than salary. The practice of tying a director's pay to the company's stock is making significant inroads. In 1997, 81% of the Standard and Poor's 500 firms awarded either stock or stock options (or both) to their board members (Bebchuk, Fried, and Walker, 2001). In theory this would align the interests of the board and the shareholders. Although the board of directors represents the shareholders, and shareholders sit on the board, it is often dominated by the manager. (In fact, managers often control the selection of the board members. And they can arrange to have their firms award lavish consulting contracts to board members. Brickley, Coles, and Terry, 1994, marshal evidence suggesting that shareholders do better when the board contains a large number of directors who have no significant business ties with the company.) However, "it has been widely agreed that the board of directors is an ineffective way of dealing with this [agency] problem" (Allen and Gale, 2000, p. 76).

An increasingly commonplace device for inducing performance that generates the maximum increase in the value of the company's shares is the franchise arrangement. A company provides a standardized product: hamburgers, fried chicken, automobile parts, retail drug outlets, and so forth. When a new outlet is opened, should the existing shareholders operate it themselves or license someone else (the franchisee) to do it according to the standard formula? If the shareholders operate it themselves they must hire a manager, in which case profit maximization cannot be taken for granted. The franchise arrangement requires the licensees (or franchisees) to put up some of their own capital and manage the outlets themselves. The franchisee must also pay a substantial fee to the parent company in return for permission to use its brand label and enjoy the benefits of its reputation and national advertising. The franchisee becomes the residual claimant, keeping every dollar of profit after paying the license fee. This gives the local manager maximum incentive to run the firm efficiently. The parent is in the business of selling franchises instead of hamburgers (etc.). The franchise fees bring in more profit than would a chain of parent-managed outlets because managers would not perform as well under the latter system-because they would have less at stake personally.

Why would an entrepreneur pay a fee to the parent *and* agree to follow the company formula strictly in return for the right to produce something that the entrepreneur could legally sell without paying the fee? Because the identification with the parent's brand is a source of extra profit, justifying the hefty licensing

fee. But the extra profit depends upon consumer confidence in the brand, which in turn depends upon *other* outlets strictly adhering to the company formula, so each franchisee agrees to the limitations because the brand identification is not worth anything unless it is clear that the parent will enforce its standards generally. Nor will there be extra profit if the franchisee is not given the exclusive right to operate in a specified area, so the parent agrees not to allow anyone else to use the brand name in the franchisee's neighborhood. Without accepting this limitation the parent would have nothing valuable to sell. Each side agrees to restrict its activities because the configuration leads to maximum profit for all the participants.

There are many situations, however, in which franchising would not solve the problem of divorce between ownership (by shareholders) and control (by management): The product may be produced at only a few locations or sold only at the wholesale level. Most significantly, the firm may be too large to be purchased by a single individual, which is essentially what franchising requires.

Capital markets

How do capital markets discipline top executives? We do not discuss the role of debt financing, although it has the potential for disciplining the management team: The idea is that if the firm were committed to pay out substantial interest on debt, the executives would be forced to avoid shirking and also to minimize the amount of revenue diverted to their own bank accounts (Grosssman and Hart, 1982; Jensen, 1986). However, American corporations rely on retained earnings to finance expansion far more than debt.

When an outside interest purchases a controlling interest in a firm we refer to this as a *takeover*. When the incumbent management is replaced we refer to it as a *hostile takeover*. If the firm's performance had been poor, then the price of its shares will be low. If the new owners replace the management team with a more effective one, and there is a big increase in the flow of profits as a result, the share price will increase. The new owners will have realized a handsome return, justifying the takeover. The possibility of dismissal may provide an incentive for managers to maximize profit in the first place. (This process can also correct deviations from profit maximization caused by management error, as opposed to management shirking.) Hostile takeovers were relatively rare until the 1960s (Hansmann, 1996). They remain rare outside of Anglo-Saxon countries.

If an outsider can determine when the management team is underperforming, why can't the firm's current owners? The research required to evaluate a firm's performance is costly. If each owner has a small fraction of the shares, the cost of research to an individual will be greater than any increase in the value of the individual's holdings as result of that research. However, if an outside interest purchases a significant fraction of the shares, it will realize a net gain if the firm's performance does in fact improve. (See Examples 4.3 and 4.4 in Subsection 4.4.)

Takeovers are not inevitable when the management team does a bad job. Easterbrook (1984) estimates that it takes an anticipated 20% increase in the value of shares to trigger a takeover. In that case, the threat of a takeover does little to discourage management from diverting profit away from the owners: A

15% increase in the CEO's pay would result in a tiny drop in the value of shares (Bebchuk, Fried, and Walker, 2001, p. 26). Also, managers often restrict the flow of information concerning the internal operation of the firm, making it even harder to determine its potential. Moreover, it often happens that dismissed managers have contracts with the original owners that provide them with multimillion-dollar parting gifts (a *golden parachute*) in case they are fired as a result of a takeover. The fact that boards of directors offer this sort of compensation may point to the unwillingness of directors to properly monitor managers. However, golden parachutes can be socially beneficial if they induce managers to accept hostile takeovers.

Some acquisitions serve the managers' interests by entrenching their positions. Shleifer and Vishny (1988) report that managers sometimes initiate takeovers. If some managers have strong reputations in the railroad industry, say, and their firm acquires a railroad, then they will be much more valuable to the shareholders. They have strengthened their positions at the head of the firm, even if the acquisition diminishes the present value of shareholder wealth.

In spite of the obstacles, takeovers are far from rare in the United States (and the United Kingdom). Almost 10% of the U.S. firms listed in the Fortune 500 in 1980 have since been taken over in a hostile transaction—or one that started out hostile (Prowse, 1995). These takeovers left a trail of data that should allow us to determine if takeovers have provided a significant corrective. For takeovers during the period 1976 to 1990, the increase in the value of shares in the target companies was about \$750 billion according to Jensen (1993). Scherer (1988) is skeptical about the social value of takeovers, but Lichtenberg (1992) finds strong evidence that a firm's total factor productivity increased after a takeover. And in a review of the empirical work on this question, Jarrell, Brickley, and Netter (1988) conclude that takeovers induce a beneficial restructuring of real capital. According to Jensen (1986), restructuring of the firm following a merger sometimes eliminates projects with negative net present value. The contemporary consensus is that the takeovers of the 1980s precipitated significant efficiency gains (Holmström and Kaplan, 2001).

There is a free rider problem that could undermine takeovers as a device to discipline managers. Existing shareholders stand to benefit from any improvement in profitability that a takeover would bring. This could make them reluctant to sell to the takeover group at the current market price or at a price low enough to render the takeover profitable to the new owners. Consequently, it could become difficult or impossible to find enough current shareholders willing to sell their shares (Grossman and Hart, 1980). That is why the constitutions of many firms include a *dilution* provision. This allows the new owner to sell part of the firm's assets to another company belonging to the new owner at terms that are beneficial to takeover group and disadvantageous to the firm's minority shareholders. Dilution can also take the form of the new owners issuing themselves new shares. Why would the original owners of the firm place such a provision in their constitution when it is potentially to their disadvantage? Because it makes takeovers more credible and thus serves to discipline the firm's manager. If the discipline is strict enough then the incumbent manager will work assiduously to maximize

profit, vitiating the need for dilution. A two-tiered offer can also eliminated the free rider problem, as we showed in Subsection 6.7 of Chapter 1.

Are there any other techniques that can be used to provide managers with appropriate incentives? One surprising possibility is *insider trading* (Manne, 1965, 1966). This term refers to the managers of firm *A* using important information about *A*'s prospects that is not available to the general public or even to trading specialists. If this information is used to purchase or sell *A*'s shares in a way that benefits the managers or their friends we have an instance of insider trading. It seems very unfair for those on the inside to profit from their privileged position. Indeed, the U.S. Securities and Exchanges Commission declared insider trading unlawful in 1961, and the courts have ratified this position. (Insider trading is not unlawful if it is based on information that is available to the general public.) Is it harmful enough to outsiders to warrant its prohibition? Banerjee and Eckard (2001) examine data from mergers that took place during "the first great merger wave" (1897 to 1903), before insider trading was outlawed. They discovered that outsiders appear not to have benefitted significantly from the ban on insider trading.

One form of insider trading is clearly harmful to society in general. If managers were able to take short positions in the shares of their own company they would have a strong incentive to ensure that their firms did *badly*. Selling short consists in selling something you don't own (shares in this case) at a price agreed upon now for delivery at a specified time in the future. The person selling short is betting that the asset will fall in value. When it is time to deliver the promised number of units of the asset and the price *has* fallen, the seller simply buys the required number of units on the "spot" market and delivers them, collecting

In July 1929 the head of the large Chase bank sold short more than 42,000 shares of Chase stock in advance of the October crash (Malkiel, 2003, pp. 47–8). the high price specified in the original contract. If managers could do this with shares in the companies then they run they could get rich by mismanaging their companies so that the stock falls in value. The flow of goods and service to consumers would be correspond-

ingly diminished. It is clearly in our interest to have short sales by managers declared illegal. Short sales by insiders have been prohibited in the United States since 1936.

What about ordinary (spot) trading by insiders? Some claim that this gives managers a strong incentive to do their utmost for the shareholders. If the company performs substantially better than expected then the price of its shares will rise on the stock exchange. Once this superior performance is public knowledge the shares will be immediately bid up in price and it will be too late for a manager to benefit from purchasing his or her company's shares. But if the manager purchases his or her company's stock at the current price in the light of advance, inside information on the company's unexpected performance then substantial capital gains are made when the share price is bid up in the wake of public realization of the enhanced profitability. This suggests that insider trading—with short sales disallowed—can help align the interests of manager and shareholder. (You might ask yourself why stockholders do not insist that managers agree to abstain

from insider trading as part of the contractual agreement between shareholder and manager.)

4.2 Examples of managerial shirking

Managers' immediate concern is their long-run well-being. Unless incentives or personal integrity take them in a different direction, their performance will be designed to enhance their present income, nonmonetary rewards, and future monetary rewards on the job (involving the use of a company airplane, etc.), perceived value to other companies (to enhance job prospects elsewhere), and retirement package. Studies of the agency problem have uncovered a long list of avoidable deviations from profit maximization. Some are deliberate, and some are the result of poor judgment. Ideally, both can be corrected by means of contracts that provide appropriate incentives to the decision makers.

We have grouped a variety examples of departures from profit maximization into four categories: Deliberate mismanagement, wealth diversion from the owners to the manager, bad judgement, and sins of omission. It is usually easy to rationalize the first two in terms of the *manager's* welfare.

Deliberate mismanagement

Managers have been known to restrict the flow of information to the board of directors to make it harder to determine when the managers are acting in the interest of the shareholders. Managers may even reduce the present value of the annual profit stream by tapping a source of profit slowly, so that it provides a steady flow of comfortable returns over the long haul. This can yield an annual profit that is high enough to survive owner scrutiny but not so high as to raise expectations for a repeat of the previous year's record return. (This form of shirking was said to be a common practice of managers in the former Soviet Union.) In the 1970s the management of H. J. Heinz delayed declaring some of its profit in one year so that profit in subsequent years would be artificially higher, to allow bonuses to kick in. (If a bonus is paid in any year in which profit increases by 5%, and profit is increasing at the rate of 4% a year, then by declaring a 2% increase in one year and a 6% increase the next, the executives qualify for a bonus in the second year.) It is not uncommon for managers to delay an announcement that would have a positive effect on the share pricefor instance, the discovery of a new drug—until after they have been granted stock options (Yermack, 1997).

We have seen that the threat of a hostile takeover imposes considerable discipline on managers. That discipline is undermined when managers adopt strategies to make takeovers costly. In some cases they can even block them. There is evidence that their defensive strategies often work (Jarrell, Brickley, and Netter, 1988). Because managers control the flow of information, they may be able to persuade shareholders that the company attempting a takeover is not offering enough for the shares and that they should continue with the present management or wait for a better offer. Managers can use shareholder wealth—that is, company cash—to buy back the shares acquired by a firm attempting a takeover. This usually requires a payment in excess of the market value of the

shares. In the 1980s the Disney management (prior to the arrival of Michael J. Eisner) paid \$31 million in excess of market value to buy back shares. This is called *greenmail*. Managers have also used their time and shareholder wealth to lobby state governments for antitakeover legislation. They have been enormously successful. More than 50% of U.S. states have recently passed legislation making hostile takeovers more costly.

The constitutions of many firms include provisions that are activated when the firm is taken over without the endorsement of the board of directors. The purpose of these *poison pill* clauses is to preempt a hostile takeover by substantially reducing the value of the company to an outsider. Poison pills appeared for the first time in 1982. By making hostile takeovers excessively costly, poison pills entrench management at the expense of shareholders, as Malatesta and Walking (1988) and Ryngaert (1988) have demonstrated. (Both papers are good introductions to the poison pill technique; in particular they have insightful examples.) One poison pill strategy requires the new owner to make large payments to the incumbent management of the company. By far the most common strategy is dilution-a clause in the firm's constitution that permits the board of directors of the target firm to sell new shares to incumbent owners, at 50% of current market price, when an offer is made for the company. With more outstanding shares, the firm attempting the takeover finds that control of the target company would be worth less because it gets a smaller fraction of profits. Some have argued that this benefits the shareholders of the target firm because it gives its management bargaining power: Management can threaten dilution unless the shareholders are given a better deal by the takeover firm. If that were the case we would expect to see a company's shares rise in value after the adoption of a dilution clause, but the share price usually falls on the stock market. Another significant poison pill strategy gives the board the power to reject any offer that it considers not in the company's interest. If the board is in thrall to incumbent management it may use that provision to block a takeover that would benefit the shareholders but cause the incumbent management to be dismissed. (These examples are taken from Dutta, 2000, p. 172.) Comment and Schwert (1995) report that 87% of firms listed on the New York Stock Exchange have a poison pill statute of some kind on the books. A recent court decision in the state of Delaware (where many U.S. corporations are based) did away with *dead hand* pills that remained in effect even after an entire board was dismissed (The Economist, June 1, 2002, p. 61).

Self-dealing

Even when the managers are doing everything to maximize profit, there is much that they can do to increase the share of that profit going to the top executives by decreasing the share going to owners. In 1985 Victor Posner of Miami held a controlling interest in DWG, but he was not the sole shareholder. He extracted \$8 million in salary from DWG that year, even though the firm did not make a profit (Shleifer and Vishny, 1997, p. 742). However, most of the U.S. examples of wealth diversion from shareholders to CEOs are less direct, thanks in part to the intervention of the courts.

Management may buy an expensive fleet of corporate jets and use them primarily to fly executives to a trendy resort, or buy an expensive apartment in Manhattan for the use of the executives when in New York on business. Both purchases can often be justified as sound business practice but sometimes they are made to enhance the executives' leisure consumption. It is not unheard of for corporate jets to be used to fly executives to Superbowl games, baseball spring training sessions, and the like (McMillan, 1992, p. 121). As head of RJR Nabisco in the 1980s F. Ross Johnson bought ten corporate jets and hired thirtysix pilots, and that was just the tip of the Johnson iceberg (Milgrom and Roberts, 1992, p. 493). In some firms the executives reward themselves with exquisite amenities, such as an opulent executive dining room, that cost the company millions of dollars a year. At least one CEO is known to have kept celebrities and athletes on the payroll for retainers of a million dollars a year, apparently for no other reason than to give the executives an opportunity to play golf with the luminaries.

Armand Hammer, the founder and CEO of Occidental Petroleum, used \$120 million of company funds to build a museum to house his personal art collection despite being challenged in court by the shareholders (Milgrom and Roberts, 1992, p. 493). (The dispute was settled out of court.) The direct approach is for a manager to persuade the firm's board to grant an enormous pay raise, far beyond what has been established by convention or is required for appropriate incentives. There are a number of reasons why this strategy often succeeds. For one thing, the board members are often CEOs of other companies, and if they grant an extravagant raise to the manager under their aegis, the bar is raised and thus so is the probability that their own salary will be matched. (The manner in which CEO pay is determined is intensively studied in Bebchuk, Fried, and Walker, 2002.)

Suppose a bank manager, who is also the bank's largest shareholder, makes a loan to a friend on terms guaranteed to result in a loss to the bank. If the borrower makes a secret payment to the manager there can be a net gain for both—at the expense of the other owners, of course. See Akerlof and Romer (1994) for evidence of this kind of fraud. The manager of manufacturing firm Mcan establish a company to supply M with key inputs. If these are priced above the market level—that is, more than other suppliers charge—then the manager will have successfully transferred some of the profit from M to the company the manager owns (Vives, 2000, p. 4). Russian oil companies have been known to sell their oil at absurdly low prices to companies owned by the managers of the oil companies. Korean conglomerates (called *chaebols*) have sold entire subsidiaries to relatives of the founder at low prices. Similar stories have surfaced from Italy (Shleifer and Vishny, 1997, p. 742). For the most part, American courts thwart this extreme form of self-dealing.

As Section 4.3 on the Enron story demonstrates, the top executives can manipulate the price of their company's stock to take advantage of stock options. This is not a new practice. Early in the history of the Ford Motor Company, Henry Ford announced that the company would soon cease paying dividends so that it could provide enhanced benefits to the firm's workers. This maneuver was

successfully challenged in court by the shareholders. It appears that Ford had no intention of carrying out his plan but was attempting to manipulate the price of shares so that he could purchase Ford stock at a reduced price (Allen and Gale, 2000, p. 26).

Bad judgement

In 1921 Ford made 55% of the cars sold in the United States and General Motors (GM) made 11%. GM's business strategy had a number of fundamental flaws. The divisions (Chevrolet, Pontiac, Buick, Oldsmobile, and Cadillac) made very similar cars, so the divisions were competing with each other. The economy was in recession, and car sales were sluggish. Nevertheless, each division continued to overproduce, resulting in unprofitable inventory accumulation. The company did not have a strategy for making division managers take into consideration the cost that inventory accumulation imposed upon GM.

When Alfred P. Sloane took over the helm at GM the company was transformed. Decision making was decentralized. The head of GM made policy—for instance, each division was told to make a car targeted to a particular segment of the market—and each division manager was required to maximize the division's profit subject to guidelines set by the head. In particular, the division's inventory was charged to the division as a cost. Henry Ford, who was still the chairman of Ford and its largest stockholder in the 1920s, vigorously resisted the notion of decentralization. Ford felt that absolute control should flow from the top down. However, a large firm runs more efficiently if it takes advantage of the reduction in agency costs when decentralization is used. (All large modern corporations decentralize, at least to some extent.) By 1940 Ford's market share had fallen to16% and GM's had risen to 45%. (The last two paragraphs are based on Milgrom and Roberts, 1992, pp. 2–4.)

Between 1980 and 1990 GM spent \$67.2 billion on research and development. GM could have purchased Toyota *plus* Honda for that, but by 1990 equity in GM was only \$26.2 billion. The CEO was fired in 1992. General Tire (owned by General Corporation) had substantial excess capacity in 1985 due primarily to the introduction of radial tires, which last three to five times longer than bias-ply tires. Nevertheless, the General Tire management *expanded* capacity.

Incentives can be too strong. Consider the case of Salomon Brothers, the bond trading firm. In the 1980s they had a very comprehensive bonus system involving employees from top to bottom. The firm calculated an employee's contribution to profit from almost every transaction, and bonuses were based to a great extent on an employee's annual contribution to profit. This induced people to work very hard, but it did not yield the best outcome for the firm as a whole. Department *A* might withhold key information from Department *B* if disclosure would benefit *B*. On occasion, a department would "steal" another department's profit. In 1990 Salomon hired Myron Scholes, a Stanford professor who would win the Nobel Prize in Economics seven years later, to reform the incentive system. Scholes's key innovation was to have the employee's bonus money used to buy company stock, with the proviso that it could not be sold for five years. This gives the employee a sufficient interest in the profit of the firm

as a whole, eliminating the incentive for dysfunctional behavior. (This story is from Milgrom and Roberts, 1992, pp. 10–1.)

We explain in Section 5 why a firm's owners may be assumed to be risk neutral: They want the firm to maximize the expected value of profit. The managers, however, are risk averse because a large fraction of their income comes from the firm that they manage. If the managers' pay is a function of their firms' profit, they may avoid decisions that increase the expected value of profit when that would result in a big increase in the variability of profit. This may be the rationale behind *golden parachutes*, which give a manager who is dismissed a huge severance payment. However, if the manager's pay is not sufficiently sensitive to profit then he or she may cause the firm to take *excessive* risk. In the 1980s, managers in the oil industry spent billions of dollars exploring for oil when proven reserves could have been purchased for less than a third of the money. Alternatively, the money could have been passed on to shareholders (Jensen, 1986; see also, McConnell and Muscarella, 1986).

Inertia—sins of omission

Managers who sacrifice shareholder value for personal gain are more likely to take it easy than engage in empire building (Bertrand and Mullainathan, 2003). Profit maximization is a journey into uncharted territory. Managers have to put pressure on themselves to be creative in many dimensions. They have to be on the lookout for new products, new production techniques, and so on. Just because profit is high doesn't mean it has been maximized. Sometimes, an opportunity for increasing profit has already been demonstrated by another firm, yet the manager doesn't adopt it. For instance, banks in Australia and the United Kingdom offer personal accounts that automatically move a customer's money into the highest yielding account—including paying down one's mortgage. This service is available to business customers in the United States but not to individuals (Nalebuff and Ayres, 2003). This personal service has been a great success in countries that have tried it. One would think that an American bank could attract customers away from rivals by introducing it. Why don't they? Surely a bank's owners would favor such an innovation.

Providing appropriate incentives to the firm's other workers is a key part of the management team's assignment. Managers can be considered to be shirking if they do not put much effort into solving the problem of shirking by the firm's other employees. Here are two examples: The Safelite Glass Corporation, which installs car windshields, began using piece rates in the mid-1990s. It now pays a worker according to the number of windshields installed. The firm's productivity (output per worker) increased by 44% as a result, and profit also went up. This was due in part to the incentive to work quickly, and in part to self-selection because workers who knew themselves to be unwilling or unable to pick up the pace left the firm for jobs that did not involve piece rates. The danger with piece rates is that workers might skimp on quality to increase the rate of output. But Safelite used a computer chip to tag a windshield so that the worker who installed it could be identified. U.S. shoe manufacturers switched *away* from piece rates to an hourly wage because of problems such as unreliable quality. (The Safelite story

is based on Lazear, 2000, and the shoe manufacturing story is based on Freeman and Kleiner, 1998.) Motivating workers is becoming increasingly important to the modern firm because human capital is becoming more and more central to the firm's operations (Rajan and Zingales, 2000).

Less than half of the value of stock options granted to employees in the United States are awarded to individuals in or near the top executive category. This means that more than half of the stock options are granted to workers who have little or no ability to affect the firm's overall performance. Moreover, this is an expensive form of compensation for the firm. Hall and Murphy (2000, 2002) estimate a firm's cost of granting an option to one of its employees—essentially, the revenue that would have been earned by selling the option on

In which category would you put the production of defective tires by Firestone in the mid-1990s? The defects resulted in 271 deaths and 800 injuries. The problem has been traced to the hiring of replacement workers at the company's Decatur plant during a labor dispute (Krueger and Mas, 2004). the market—and compare it to the value to the employee receiving the option, which is roughly half of the cost to the firm. Why is the value to the employee so much lower? Because the employee is undiversified and is prevented by law and employer policy from hedging the risk of holding so much in the stock of one company. Why are stock options granted to subexecutive workers when they are so costly to the firm and have no incentive effect? Hall and

Murphy (2003) argue that it is because managers and boards of directors are too enamored of the fact that granting options does not require an *immediate* cash outlay. (Other accounting considerations play a role as well.)

4.3 The Enron debacle

According to its own financial statements, Enron was the seventh-largest American corporation in December of 2000 when its shares were trading for \$84.87. By November 28, 2001, the share price was below \$1 and the company filed for bankruptcy a few days later.

Enron's profit came primarily from arbitrage—buying energy where it was priced low and selling it where it commanded a high price. The Enron

Arbitrage *per se* is socially valuable: A commodity will command a high price in a market where it is in short supply. It is more abundant in a market where the price is low. By buying in a low-price market and selling in a high-price market the arbitrageur moves some of the good from the high-supply area to the low-supply area.

management team explicitly adopted the arbitrage strategy, in preference to actually *producing* electricity, which requires a large stock of expensive equipment.

Arbitrage can be very profitable, as it was initially for Enron. Enron got a head start in newly formed energy markets of the 1980s, as countries around the world restructured their former state monopoly energy industries—the United States included, of course. However, as other companies followed suit, learning from

their own experience and from Enron's, the opportunities for Enron to buy cheap and sell dear greatly diminished. Couple that with the emphasis that Wall Street placed on revenue growth in the 1990s, and you have the seeds of the debacle.

Enron executives exploited the considerable discretion available to them under GAAP (generally accepted accounting practices). For instance, they reported gross revenue from future electricity deliveries as if it were net revenue—that is, they did not deduct the cost of buying the electricity. They made a large sale to at least one company while promising to reverse the transaction at a future date but recorded the proceeds of the sale as revenue. They produced an indecipherable balance sheet and took advantage of Wall Street's preoccupation with revenue growth. In many cases, Enron covered losses from particular ventures by borrowing hundreds of millions of dollars, adding the proceeds of the loan to reported profit while keeping the loss off its books by attributing it to a "partnership" (Malkiel, 2003, pp. 99–100). "The desire of Enron's management to maintain initial revenue and profit growth rates despite the growing sophistication of its competitors created very strong incentives for its management to engage in many of the dubious accounting practices and risky business ventures that ultimately led to Enron's bankruptcy" (Wolak, 2002).

Because Enron's management strategy led to bankruptcy it was certainly not in the owners' long-run interest. How did it benefit the top executives? Enron's spurious claims of high revenue growth led initially to large increases in the price of Enron shares. That made it enormously profitable for management to exercise stock options, making the top executives fabulously wealthy. In principle, stock options give management the incentive to maximize profit because increases in profit lead to increases in the price of the shares on the stock exchange. Because managers are rarely required to hold their shares for any length of time, they have an inordinate interest in short-run profit maximization. They can "earn" tens of millions of dollars in a few years, and when that is a possibility the interests of the executives and the owners diverge. Not that every CEO will exploit the opportunity to acquire vast wealth with reckless disregard for the value and longrun viability of the firm. Presumably, most executives strive to carry out their responsibilities faithfully, and the stock option carrot works in the shareholders' interest if management feels itself ethically constrained to exploit stock options in a way that also enhances the welfare of the firm's owners. However, the stock option carrot can attract unscrupulous individuals whose guiding principle in life is to take as much as they can get away with.

Source

The subsection is based on Wolak (2002).

Link

Holmström and Kaplan (2003) demonstrate that in spite of the corporate board and governance scandals that shook the public's faith in the management of American companies, the system has performed well overall, both in comparison with the periods before and after the scandals broke in 2001 and relative to other countries.

4.4 Why shareholders allow managerial shirking

Why are managers able to take decisions that enhance their own financial positions at the expense of the owners? Why don't the directors prevent it? In part

because executives control the flow of information about their company and can sometimes withhold information from the board if it reflects adversely on management. Also, managers often award consulting contracts to directors and the firms with which they are associated. CEOs often select the directors themselves, and these directors are often CEOs of other companies. Moreover, they frequently serve on a large number of boards and are stretched thin. Directors with a reputation for challenging CEOs will find their invitations to serve on boards drying up (Bebchuk, Fried, and Walker, 2001).

A key point is that a shareholder who owns a small fraction of the company has no incentive to incur the costs of monitoring a management team—the potential benefit to the group of shareholders as a whole is enormous, but the gain to a small individual shareholder will be small. This is an instance of the *free rider problem*: If monitoring requires a high fixed cost and yields a relatively small benefit to the individual, then no owner can gain by absorbing the monitoring cost. But the net gain to the owners as a whole can be vast.

Example 4.3: No owner has an incentive to monitor the manager

Firm *X* has a large number of shareholders, each of whom owns 2500 shares. The shares of firm *X* are currently trading for \$40. Therefore, the value of each person's holding is $2500 \times $40 = $100,000$. Monitoring the manager would increase the value of shares by 50%, but monitoring would cost \$120,000. (A consulting firm would have to be engaged to conduct in-depth research.) No single shareholder is willing to pay \$120,000 to increase his or her wealth by \$50,000.

Clearly, if an individual had enough at stake the entire cost of monitoring could be absorbed by that person and still leave a net gain.

Example 4.4: A firm with a large stakeholder

Firm *Y* has many shareholders, one of whom (individual *J*) holds 12,500 shares. *Y*'s shares are currently trading for \$40, and thus *J*'s holdings are worth 12,500 × 40 = 500,000. If *J* incurred the \$120,000 monitoring cost, and the value of each share rose by 50% as a result, J's wealth would increase by \$250,000 - \$120,000 = \$130,000.

Firm *Y* will be monitored by one of the owners, but firm *X* will not be. Is there evidence for this? Bertrand and Mullainathan (2000) examined CEO contracts before and after the introduction of legislation that made it more costly for an outsider to mount a successful hostile takeover. (More than half of the states in the United States have adopted such laws.) When hostile takeovers become more costly, managers are subject to weaker discipline. Will the owners substitute another form of discipline, or will the CEOs seize the opportunity to increase their pay? Firms with at least one owner holding a fairly substantial fraction of the shares responded to the change in the legal environment that diminishes the market discipline on CEOs by increasing the incentive component of executive contracts, but other firms tended not to do so. In fact, in firms without a large

shareholder the salary part of the manager's pay tends to increase when external discipline weakens. (See also Shleifer and Vishny, 1986, and Bertrand and Mullainathan, 2001.)

Performance bonuses have some effect on pay and hence on performance. In the United States, however, the effect on performance is too narrowly focused on the short run. From the standpoint of both consumer and shareholder welfare, it is long-run profit that should be maximized. If the U.S. stock market is sensitive to short-run profit maximization more than long run, then to the extent that changes in the value of a company's stock affect its managers' performance, it is short-run profit maximization that is encouraged. (We might expect to see a reduction in research and development spending by newly acquired U.S. firms. The evidence is mixed, according to Hall, 1988.) Why might the U.S. stock market be too insensitive to the long run? More than 50% of the common stock is held by pension funds, mutual funds, educational endowments, and charitable foundations, and these institutions account for 80% of the trading (Bernstein, 1992). A mutual fund seldom holds more than 1% of the outstanding stock of a company, and-to ensure diversification-it is illegal for a mutual fund or pension fund to hold more than 10% of the stock of its sponsoring company. The significance of this is demonstrated by Examples 4.3 and 4.4. Management would be more intensely scrutinized if ownership were more concentrated.

In short, most of the stock in a large U.S. company is held by institutions who hold only a tiny fraction of its shares and who trade them frequently. This means that the majority of owners have only a very short-run interest in the company, and the executives themselves stay with the company for only five years on average. (In Japan it is typically a lifetime. Worker-managed firms are springing up across the United States, and the worker-managers typically have a long-term interest in their business. See Harrison, 1993.) Who, then, will put pressure on management to consider the long view? In the United States only 21% of research and development funding in the private sector is targeted for long-run projects; this contrasts with 47% in Japan and 61% in Europe. (The profitability of a randomly selected firm may not increase by anything close to 10% as a result of an investigation of management practices. But firms that are suspected of being poorly managed may well be capable of yielding 10% more profit.)

Sources

Because more than a hundred articles were used in preparing Section 4, the citations have been inserted into the text at the relevant points.

Links

Kotowitz (1989) is a general but brief introduction to hidden action problems. Radner (1992) examines the role of hierarchy in the managerial process. See Easterbrook (1986); Jarrell, Brickley, and Netter (1988); Jensen (1988); Leland (1992); Scherer (1988); and Shleifer and Vishny (1988) for more on takeovers. Hall and Murphy (2003) provide a thorough examination of the role of stock options in American executive compensation and employee compensation in general. See Kanter (1989) for examples of other devices for motivating workers.

5. Agency Theory

Carmichael and McLeod (2000) is a superb treatment of one aspect of this issue. Bebchuk, Fried, and Walker (2002) consider the extent to which managers are governed by incentives and the extent to which they are able to get their own way. They conclude that managers often set the terms of their own compensation, constrained only by the fear of provoking public outrage. Murphy (2002) examines their argument carefully and finds it inadequate.

5 AGENCY THEORY

Consider fictional Hightech Corporation. The manager is the agent, and the set of shareholders constitute the principal. Each owner holds a fraction of the outstanding shares of Hightech, as well as ownership shares in other firms. That is, each Hightech owner has a diversified portfolio. Thus, we assume that "the" principal is risk neutral and simply wants the manager to maximize expected profit.

Example 5.1: The benefits of diversification

A risk-averse individual owns one share each in 50 identical but separate firms. In the case of any firm, if a manager's strategy is passive the profit will be 105 with probability $\frac{1}{2}$ and 95 with probability $\frac{1}{2}$. However, if the manager's strategy is aggressive a firm's profit will be 180 with probability $\frac{1}{2}$ and 60 with probability $\frac{1}{2}$. A passive strategy is less risky but it results in an expected value of 100, whereas the expected value of the aggressive strategy is 120. On one hand, if the manager of each firm is aggressive, then with very high probability close to half of the firms will have profit of 180 and the rest will have 60. Therefore, with very high probability the profit per firm will be close to 120. With very high probability the owner will get a share in a total profit of $50 \times 120 = 6000$ if each manager is aggressive. On the other hand, if each manager is passive, the total profit will be very close to $50 \times 100 = 5000$. The owner is much better off when each manager pursues the risky strategy, even though the owner is risk averse. Diversification reduces the risk of the *portfolio*, even when the individual shares incorporate a lot of risk. Therefore, the diversified owner is risk neutral from the standpoint of the performance of the firms in which he or she owns a share.

In most cases, the *manager's* consumption and utility depends crucially on the pay received for managing the firm. Hence, the manager is typically riskaverse. The manager's effort has a strong influence on the firm's profit but so do random forces. If the manager's pay went up by a dollar every time profit went up by a dollar, and went down by a dollar every time profit fell by a dollar, then the manager would have the strongest possible incentive to maximize expected profit. We say that the manager has *maximum incentive* in that case. But when the manager's pay moves perfectly in step with the firm's profit, that pay is most strongly influenced by the random component of profit. Because the manager is risk averse, that will lower his or her expected utility (EU) unless the manager

is compensated in the form of higher expected pay. The higher expected pay can lower the expected return to the firm's owners.

With maximum incentive, the gross expected profit is highest, but the manager's share of that profit would have to be higher on average because of the manager's exposure to risk. The owner's net expected profit-net of the manager's compensation-is not maximized under maximum incentives. Shareholders face a trade-off between incentives and risk spreading. Compared to a contract in which variations in profit have their full effect on the manager's pay, the shareholders do better when they reduce the manager's exposure to risk by providing an insurance element in the pay package. This weakens the manager's incentive, of course. The insurance market provides an extreme example. The consumer who purchases health insurance can influence the size of claims submitted by means of preventive medicine and by eschewing frills when illness does strike. But the random forces that select one person as a victim of ill health rather than another play a vastly more important role in determining individual medical expenses. Therefore, insurance contracts give relatively little scope for incentives and go a long way toward protecting the individual from random events. (This is discussed in more detail in Section 9 of Chapter 3.) At the other extreme, fast food chains commonly employ franchising: The manager of the local outlet absorbs most of the risk to enable incentives to have a big impact.

When we model the principal-agent relationship, we assume that effort is one dimensional. The agent can supply an additional unit of effort by reducing leisure consumption by one unit. This is the only way that the manager can affect the firm's profit in our formal model. In the real world, managers' activities can deviate substantially from maximizing the owner's return even when managers put in long hours. For instance, a manager can devote considerable effort to concealing data from the directors and shareholders, knowingly undermining the principal's welfare. Happily, we can draw a great deal of insight from a model in which the manager has a simple one-dimensional trade-off between effort and leisure.

5.1 A diagrammatic introduction

We model the principal-agent relationship by abstracting from everything but the inability of shareholders to determine the amount of effort contributed by the manager of their firm, even though effort is correlated with profit. Because profit is also influenced by random forces, the correlation between managerial effort and the firm's profit is not perfect. The owner can only observe profit and thus has to offer the manager a wage schedule that features a dependence of the compensation package on profit alone and will endeavor to structure compensation in a way that induces the manager to apply a high level of effort not the highest possible level of effort, but the level that maximizes the return to the owners.

Although we speak in terms of a manager in relation to the firm's owners, the analysis applies just as well to any principal-agent relationship. The principal can be a university designing a contract for its agent, a football coach. The manager of a privately owned firm is the principal when he or she employs a salesperson.



Figure 4.4

Should the salesperson be paid on commission, and if so at what rate? The agent could be a professor hired by a university, the principal, and so on. In any principal-agent relationship there will be a wide variety of opportunities for shirking. In spite of the fact that shirking is multifaceted, it is modeled here as a one-dimensional sacrifice of effort in return for increased leisure consumption.

We begin by supposing that profit has no random component. It is easy to adapt the argument to cover uncertainty with risk-neutral individuals when we have analyzed the deterministic case. The firm's profit *R* is βe if the manager supplies *e* units of effort. *T* is the time endowment, and *x* is the manager's consumption of leisure. Of course x = T - e.

DEFINITION: Profit in the agency model

When we use the term profit (*R*) in this section we mean revenue minus all costs except the manager's pay. The owner's *net return N* is profit in the usual sense—revenue minus all costs, including the manager's pay.

Figure 4.4 shows the profit function $R = \beta e$ as line *L*. The owner's net return *N* is the difference between *R* and the payment *y* to the manager. Profit

maximization by the owner keeps the manager on the indifference curve u^0 , representing the utility that would be realized by the manager in his or her best alternative. (This is explained carefully in Section 5.2.) The diagram shows three net return levels N_1 , N_2 , and N_3 corresponding to three respective budget constraints for the manager, B_1 , B_2 , and B_3 . For any budget line B, the owner's net return N is the vertical distance between L and the point where B is tangent to the indifference curve; this gives us R - y. If -p is the slope of the budget line then the line can be expressed as px + y = C, or y = p(T - x) + F, or y = pe + F, where F is the constant C - pT.

It is clear that the owner's net return is highest with budget line B_2 parallel to L. Here's why: Assume for a moment that effort is observable and can be mandated in a contract offered by the owner. Consider B_1 , which is tangent to u_0 at C_1 . In other words, the manager will choose basket C_1 if his or her budget line is B_1 . The owner's net return is N_1 . Now, increase the manager's input of effort by changing the budget line, making it steeper, so that the manager has a new consumption plan C' along u^0 to the left of C_1 . B_1 is flatter than L, which means that beginning at C_1 , an increase in effort will increase income more quickly along L than along u^0 . In other words, a reduction in x (caused by an increase in effort) will cause R to increase faster than y. (R is on L, and y is on the indifference curve u^0 .) This means that profit R increases faster than the manager's pay y. Therefore, the owner's net return will increase. This argument applies at any point on u^0 to the right of C_2 , where the tangent to u^0 is parallel to L. Therefore, to the right of C_2 the owner's net return $N = R - \gamma$ increases with e because the manager's consumption plan moves along u^0 from right to left but R increases at a faster rate.

To the right of C_2 on u^0 , the tangent to u^0 (the manager's budget line) gets steeper as we move toward C_2 by increasing the amount of effort required by the agent. Increasing p for budget line y = pe = p(T - x) + F is equivalent to increasing the manager's reward per unit of effort supplied. This increase in p is advantageous to the principal because it allows N = R - y to increase. But if we move beyond C_2 by making p larger than β , the owner's profit will fall. Why? Because u^0 is steeper than L to the left of C_2 , and thus as we move the manager along u^0 to the left of C_2 the manager's consumption y will increase faster than R. Even though R increases, because the manager supplies more effort, y increases at a faster rate so the owner's net return falls to the left of C_2 .

The owner's net return is highest with budget line B_2 parallel to L. (Exercises 8–10 at the end of this section take you through an algebraic proof.) L is the line $R = \beta e = \beta (T - x)$. Lines B_2 and L have the same slope, so B_2 has slope $-\beta$. Then we can write B_2 as $y = \beta (T - x) + F = \beta e + F = R + F$ where R is the firm's realized profit. R depends on the manager's effort, and the manager knows the functional relationship between R and e. Therefore, the contract y = R + F offered by the owner will induce the manager to supply the amount of effort that leaves N_2 for the owner, *even if the owner cannot observe and enforce e*. The contract y = R + F reads "the manager gets all the profit R after delivering the fixed amount -F to the owner."

5. Agency Theory

We can now drop the assumption that effort is observable, because the contract y = R + F transfers all of the social gains or losses from a change in the manager's effort level directly to the manager, who is now the sole residual claimant on the firm's profit. In other words, y = R + F is optimal for the owners because, under that contract, the cost of leisure consumption to the manager is equal to the cost to the firm.

Note that *F* is negative (Figure 4.4). The manager pays a *franchise fee* of -F to the owner and then keeps all profit net of the fee.

DEFINITION: Residual claimant

If two individuals share an amount of revenue that is a function of the input of one or both of those individuals, and the share of one of them is fixed independently of the amount of revenue generated, then the other individual is the residual claimant, receiving whatever revenue is left over after the fixed payment is made to the other person.

This argument applies to production with uncertainty as long as the manager and the owner are risk neutral and the expected value of the random component

Recall the story of Chinese agricultural reform of the 1980s (p. 7 in Chapter 1). Before the reform, the farm had to deliver all of its surplus to the state and hence agricultural output was very low. When the rule changed, allowing the farm to keep the surplus after delivering a fixed amount to the central government, output soared. Under the new rule the agent-the farmer-is the residual claimant and hence has maximum incentive to work efficiently. Consequently, the central government collects more output because it can require a fairly high fixed quota to be supplied by the farm. The same principle explains contemporary amusement park pricing in the United States: The rides are free, so visitors to the park derive a high level of consumer benefit. This allows the park owner to collect a high entry fee at the gate. The owner receives more revenue by giving the rides away and collecting a large fixed fee as the patron passes through the park gate.

is zero. Suppose that $R = \beta e + \xi$, where ξ is a random variable with expected value zero. Then the expected value of *R* is βe , and we apply the analysis to the expected value of *R*, which the owner wants to maximize, net of the payment to the manager. If the manager is risk neutral then *y* enters the manager's utility function linearly. That is, U(x, y) = B(x) + y. If *y* is the expected value of the manager's pay then B(x) + y is the manager's expected utility, and the argument above goes through. In fact, this holds even if $E(\xi)$ is not zero. We assume that $E(\xi) = 0$ to simplify the calculations.

We can apply this discussion to *any* of the firm's workers. The optimal contract requires a wage W = R + F, where *F* is negative. But would we really expect the worker to pay the employer? This incentive scheme would actually provide more utility for the worker. Because it induces efficiency there would be more output per capita in the economy, and competition among employers for workers would result in a higher u^0 (utility from alternative employment). But suppose there is a cash constraint preventing a payment by workers to





Fruit pickers in orange groves are paid a piece rate-a fee per box of oranges. This motivates them to pick quickly. (If they were on salary they would have an incentive to dawdle.) They supply maximum effort in the everyday sense of the word. But the piece rate formula gives workers an incentive to pick the ground fruit first, although oranges on the ground are high in bacteria. Also there is a tendency to take the most accessible fruit from the branches and leave the rest to rot on the tree. Hence, there is shirking in a more general sense, and it is handled by direct monitoring of the workers (McPhee, 1966, p. 55).

employers, or hidden information problems standing in the way of a loan of *F* dollars from the owner to the worker. We can achieve the same outcome by means of progressive piece rates. This is illustrated in Figure 4.5 with budget line ACD. The worker receives a basic salary of S. For (gross) profit levels less than R_C the worker is paid *p* dollars per unit of additional effort supplied. (The slope of the AC segment of the budget line is -p.) For profit above R_C the worker receives β dollars per unit of additional effort supplied. The contract would actually be written so that for output levels less than R_C the worker is paid p/β dollars per dollar of additional profit generated, and for output above R_C the worker receives the whole of each dollar of additional profit generated. That way, the con-

tract does not mention the unobservable *e*. This progressive piece rate system and the contract y = R + F induce identical decisions. But the progressive piece rate system has a serious hidden action defect. Unless the quality of output can

5. Agency Theory

be easily verified—bushels of wheat, for example—the worker has an incentive to work quickly, sacrificing quality, to reach the output level R_C where the higher piece rate is available. There is no danger of this with the contract y = R + F because the agent bears the full brunt of any production decision that affects profitability.

5.2 The basic agency model

We assume that the owner of the firm is risk neutral. The manager contributes a level of effort *e* that is unobservable. Before making a payment to the manager the firm's profit is $R(e, \xi)$, a function of the effort *e* supplied by the manager and of a random variable ξ . The firm offers the manager a compensation package *w* that is a function w(R) of the realized profit. Although *R* will depend in part on *e*, the effort level *e* is unobservable so the manager's contract will depend only on the actual, observable profit *R*. The manger can achieve a utility level u^0 by working elsewhere so the compensation schedule must allow the manager to achieve a level of expected utility at least as high as u^0 . This is called the *participation constraint*.

DEFINITION: The participation constraint

Managers will not accept contracts if they do not allow them to reach the highest expected utility level u^0 that can be attained by working elsewhere.

The manager's utility is U(x, y), where x is leisure consumption and y is monetary compensation—think of it as income. We let *EU* denote the manager's expected utility. We let T represent the initial endowment of time, a constant. Therefore, x = T - e. Note that y is a random variable, because it equals W(R), and R depends on e and ξ . (Assume for convenience that the manager does not have an endowment of Y.) The manager will maximize EU—that is, the expected value of

$$U(x, y) = U(T - e, w[R(e, \xi)]),$$

subject to EU being at least u^0 . The maximization exercise induces a dependence of effort on the compensation schedule, and the owner can use that relationship in designing a contract.

Suppose that U(x, y) = B(x) + y and the compensation schedule is a member of the linear family $\theta R + F$, where θ and F are constants. (Think of θ as the commission rate paid to a salesperson, the share of taxi revenues going to the driver, or the royalty rate paid to a textbook author. In each case, the individual in question appears in our model as the manager.) The manager is risk neutral in this case, because x does not depend on the random variable ξ , so the manger's EU is B(x) + E(y), where E is the expectation operator. Let $E(y) = E(\theta R + F) = \theta E(R) + F$. Now, maximize

$$EU = B(T - e) + \theta E[R(e, \xi)] + F.$$

This is a function of *e*, which the manager controls, given the owner's choice of θ and *F*. (The random variable ξ disappears when we take the expected value.)

DEFINITION: The agency model

Variables *x*, *e*, and *y* denote, respectively, the manager's leisure consumption, effort, and income. $R = \beta e + \xi$, where *R* is the firm's profit before deducting the manager's pay, β is a given positive constant, and ξ is a random variable. The time endowment is *T*, and thus e = T - x. If the manager's utility function U(x, y) is quasi linear, with U = B(x) + y, then the manager is risk neutral. The manager's best alternative employment yields an expected utility of u^0 . The manager will be offered a contract that determines his or her pay as a function w(R) of profit. If $w(R) = \theta R + F$ then it is a member of the linear family of contracts.

Given θ , let e_{θ} be the value of e that maximizes EU. (The constant F does not influence the maximizing value of EU. But it does play a role via the participation constraint.) Then e_{θ} is the amount of effort supplied by the manager when facing the compensation schedule $\theta R + F$. If θ were increased then the manager's opportunity cost of leisure consumption increases because the manager now gets a larger fraction of an additional dollar of profit generated by increased effort. We would expect effort supply to increase.

Example 5.2: The effort supply function

Let $U(x, y) = \alpha x - \frac{1}{2}x^2 + y$ and $R(e, \xi) = \beta e + \xi$. Assume also that $E(\xi) = 0$. Then $E[R(e, \xi)] = \beta e$. If $y = \theta R + F$ then $E(y) = \theta E[R(e, \xi)] + F$ and we have $E(y) = \theta \beta e + F$, and thus

$$EU = \alpha x - \frac{1}{2}x^{2} + \theta \beta e + F = \alpha (T - e) - \frac{1}{2}(T - e)^{2} + \theta \beta e + F$$

because x = T - e. And $\alpha (T - e) - \frac{1}{2}(T - e)^2 = \alpha T - \alpha e - \frac{1}{2}T^2 + Te - \frac{1}{2}e^2$, so we have

$$EU = (\theta\beta - \alpha + T)e - \frac{1}{2}e^2 + \left[\alpha T - \frac{1}{2}T^2 + F\right].$$

The terms inside the square brackets are constant—that is, independent of the manager's choice of *e*. We maximize *EU* by employing calculus or the formula for maximizing a quadratic. We get the effort supply function

$$e(\theta) = \theta\beta - \alpha + T.$$

This does increase when θ , the manager's share of a marginal dollar of profit, increases.

Note that when the manager is offered the contract $y = \theta R + F$, the parameters θ and *F* are constants from the manager's perspective. From the owner's
5. Agency Theory

standpoint, θ and *F* are variables, chosen by the owner to maximize the owner's net return subject to incentive compatibility (given the contract, the manager will choose *e* to maximize *EU*) and the participation constraint ($EU \ge u^0$). Because the maximization of the owner's net return requires $EU = u^0$, this induces a functional dependence of *F* on θ because *F* is a part of *y*, which is a part of *EU*.

Now, let's relax the assumption that the manager's contract belongs to the linear family $y = \theta R + F$. It can be any function of R. What is the owner's objective? The owner is risk neutral and supplies no effort, so the owner simply wants to maximize the expected value of R net of the payment to the manager. In other words, the owner seeks to maximize the expected value of $N = R(e, \xi) - w[R(e, \xi)]$. We have seen that e depends on the compensation schedule w, so we let $e^*(w)$ denote the manager's effort supply when the manager is offered the contract w. Therefore, the owner chooses w to maximize the expected value of

$$R[e^{*}(w), \xi] - w(R[e^{*}(w), \xi])$$

subject to $EU \ge u^0$. How do we solve for *e*? Effort depends on the compensation schedule via the manager's optimization problem, but the owner's optimization problem causes the compensation schedule to depend on the manager's effort supply function. We'll start with an easy case, that of a risk-neutral manager.

5.3 Risk-neutral managers

The owner of the firm is risk neutral, and in this subsection we suppose that the manager is risk neutral as well. Then U(x, y) = B(x) + y for some function *B*. We start by observing that E(R) is a function of *e*. The random variable influences the value of the expected value E(R), and E(R) itself depends on *e*. The participation constraint is $E[B(x) + y] \ge u^0$. Note that E[B(x) + y] = B(x) + E(y).

Profit maximization implies that the owner will choose a compensation schedule that equates the manager's EU with u^0 . Why? We know that $EU \ge u^0$ must hold. If the manager's EU actually exceeded u^0 the owner could reduce the compensation offered for each realization of the random variable without violating the participation constraint $EU \ge u^0$. This would increase the owner's return. Therefore, at equilibrium we must have $EU = u^0$.

We assume temporarily that the owner can observe and mandate *e*. This allows us to find e^* , the level of effort that maximizes the owner's net return subject to the participation constraint (but without imposing the incentive compatibility constraint, which recognizes that the agent must have an incentive to set $e = e^*$). Then we will discover that there is a contract that induces the manager to choose e^* even though *e* is not observable and the manager knows it. Because profit maximization implies $EU = u^0$, we have $B(x) + E(y) = u^0$, and thus $-E(y) = B(x) - u^0$. The owner then will maximize E(N) subject to $-E(y) = B(x) - u^0$. Now, $E(N) = E(R) - E(y) = E(R) + B(x) - u^0$. Because x = T - e, this can be considered a function of *e*. The owner wants to maximize

$$f(e) = E(R) + B(T - e) - u^0$$
.

If the manager receives the contract y = R + F, where *F* is a constant, then the manager will maximize EU = E[B(x) + y] = B(x) + E(y) = B(x) + E(R + F) = B(x) + E(R) + E(R) + F. That is, the manager will maximize

$$g(e) = B(x) + E(R) + F.$$

Again, x = T - e, so g really is a function of e.

Compare *f* and *g*. They differ by a constant: $g(e) = f(e) + F + u^0$. Therefore, e^* maximizes *f* if and only if e^* maximizes *g*. This means that the contract y = R + F induces the manager to select the effort level that the owner would insist on if the owner could observe and enforce *e*. With effort supply determined, *F* is the solution to $B(T - e^*) + E[R(e^*, \xi)] + F = u^0$.

The effort supply e^* that maximizes both f and g satisfies the participation constraint because it is built into g. Therefore, the manager would accept the contract y = R + F. Having done so, the manager maximizes her expected utility by setting $e = e^*$. Even though e^* is the level of effort that would be mandated if the owner had full information, it is chosen by the manager even when effort is not observable. Let M = -F. Then $M = E[R(e^*, \xi)] - u^0 + B(T - e^*)$. The compensation schedule would give the manager the actual realized profit minus the constant M. To verify that it would be in the manager's interest to supply e^* *if* she accepted the contract let's compute the manager's EU for the compensation schedule w(R) - M:

$$EU = B(T - e) + E[R(e, \xi)] - M.$$

Although the actual return *R* varies with ξ , *M* is a number—an expected value so any value of *e* that maximizes the manager's EU also maximizes the owner's profit *f*(*e*). Because *e*^{*} was defined as the value of *e* that maximizes $B(T - e) + E[R(e, \xi)]$ it is in the manager's interest to set $e = e^*$. *If* the manger accepts the contract she will choose the effort level *e*^{*}. But will she accept? By definition of *M* we have

$$B(T - e^*) + E[R(e^*, \xi)] - M = u^0,$$

so the compensation contract w(R) = R - M does allow the manger to achieve the EU level u^0 . Note that the profit-maximizing pay schedule is $y = \theta R + F$ for F = -M and $\theta = 1$. In practice, the compensation contract would offer slightly more utility than u^0 to ensure that the manager will accept the contract in preference to the best alternative, which yields a utility level of u^0 .

Optimal contract for risk-neutral managers

The manager pays a lump sum to the owner and keeps all remaining profit. In other words, the manager becomes the residual claimant. But the payment to the owner is set so that the manager's participation constraint is satisfied as an equality.

5. Agency Theory

Here is a simple example that allows us to explicitly solve for the manager's choice of e as a function of w and then to solve for the profit-maximizing pay schedule w.

Example 5.3: Deriving the optimal contract

The manager's utility function is $U(x, y) = 20x - \frac{1}{2}x^2 + y$. That is, $B(x) = 20x - \frac{1}{2}x^2$. Set T = 24. (The manager is endowed with 24 units of *X* and 0 units of *Y*.) We assume that the manager's best alternative is to consume 24 units of *X*, so $u^0 = 20(24) - \frac{1}{2}(24)^2 = 192$. The production function is $R = 10e + \xi$, with $E(\xi) = 0$. Then E(R) = 10e. Assume for a moment that the owner can observe and mandate the effort level *e*. What *e* would he select? We know that the contract that maximizes the expected value of the owner's net return satisfies $EU = u^0$. In this case we have EU = B(x) + E(y) = 192. Then E(N) = E(R) - E(y) = E(R) + B(x) - 192. Therefore, the owner maximizes

$$f(e) = 10e + B(x) - 192 = 10e + 20(24 - e) - \frac{1}{2}(24 - e)^2 - 192$$
$$= 14e - \frac{1}{2}e^2.$$

Note that *f* is a quadratic, and when we apply the formula for maximizing a quadratic we get $e^* = 14$. (Here is the calculus derivation: f'(e) = 14 - e, and f''(e) < 0. Therefore, we set f'(e) = 0 to maximize the owner's expected profit. This yields $e^* = 14$.)

Return to the case of unobservable effort. We show that the contract y = R + F induces the manager to set e = 14: If y = R + F then the manager's EU is

$$EU = B(x) + E(y)$$

= B(x) + 10e + F
= 20(24 - e) - $\frac{1}{2}(24 - e)^2 + 10e + F$
= 14e - $\frac{1}{2}e^2 + 192 + F$.

This function is maximized at $e^* = 14$. (We could have employed a shortcut. In Example 5.2 we derived the effort supply function. It is $e = 4 + 10\theta$ when $\alpha = 20$, T = 24, and $\beta = 10$. Therefore, when $\theta = 1$ the manager will supply the effort $e^* = 14$ that maximizes the owner's profit, even though the owner cannot observe or enforce *e*.)

Now, compute *F* under the profit-maximizing contract: E(w) = 10e + F = 140 + F and x = 24 - 14 = 10. Therefore, the manager's EU is

$$20(10) - \frac{1}{2}(10)^2 + 140 + F = 192.$$

Then F = -98. The manager pays the owner a license fee of \$98 and then keeps the remaining profit.

The manager's contract is y = R - 98. We have E(R) = 10e, so the contract is equivalent to $\bar{y} = 10e - 98$, where \bar{y} is the expected value of y. (The manager can observe her own effort supply of course.) If $\bar{y} = 10e - 98$ then $\bar{y} = 10(24 - x) - 98$. Therefore, $\bar{y} = 240 - 10x - 98$, and thus the contract that maximizes the

242 Corporate Governance

owner's net return allows the manager to choose a commodity bundle from the budget line $10x + \bar{y} = 142$. Now, show that an individual with utility function $U(x, \bar{y}) = 20x - \frac{1}{2}x^2 + \bar{y}$ and budget constraint $10x + \bar{y} = 142$ will choose the bundle $(x, \bar{y}) = (10, 42)$.

Note that for Example 5.3 we have proved that the contract y = R - 98 yields a higher expected profit to the owner than *any* other contract. This is much stronger than merely proving that y = R - 98 is profit maximizing within the family of linear contracts. Nevertheless, you might benefit from solving directly for the profit-maximizing values of θ and *F* within the linear family.

Example 5.4: Using the effort supply function to derive the optimal contract

Assume the setup of Example 5.3. The effort supply function is $e = 4 + 10 \theta$ when $y = \theta R + F$. (Recall Example 5.2.) Profit maximization and the participation constraint imply $B(x) + E(y) = u^0 = 192$. Therefore,

$$E(y) = 192 - B(x) = 192 - 20x + \frac{1}{2}x^{2}$$

= 192 - 20(24 - e) + $\frac{1}{2}(24 - e)^{2}$
= 192 - 20(24 - 4 - 10 θ) + $\frac{1}{2}(24 - 4 - 10 \theta)^{2}$

because $e = 4 + 10 \theta$. Therefore,

$$E(y) = 192 - 20(20 - 10\theta) + \frac{1}{2}(20 - 10\theta)^2 = -8 + 50\theta^2.$$

Now, the owner wants to maximize E(R) - E(y). Using the last equation and the fact that E(R) = 10e,

$$E(R) - E(y) = 10e + 8 - 50 \theta^{2}$$

= 10(4 + 10 \theta) + 8 - 50 \theta^{2}
= 48 + 100 \theta - 50 \theta^{2}.

This is a function of one variable, θ . Using the formula for maximizing a quadratic, we get a maximum at $\theta = 100/100 = 1$. (The first derivative of E(R) - E(y) is $100 - 100 \theta$, and the second derivative is negative. Therefore, we achieve a maximum by setting $100 - 100 \theta = 0$.) We see that $\theta = 1$ maximizes the owner's expected profit. We solve for F = -98 as in Example 5.3.

The manager bears all the risk; the owners receive a constant return of M(=-F). This is equivalent to an arrangement in which the owners sell the firm to the manager for a price of *M* dollars. The incentive scheme that induces the optimal effort is equivalent to having the owner manage the firm herself. Is this plausible? Risk neutrality itself is plausible only if the manager's income

5. Agency Theory

from the firm is a small component of her portfolio (or income sources). In other words, the manager is diversified and *M* represents a small fraction of her assets.

In the Middle Ages, when a monarch wanted to collect taxes from subjects in remote regions, he or she would sometimes give the job to a senior church official, who would pay the monarch a fee for the privilege and then keep the taxes that the official collected. This is clearly an early example of franchising (Thompson, 1971). This is implausible in the case of a typical firm and a typical manager. It is usually the other way around. The value of a typical firm is many times greater than an executive's wealth.

Franchises, however, are relatively small. The franchise situation, in which the manager pays a fee *M* to the parent and keeps the residual profit, comes close to the outcome outlined in this section. Moreover, it is more important for the manager to be the residual claimant in the case of a chain of franchises because the pro-

duction units—the local franchises—are widely scattered and hence difficult for the head office to monitor. In the United States, 10% of consumer all retail spending is received by franchises, which employ more than 6% of the workforce.

With risk-neutral management, the equilibrium resulting from maximization of the manager's utility and of the owner's profit will be efficient.

The optimal contract y = R + F is efficient if the manager is risk neutral.

We have three ways of demonstrating this:

- We show that the equilibrium e^* maximizes the *sum* of the expected utilities. (There may be other efficient outcomes, but anything that maximizes the sum of utilities will belong to the set of efficient outcomes.) The owner's EU is just the expected value of y_1 , the net return to the owner. The manager's EU is $B(x_2) + E(y_2)$, where y_2 is the payment from the owner to the manager, and $E(y_2)$ is its expected value. Therefore, we can find an efficient outcome by maximizing $E(y_1) + B(x_2) + E(y_2) = B(T - e) +$ $E(y_2 + y_1)$. Now, $y_1 + y_2$ is $R(e, \xi)$, the gross return to effort. We get an efficient level of *e* when we maximize B(T - e) + E(R). But this differs from *f* and *g* only by a constant, so all three functions are maximized by the same e^* . Hence, e^* is efficient.
- The optimal contract is efficient because the social cost of leisure consumption equals the private cost of leisure consumption when $\theta = 1$. The social cost of leisure consumption is always the change in E(R) when the manager reduces *e* by one unit. Of course, when $\theta = 1$ this is also the cost to the manager of increasing her leisure consumption by one unit.
- The optimal contract is obtained by maximizing the owner's utility subject to the manager's utility not falling below a specified level. In any context, any solution s^* to the problem "maximize U_1 subject to $U_h \ge u_h^0$ for all $h \ne 1$ " is efficient. (Note that u_h^0 is a constant for each h.) If s^* were not efficient there would either be an alternative s such that $U_1(s) > U_1(s^*)$ and $U_h(s) \ge U_h(s^*)$ for all $h \ne 1$, contradicting the fact that s^* solves the constrained maximization problem, or else an alternative s such that $U_1(s) \ge U_1(s^*)$, $U_h(s) \ge U_h(s^*)$

for all $h \neq 1$, $U_j(s) > U_j(s^*)$ for some $j \neq 1$. In the latter case, we could transfer a positive but sufficiently small amount of money from j to 1, resulting in an alternative s' such that $U_1(s') > U_1(s) \ge U_1(s^*)$ and $U_h(s') \ge U_h(s^*)$ for all $h \neq 1$, contradicting the fact that s^* solves the constrained maximization problem. Therefore, s^* is efficient.

Finally, we can drop the assumption that effort is one dimensional. Because the optimal contract makes the manager the residual claimant, the manager

Suppose that effort has two dimensions, quality and quantity. If the agent is not the residual claimant, giving a strong quantity incentive can result in severe shirking on quality. Consider this report from a worker in a Baltic firm producing television sets. It describes conditionsprior to the collapse of communismtoward the end of the month as the employees strive to earn bonuses: "We never use a screwdriver in the last week. We hammer the screws in. We slam solder on the connections, cannibalize parts from other televisions if we run out of the right ones, use glue or hammers to fix switches that were never meant for that model. All the time the management is pressing us to work faster, to make the target so we all get our bonuses" (Cook, 1990, quoted in Milgrom and Roberts, 1992, p. 14).

has an incentive to adopt any measure, and to modify any part of her decision strategy, that will increase profit.

To highlight the significance of making the agent the residual claimant we have oversimplified the relationship between the franchise and the parent corporation. In fact, the parent supplies important inputs, primarily national advertising and the enforcement of standards. It's obvious why the franchisee benefits from an advertising campaign. The franchisee also benefits when the parent enforces standards across the board because the customer then comes to expect a uniform product at each of the franchise outlets. In other words, risk-averse consumers benefit from the reduced uncertainty that results from the parent enforcing standards, and this makes each franchise's output more valuable to consumers. In return for supplying these inputs the parent receives a royalty of 2% or 3% of the franchisee's revenue, in addition to the fixed franchise fee. The royalty

payment gives the parent a direct financial stake in the franchise and hence an incentive to supply advertising and standards enforcement optimally—or close to optimally.

5.4 Risk-averse managers and binary effort supply

We now turn to the case of a risk-averse manager of a large corporation. If you are unfamiliar with the elements of decision making under uncertainty you will need to read Section 6 of Chapter 2 before continuing.

To give us a point of comparison, suppose (temporarily) that *e* can be observed by the owner and verified by a court. This means that a contract can specify the effort level contributed by the manager. The owner can insist on a particular effort level e^* . Let $w(e^*, \xi)$ represent the compensation package offered to the manager. If this is not a constant, independent of ξ , let $C = E[w(e^*, \xi)]$, which is a constant. If the risk-averse manager had a choice between $w(e^*, \xi)$ and a constant salary that paid *C* whether profit was high or low then she would choose *C* because it has a higher expected utility. That follows from risk aversion

5. Agency Theory

Table	4.	1
-------	----	---

	<i>e</i> = 1		<i>e</i> = 0	
	State 1	State 2	State 1	State 2
<i>R</i> Probability	70 ¹ / ₂	46 ¹ /2	29 ¹ / ₂	22 ¹ / ₂

and fact that the constant salary *C* has the same expected monetary value as *w* but *C* offers complete certainty. Therefore, for $\delta > 0$ sufficiently small, a constant wage of $C - \delta$ would yield a higher expected utility than $w(e^*, \xi)$, and it would satisfy the participation constraint for δ sufficiently small, because $U(T - e^*, C)$ is higher than the expected utility yielded by the contract $w(e^*, \xi)$ which itself satisfies the participation constraint. The return to the owner from $w(e^*, \xi)$ is $E[R(e^*, \xi)] - E[w(e^*, \xi)] = E[R(e^*, \xi)] - C$, and the return from $C - \delta$ is $E[R(e^*, \xi)] - C + \delta$. The constant salary $C - \delta$ would give the owner a higher expected profit than a variable schedule that gave the manager *C* in expectation. Therefore, the manager's compensation would be constant if monitoring were costless.

With observable and verifiable effort, incentives play no role because the profit-maximizing effort level can be mandated by the owner. Therefore, the manager receives a fixed payment (the constant salary), independent of random forces, and hence is fully insured. At the other extreme, with unobservable effort and risk-neutral management, the *owner's* return is fixed and the manager bears the full brunt of the vicissitudes of nature. This gives the manager the optimal incentive, from the standpoint of both society and the owner. We expect that if the manager were just a tiny bit risk averse then there would be a small constant element to the compensation, with the manager bearing almost all of the brunt of uncertainty.

We begin with a binary version of the model: The time endowment is T = 1, and the manager can either work (e = 1) or shirk (e = 0). We present an example in which the manager is risk averse and the contract that maximizes the owner's return does *not* make the manager the residual claimant. In fact, the optimal contract will pay the manager a fixed salary, to which the manager's best response is to set e = 0. After proving this we will go on to a richer model in which the optimal contract offers that manager a share θ of the profits strictly between 0 and 1.

Example 5.5: Binary choice of effort level

A risk-averse manager has the utility function $U = 4x + \sqrt{y}$, where *x* is leisure consumption and *y* is the manager's pay. The time endowment (i.e., the leisure consumption when effort is 0) is 1. Effort, *e*, supplied by the manager is either 0 or 1. The manager's reservation utility u^0 is 6. Recall that *R* denotes profit before deducting the manager's pay or the return to the owner. Table 4.1 specifies the

return to the manager's effort. (If e = 1 then R = 70 with probability $\frac{1}{2}$ and R = 46 with probability $\frac{1}{2}$. If e = 0 then R = 29 with probability $\frac{1}{2}$ and R = 22 with probability $\frac{1}{2}$.)

To confirm that the manager is risk averse consider an asset that pays \$64 with probability $\frac{1}{2}$ and \$36 with probability $\frac{1}{2}$. The expected monetary value is \$50. Given *x*, the EU from \$50 for sure is 4x + 7.07 and the EU from the asset is $4x + \frac{1}{2}\sqrt{64} + \frac{1}{2}\sqrt{36} = 4x + 7.0$, which is less than the EU of \$50 for sure, as required by risk aversion. (The derivative of utility with respect to wealth is $\frac{1}{2}y^{-0.5}$, given *x*, and thus the second derivative is $-\frac{1}{4}y^{-1.5}$ which is negative.)

Which linear contract $y = \theta R + F$ maximizes the owner's net return subject to the participation and incentive compatibility constraints? If e = 0 then incentives play no role, in which case E(N) is maximized by a constant salary, with $\theta = 0$. The participation constraint then requires $EU = 4 + \sqrt{F} = 6$. Therefore, F = 4. When e = 0 we have $E(R) = \frac{1}{2} \times 29 + \frac{1}{2} \times 22 = 25.5$. Therefore, if e = 0 the owner's maximum net return is 25.5 - 4 = 21.5. Of course, $\theta = 0$ implies e = 0. Therefore, it remains to determine if there is a contract with $\theta > 0$ that yields an expected net return to the owner that is greater than 21.5. That can be the case only if it induces the manager to set e = 1.

If e = 1 and EU = 6 then we have

$$EU = 0 + \frac{1}{2}\sqrt{70\theta + F} + \frac{1}{2}\sqrt{46\theta + F} = 6.$$

Then $\sqrt{70\theta + F} = 12 - \sqrt{46\theta + F}$. Squaring both sides yields $70\theta + F = 144 - 24\sqrt{46\theta + F} + 46\theta + F$, and hence $\sqrt{46\theta + F} = 6 - \theta$. Thus we have $46\theta + F = 36 - 12\theta + \theta^2$. Finally,

$$F = \theta^2 - 58\,\theta + 36.$$

Now that we have solved for *F* as a function of θ we can express E(N) as a function of θ :

$$E(N) = \frac{1}{2} \times 70 + \frac{1}{2} \times 46 - \frac{1}{2} \times [70\theta + F] - \frac{1}{2} \times [46\theta + F]$$

= 58 - 58\theta - \theta^2 + 58\theta - 36 = 22 - \theta^2.

Because e = 1 and the participation constraint imply $E(N) = 22 - \theta^2$, we see that E(N) increases as θ decreases. However, when $\theta = 0$ the manager will set e = 0. And when $\theta^2 > 1/2$ we have $E(N) = 22 - \theta^2 < 21.5$, which is the owner's maximum expected return when e = 0. Therefore, we just need to determine whether we can find $\theta^2 \le 1/2$ such that the manager's EU is higher with e = 1 than with e = 0. To this end we calculate *EU* with e = 0 and $\theta^2 = 1/2$. This will exceed 6, which is *EU* with e = 1 and $F = \theta^2 - 58\theta + 36$. When $\theta^2 = 1/2$ we have $\theta = 0.707$. When $\theta = 0.707$ and $F = \theta^2 - 58\theta + 36$ we have F = 0.5 - 41.012 + 36 = -4.512. Therefore,

$$EU = 4 + 0.5\sqrt{29 \times 0.707 - 4.512} + 0.5\sqrt{22 \times 0.707 - 4.512} = 7.66$$

Therefore, if $y = \theta R + F$, and *F* is set at the value for which EU = 6 when e = 1, the manager will actually choose e = 0 if $\theta^2 = \frac{1}{2}$. If $\theta^2 > \frac{1}{2}$ then the owner's net return will be higher with $\theta = 0$ and e = 0.

5. Agency Theory

What about $\theta^2 < \frac{1}{2}$? Recall that $F = \theta^2 - 58\theta + 36$ when e = 1 and EU = 6. If e = 0 then when R = 29 we have $y = 29\theta + \theta^2 - 58\theta + 36 = \theta^2 - 29\theta + 36$. If e = 0 and R = 22 we have $y = 22\theta + \theta^2 - 58\theta + 36 = \theta^2 - 36\theta + 36$. The quadratic $\theta^2 - 29\theta + 36$ is *minimized* when $\theta = 29/2 = 14.5$ and the quadratic $\theta^2 - 36\theta + 36$ is minimized when $\theta = 36/2 = 18$. The graph of each quadratic is a valley, so with the constraint $\theta^2 \le \frac{1}{2}$ we reach the constrained minimum in each case when $\theta^2 = \frac{1}{2}$. Therefore, if $\theta^2 < \frac{1}{2}$ the manager's EU from e = 0 cannot be any lower than it is when $\theta^2 = \frac{1}{2}$, and we already know that for $\theta^2 = \frac{1}{2}$ the manager's EU is higher with e = 0 than with e = 1 (which results in EU = 6).

We have shown that if the participation constraint is satisfied as a strict equality then the owner's net return is maximized by the contract y = 4, a constant. Is it possible to have the participation constraint satisfied as an inequality, with the manager choosing e = 1 and E(N) > 21.5? Given θ , when we increase the manager's EU by (algebraically) increasing *F* we increase the EU that results from e = 0 faster than we increase the EU that results from e = 1. That is a consequence of the fact that the \sqrt{C} increases faster as *C* is smaller. (You really need calculus to establish that, but try some examples if you're not convinced.) Therefore, the manager's EU will still be higher with e = 0. If we hold *F* constant and reduce θ then that *increases* the manager's incentive to set e = 0 because her pay becomes less sensitive to profit.

The manager will be offered the contract y = 4 for the firm of Example 5.5 and that means that the contract will not be efficient. The optimal contract will give the manager an EU of 6. If the manager could be relied on to set e = 1 and were paid a fixed salary of \$36, then the manager's EU would still be 6, but the return to the owner would be $\frac{1}{2} \times 70 + \frac{1}{2} \times 46 - 36 = 22$, which is higher than 21.5, the owner's net return from the optimal contract.

Our search for the optimal contract in Example 5.5 was confined to the linear family $y = \theta R + F$. But there is a *nonlinear* contract that will give the owner an even higher return than 21.5 and will also induce the manager to supply the efficient level of effort. (The manager's utility will remain at the reservation level $u^0 = 6$. If we increase E(N) we will have made the owner better off without lowering the manager's utility.)

Example 5.6: A nonlinear contract for Example 5.5

Consider the contract

y = 36 if R = 70 or 46, y = 8 if R = 29,

and

$$y = 1$$
 if $R = 22$.

If e = 1 then the manager gets 36 for sure, in which case her utility will be $0 + \sqrt{36} = 6$. The participation constraint is satisfied. But will the manager set e = 1?

If e = 0 then we will have R = 29 with probability $\frac{1}{2}$ and R = 22 with probability $\frac{1}{2}$. In that case the manager's EU will be $4 + \frac{1}{2}\sqrt{8} + \frac{1}{2}\sqrt{1} = 5.914$. Therefore, the manager's EU is higher when e = 1. Then $E(R) = \frac{1}{2} \times 70 + \frac{1}{2} \times 46 = 58$. Therefore, this nonlinear contract yields E(N) = 58 - 36 = 22, which is higher than 21.5, the return to the owner from the best linear contract.

Note that the optimal nonlinear contract requires the owner to know the distribution of *R*. This is an extreme assumption, and for the rest of this section we restrict our attention to contracts of the form $y = \theta R + F$.

There is a trade-off between incentives and insurance. As the degree of risk aversion increases, the amount of insurance afforded to the manager by the optimal contract also increases. Let's return to production function of Example 5.5 to investigate this. We assume that the firm is managed by someone who is less risk averse than the CEO of Example 5.5. The new manager will be given a contract for which $\theta = 1$ and the resulting effort supply is one.

Example 5.7: A less risk-averse manager

The manager's utility function is $U = 21x + \sqrt{y} + y$ and $u^0 = 33.2$, but the problem is otherwise the same as in Example 5.5: If e = 1 then R = 70 with probability $\frac{1}{2}$ and R = 46 with probability $\frac{1}{2}$. If e = 0 then R = 29 with probability $\frac{1}{2}$ and R = 22 with probability $\frac{1}{2}$. (To confirm that this manager is risk averse, consider an asset that pays \$64 with probability $\frac{1}{2}$ and \$36 with probability $\frac{1}{2}$. The expected monetary value is \$50. Given *x*, the EU from the asset is 21x + 57, but the EU from \$50 for sure is 21x + 57.07.)

Suppose that the manager is offered the contract y = R - 30. The manager cannot set e = 0 because that will not result in enough profit for the \$30 payment to the owner. She will set e = 1 and her EU will be

$$0 + \frac{1}{2}\sqrt{70 - 30} + \frac{1}{2}(70 - 30) + \frac{1}{2}\sqrt{46 - 30} + \frac{1}{2}(46 - 30) = 33.2$$

In that case, E(N) = 30.

If y = S, a constant, then the manager will set e = 0. If S = 9 then $EU = 21 + \sqrt{9} + 9 = 33$. Thus, *S* is not quite high enough to satisfy the participation constraint. Therefore, *S* will have to be greater than 9, which means that E(N) will be less than 25.5 - 9 = 16.5. Therefore, $\theta = 1$ results in a higher return to the owner than $\theta = 0$. The optimal contract will induce the manager to set e = 1. (Given e = 0, the owner will maximize E(N) by offering a constant salary. We have seen that a constant salary cannot provide a higher return to the owners than 16.5.)

Even though the optimal contract has the manager choosing e = 1 in Example 5.7, it is not efficient. Under the optimal contract the manager's expected pay is E(R) - 30 = 28 and the owner's expected return is 30. *If* the manager had a constant salary of 28 and were to set e = 1 then E(N) would still be 30 but *EU* would be $\sqrt{28} + 28 = 33.3$, which exceeds her utility under the optimal contract.

5. Agency Theory

5.5 Risk-averse managers and a continuum of effort levels

We continue to assume that only linear contracts are practicable, but we now allow *e* to vary along a continuum.

We simplify the expression of the manager's expected utility, to make the model more amenable to the realistic case of costly monitoring of a risk-averse manager. If *w* is a compensation package let E(w) denote its expected monetary value (EMV). This expected value depends on the effort supplied by the manager, because effort influences profit *R*, which has a bearing on the manager's pay. Now, instead of explicitly writing utility in terms of *x* (the amount of leisure consumption) and *y* (income from *w* and a particular realization of the random variable ξ), we write the manager's expected utility as

$$EU = E(w) - \theta^2 K - \frac{1}{2}e^2$$

E(w) is the EMV of the compensation contract, and that is affected by *e*. A contribution of *e* units of effort by the manager causes her utility to fall on that account, because leisure consumption falls by *e*. We subtract this loss of utility, $1/2 e^2$, directly from the expected pay to determine the manager's net utility. Similarly, the manager's exposure to risk diminishes her utility, and the term $-\theta^2 K$ reflects the utility cost of this risk. The larger is *K* the more risk averse is the manager. (Here *K* is a nonnegative parameter and is constant for a particular manager.) We assume that the contract has the form $y = \theta R + F$, where $R = \beta e + \xi$. As usual, ξ is the random component and has expected value zero. The larger is θ the greater are the swings in the manager's realized pay as the random variable moves up and down. Therefore, the larger is θ the greater is the negative impact of risk on utility. Let's determine the contract that maximizes the owner's net return. (We do not specify the time endowment but simply assume that the solution value of *e* is feasible.)

Example 5.8: A continuum of effort levels

Incentive compatibility is incorporated by maximizing the manager's EU. If $R = 10e + \xi$ and $y = \theta R + F$ then $E(y) = \theta 10e + F$, so

$$EU = 10\,\theta e + F - \theta^2 K - \frac{1}{2}e^2,$$

which is a function V(e) of e. The parameter θ is determined by the owner. The manager responds by selecting e, the only variable that she can control. Therefore, from the manager's standpoint, V(e) is a simple quadratic function of e. From the formula for maximizing a quadratic we get $e = 10\theta$, the effort supply function. Of course, e increases as θ increases. (In calculus terms, V'(e) = $10\theta - e$. Obviously, V'' < 0 at every point, so we want to set V'(e) = 0, and this yields $e = 10\theta$.)

To calculate the owner's profit-maximizing values of *F* and θ we again recognize that profit maximization causes the participation constraint to be satisfied as a strict equality at equilibrium. (If $EU > u^0$ the owner can reduce *F* and that will reduce *EU* by the same amount for every value of *e*. Hence, the manager's choice of *e* will not be affected, but E(N) will increase, and the participation

250 Corporate Governance

constraint will still be satisfied if the reduction in *F* is not too large.) Because $EU = E(y) - \theta^2 K - \frac{1}{2}e^2 = u^0$ we have

$$-E(y) = -\theta^2 K - \frac{1}{2}e^2 - u^0.$$

Because E(R) = 10e, the owner's expected profit is

$$E(R) - E(y) = 10e - \theta^2 K - \frac{1}{2}e^2 - u^0$$

= 10(10 \theta) - \theta^2 K - \frac{1}{2}(10 \theta)^2 - u^0
= 100 \theta - (50 + K)\theta^2 - u^0,

a quadratic function $G(\theta)$ of θ . By the formula for maximizing a quadratic, the function is maximized at

$$\theta^* = \frac{100}{100 + 2K}.$$

(Alternatively, $G'(\theta) = 100 - 2\theta K - 100\theta$ and $G''(\theta) < 0$ at every point. Therefore, we set $G'(\theta) = 100 - 2(50 + K)\theta = 0$.) If the manager is not risk averse then K = 0 and hence $\theta^* = 1$. But for all K > 0 we have $0 < \theta^* < 1$. Because *G* is a quadratic, profit rises as θ increases to θ^* and then falls as θ increases beyond that point (Figure 4.6). The expected return E(R) is higher for θ greater than θ^* because more effort is supplied. But the manager is exposed to greater risk when θ is higher, and the participation constraint forces the owner to compensate the risk-averse manager for the increased risk. The expected return is higher but the manager's pay is higher still when $\theta > \theta^*$. We no longer have maximum incentive ($\theta = 1$) because the owner has to trade off insurance and incentive. Note that θ falls as K increases: The greater the degree of risk aversion the lower is the profit-maximizing value of θ and the more insurance is provided to the manager by the profit-maximizing owner. To determine F we return to the participation constraint, $EU = E(y) - \theta^2 K - \frac{1}{2}e^2 = u^0$ and $E(y) = \theta^* E(R) + F$. Expected gross profit E(R) depends on *e*, which in turn is a function of θ , which equals θ^* . Therefore, with θ^* and u^0 specified we can solve for F.

To summarize, when effort is unobservable the shareholders will have to provide the manger with an incentive to supply effort, and this means that the manager's compensation must be correlated with observed profit. Because profit is influenced by random forces as well as the manager's effort, the incentive effect prevents her from being fully insured against risk, even though the owners bear all the risk in the ideal case of costlessly observable effort. Because the manager is not fully insured, her expected pay must be higher than in the full information case to elicit her participation. Accordingly, the owner's expected return is lower. The manager is not fully insured but does not assume all risk—much of it falls on the shoulders of the owners. To the extent that the manager is insured, the contract diminishes the manager's incentive to maximize the owner's expected profit. The agent is no longer the sole residual claimant. Because she is insured against bad outcomes she will work less assiduously to avoid bad outcomes.





Ken O'Brien was a National Football League quarterback in the 1980s. Because he threw a lot of interceptions early in his career he was given a contract that penalized him every time an opponent caught one of his passes. This incentive clause succeeded in reducing the number of interceptions, but that was probably because he rarely passed (Prendergast, 1999). (This explains why managers are not paid solely in the form of stock options.) In short, there is a trade-off between insurance and incentives.

We saw in Section 5.3 that when the manager is the residual claimant she has an incentive to supply the efficient amount of effort in every dimension. When the manager is risk neutral, it is the owner's interest to modify the managerial incentives to shelter the manager from risk, at least to a degree. That means that it cannot be taken for granted that the man-

ager will do a reasonable job of looking out for shareholder welfare in every dimension.

Sources

The foundations of optimal incentive contracts were laid by Ross (1973), Mirlees (1974, 1976), Stiglitz (1975), and Holmström (1979a). Example 5.8 is from McMillan (1992, pp. 205–8). The progressive piece rate idea is due to Olson (1993).

Links

The following parallel treatments of agency theory are listed in order of increasing difficulty: Chapters 8–10 in McMillan (1992); Sappington (1991); Chapter 1 in Tirole (1988); Laffont and Martimort (2002). Baker (2002) and Baker, Gibbons, and Murphy (2002) consider optimal contract design when *profit* is not verifiable. Their model and results are also presented in Dixit (2004), beginning on page 32.

252 Corporate Governance

Problem set

- **1.** A risk-neutral manager has utility function $U(x, y) = 20 \ln(x + 1) + y$. Units have been chosen so that T = 3. (The individual is endowed with 3 units of *X* and 0 units of *Y*. We could have a positive endowment of *Y* but we assume that it has been netted out of both *F* and u_0 .) The manager's best alternative opportunity is to consume 3 units of leisure and not work. If the manager supplies *e* units of effort then the firm's profit *R* will be $10e + \xi$, where ξ is a random variable with expected value zero. (*R* is profit before deducting the manager's pay.)
 - **A.** Suppose that the owner offers the manager the compensation contract $y = \theta R + F$. Determine the manager's effort supply function. Show that *e* increases when θ increases.
 - **B.** Solve for the contract that maximizes the owner's expected profit.
 - **C.** What is the owner's expected profit, the manager's expected utility, and the effort supplied by the manager under the contract that maximizes the owner's expected profit?
 - **D.** Is the outcome that maximizes the owner's expected profit efficient? Explain.
- **2.** A risk-neutral manager has utility function $U(x, y) = 2\sqrt{x} + y$. The time endowment is *T*. (The manager is endowed with *T* units of *X* and 0 units of *Y*.) The manager's best alternative opportunity provides a level of utility of $u_0 = 2\sqrt{T}$. If the manger supplies *e* units of effort then the firm's profit *R* will be $\beta e + \xi$, where ξ is a random variable with expected value zero and *R* is profit before deducting the manager's pay.
 - **A.** Suppose that the owner offers the manager the compensation contract $y = \theta R + F$. Determine the manager's effort supply function. Show that *e* increases when θ increases.
 - **B.** Solve for the contract that maximizes the owner's expected profit when the manager cannot be monitored.
 - **C.** What is the owner's expected profit, the manager's expected utility, and the effort supplied by the manager under the contract that maximizes the owner's expected profit?
- **3.** Figure 4.7 shows the indifference curve for a risk-neutral manager when the participation constraint is satisfied as an equality. (The two straight lines are parallel.) What is the output per unit of input coefficient β ? Assuming that the manager is offered the contract that maximizes the owner's net return, determine the manager's effort supply, the firm's expected profit *R*, the manager's expected income, and the owner's expected net return, at the effort level of the manager that maximizes the owner's expected net return subject to the participation constraint. What is the form of the contract? Whose income is uncertain, the manager's, the owner's, or both?





4. Figure 4.8 shows the indifference curve for a risk-neutral manager when the participation constraint is satisfied as an equality. Use the diagram to determine the manager's effort supply, the firm's expected profit, the manager's expected income, and the owner's expected net return, at the



Figure 4.8

effort level of the manager that maximizes the owner's expected net return subject to the participation constraint. (AB is parallel to FD.) Now, write the contract represented by the budget line ABCD. Write it as it would appear in the real world with unobservable effort.

- **5.** A risk-neutral manager has utility function U(x, y) = 2000 1690/x + y. Set T = 24. The manager's best alternative opportunity provides a level of utility of $u_0 = 1910$. The firm's profit *R* is $10e + \xi$, where ξ is a random variable with expected value zero and *R* is profit before deducting the manager's pay.
 - **A.** Solve for the contract that maximizes the owner's expected profit, even though the manager cannot be monitored.
 - **B.** Derive the manager's budget line, expressed in terms of *x* and *y*, that the optimal contract induces.
- **6.** A risk-neutral manager has utility function $U(x, y) = 10 \ln(x + 1) + y$. Set T = 2. The manager's best alternative opportunity provides a level of utility of $u_0 = 9.93$. The firm's profit *R* is $5e + \xi$, where ξ is a random variable with expected value zero and *R* is profit before deducting the manager's pay.
 - **A.** Solve for the contract that maximizes the owner's expected profit, even though the manager cannot be monitored.
 - **B.** Now derive the optimal contract by employing the manager's effort supply function.
- **7.** Prove that any solution to the problem "maximize u_1 subject to $u_h \ge u_h^0$ for all $h \ne 1$ " is weakly efficient in any context, where u_h^0 is a constant for each h. Prove that the solution is fully efficient if each utility function is continuous and every individual has a positive amount of some divisible private good. (Divisibility means that one individual can give an arbitrarily small amount of the good to any other person. If the good is private then only the utility of the donor and the recipient is affected.) Now show that with a risk-neutral manager and a risk-neutral owner, the owner's profit-maximizing contract solves "maximize u_1 subject to $u_2 \ge u_2^0$ " where u_1 is the owner's EU and u_2 is the manager's EU.
- **8.** This question features a manager whose utility function is nonlinear in *Y* but there are no random variables affecting production. The manager's utility function is U(x, y) = xy, and the manager's best alternative yields $u^0 = 1$. Profit is R = 4e, where *R* is profit before deducting the manager's pay. T = 2: The manager has an endowment of 2 units of *X* and 0 units of *Y*. Find the contract that maximizes the owner's profit. What is the owner's net return, the manager's consumption of *X* and *Y*, the manager's utility, and the effort supplied by the manager under the contract that maximizes the owner's net return?
- **9.** Again we have a manager whose utility function is nonlinear in *Y* and no randomness in production. The manager's utility function is U(x, y) = xy, and the manager's best alternative yields $u^0 = 1$. Profit is $R = \beta e$, where β is a positive constant and *R* is profit before deducting the manager's pay.

5. Agency Theory

The manager's time endowment is *T*. Find the contract that maximizes the owner's net return. What is the owner's profit, the manager's consumption of *X* and *Y*, the manager's utility, and the effort supplied by the manager under the contract that maximizes the owner's net return?

∂10. This question features a manager whose utility function is nonlinear in *Y*, but there are no random variables affecting production. Let U(x, y) represent the manager's utility function. The manager's best alternative yields u^0 . Profit is $R = \beta e$, where β is a positive constant and *R* is profit before deducting the manager's pay. The manager's time endowment is *T*. Initially, the manager has 0 units of *Y*. Show that the contract that maximizes the owner's profit has the form y = R + F, where *F* is fixed, independently of the profit *R*. *Hint:* Use the implicit function theorem and the participation constraint to solve for dy/dx in terms of the partial derivatives of *U*. Then compare the first-order condition from the owner's maximization problem to the solution of

maximize U(x, y) subject to $p_1x + p_2y = C$.

- **11.** Solve for the contract that maximizes the owner's net return in the model of Section 5.5 with $R = 4e + \xi$, K = 8, and $u^0 = 5$. Calculate *F* as well as θ . What is the owner's expected return?
- **12.** Solve for the contract that maximizes the owner's net return in the model of Section 5.5 when the manager's utility function is $EU = 24x \frac{1}{2}x^2 + E(y) 192\theta^2$ with $R = 16e + \xi$ and $u^0 = 320$. Calculate *F* as well as θ . What is the owner's expected return?
- **13.** Section 5.5 represents the agent's utility indirectly, as y C(e), where *C* is the cost of effort to the agent, with positive marginal cost that increases as *e* increases. Show that this leads to the same behavior as when we take U = B(x) + y, with positive marginal benefit B(x) but with marginal benefit decreasing as *x* increases. Of course x = T e. In particular, show that in both cases, utility declines as effort increases, and that the rate of decline is higher when effort is greater.
- **14.** The manager of a firm has the utility function $U = 50x x^2 + y$. The manager will cease to work for the firm if the manager's utility falls below 624. Set T = 24 and the manager's initial wealth is zero. The profit *R* realized by the owner, before deducting the manager's pay, is given by $R = 30e + \xi$, where *e* is effort supplied by the manager and ξ is a random variable with an expected value of zero. The owner cannot enforce a contract that mandates a specific input of effort. Show that the owner's net return is maximized if the owner offers the manager a contract that requires a payment of \$196 from the manager to the owner with the manager keeping any additional profit realized by the firm.

5

Hidden Characteristics

1.	Pric	e Discrimination	257 259
2.	Two	-Person Exchange	259
	2.1	Dominant strategy equilibrium	260
	∂ 2.2	Nash equilibrium Problem set	266 268
∂3 .	The	Used-Car Market	269 272
4.	Cree	dit Rationing	272
	4.1	The borrower's point of view	274
	4.2	The lender's point of view Problem set	277 279
∂ 5.	Bun	dling and Product Quality	280
	5.1	The Model	282
	∂ 5.2	Full information equilibrium	283
	∂ 5.3	Asymmetric information equilibrium Problem set	285 289
6.	Job-	Market Signaling	290
	6.1	To make a long story short	291
	6.2	A general model	292
	6.3	When education is productive Problem set	297 302
7.	Con	npetitive Insurance Markets	303
	7.1	The model	303
	7.2	The number of contracts in equilibrium	307
	∂ 7.3	Full information equilibrium	309
	∂ 7.4	Asymmetric information equilibrium Problem set	312 322

1. Price Discrimination

This and the remaining chapters investigate hidden characteristic problems, from voting to used-car markets to kidney exchanges. We see that market forces have spawned contracts and other devices that induce agents to reveal their hidden characteristics. This does not mean that the equilibrium outcome is efficient in each case, however. There are incentive schemes that *do* induce truthful revelation of the hidden information while at the same time bringing the system close to efficiency—the Vickrey auction of Chapter 6 for instance.

Markets are wonderfully creative in circumventing hidden information problems. Warranties on consumer durables provide a nice example of the market system generating its own solution to a hidden characteristic problem. The producer of a shoddy appliance cannot afford to offer a substantial warranty.

It can be in society's interest to have the hidden information remain hidden. It is often essential for communication about financial transactions to be encoded so that eavesdroppers cannot profit from the information. Electronic messages are encoded using an asymmetric form of encryption: The recipient *R* of the message publishes the key to encoding the text that R will receive. This key is the product of two very large prime numbers p and q. But only the product is published. To decode the message it is necessary to know both p and q, and only Rknows these prime factors. If they are sufficiently large, it will be well beyond the ability of even a network of huge computers to determine them in anyone's lifetime, even though the product is known.

The point of producing a low-quality item is to get more profit by keeping costs down, but if many appliances are being returned for refund or repair then costs will be high, not low. A producer who deliberately sets out to profit by misleading consumers about the quality of the product will not be able to offer the same kind of warranty as the producer of a high-quality product. The producer of the highquality item is signaling high quality to the consumer by offering a substantial warranty. Reputable manufacturers often make good on a warranty even after it has expired, as long as the appliance is returned a month or less after the expiration date.

Although not always delivering an efficient outcome, the market system often goes a long way toward eliciting the hidden information. The next section begins with a standard example of the hidden characteristic phenomenon.

Links

See Mann and Wissink (1988, 1990a, 1990b) for a more thorough discussion of warranties. For more on the technique of asymmetric encryption see Singh (1999), which is a superb history and analysis of coding and decoding from ancient Egypt to the present.

1 PRICE DISCRIMINATION

Suppose that a firm's consumers can be divided into two categories, highdemand-elasticity and low-demand-elasticity types. In that case the firm's profit can be increased by charging a higher price to the latter group. Because of this, the low-elasticity customers cannot be expected to voluntarily disclose their (elasticity) characteristic. Consequently, suppliers will endeavor to find something that is both observable and correlated with the hidden demand characteristics.

Example 1.1: Haircuts

We can simplify the analysis by selecting a commodity for which no unit after the first has utility. A haircut, for instance. To make our job really easy, suppose that everyone wants a haircut every thirty days but not more frequently. The community consists of *A* types who would be willing to pay \$20 for a haircut, but no more, and *B* types who would pay a maximum of \$9 for a haircut. There are *n* of each. There is only one barber in town, and the opportunity cost of the barber's time is \$7 per haircut. If the barber has no way of distinguishing *A* types from *B*s then the barber must charge one price *P*. If *P* = 20 then the barber's profit per haircut is 20 - 7, and because only the *A* types will come to the shop, the profit will be 13n. (The *B* types will get a friend or relative to do the job, or drive to another town.) If *P* = 9 then the profit per haircut would be 9 - 7 = 2. Everyone would be a customer at *P* = 9, resulting in a total profit of 4*n*. Clearly, if the barber can't distinguish *A* types from *B* types then the profit-maximizing price is *P* = 20.

Suppose, however, that all the *B* types are sixty-five years of age or older. If the barber charges \$9 per haircut to anyone older than sixty-four and \$20 to everyone else the profit will be 2n + 13n, which is 15% higher than the profit of 13n without price discrimination. (It is essential to our story that a *B* type is not able to buy a \$9 haircut and sell it to an *A* type for, say, \$15.)

Consider plane travel: On average, business travelers have a lower demand elasticity than recreation travelers. The former pass their travel expenses on to their companies, who in turn pass on part of the cost to taxpayers. And business trips often have an urgency that nonbusiness travel seldom does. That makes the business traveler much less responsive to a price increase. Vacationers have lots of close substitutes, which makes their demand much more price sensitive. All of this results in a relatively low elasticity of demand for plane tickets by business travelers. By charging a higher fare for travelers who don't stay over at least one Saturday night an airline company can force most business travelers to pay the higher fare. Most business trips do not extend through Saturday. A Saturday stayover is very costly to a business traveler because of the need to be back in the office as soon as the purpose of the trip has been accomplished and the desire to spend the weekend with the family.

Xerox Corporation introduced the first push-button electrostatic copying machine in 1960 and for many years the company faced very little competition. Price discriminating profit maximization implies a higher charge for machines purchased by firms that intend to use them intensively. But these firms would not willingly admit that they are high-intensity users (and the machines can be resold anyway). The chief rival to Xerox in the 1960s was Electrofax, which produced a copier that required a special coated paper. Initially, Electrofax held a monopoly on the sale of the special paper. By charging a price for the paper that was significantly above marginal cost, the company in effect charged a higher price for copying machines purchased by high-intensity users. A similar

2. Two-Person Exchange

principle applies to the charges for Polaroid film during the period when Polaroid Corporation had a monopoly on the production of self-developing film and the complementary camera. And again in the case of the early IBM computers, the punch cards sold by IBM and used to enter input provided a way for IBM to meter the use of IBM machines. Initially, IBM had a monopoly on the sale of punch cards and they were priced above marginal cost to allow IBM to extract more revenue from the high-intensity users of its computers.

The Xerox copier did not require special paper, so the Xerox Corporation solved its hidden characteristic problem by refusing to sell the copiers; the machines had to be leased from Xerox. The rental fee was based on the number of copies made, so Xerox was able to meter its customers' usage and thus force high-intensity users to pay more for the use of the copier.

There is a tension between price discrimination and the extraction of consumer surplus. If all consumers had identical demand functions for the services generated by the machines and their variable input (paper, film, punch cards) then the monopolist would want to price the variable input at marginal cost to induce the buyer to use the equipment more intensively, which in turn allows a higher price to be charged for the equipment because of the larger consumer surplus. (One can show that profit is maximized when the price of the variable input is set at marginal cost and the price of the machine is set equal to the resulting consumer surplus. In this case there is only one demand curve to consider—the demand for the services of the machine.)

Source

The copying machine details are from Phlips (1981).

Links

To see why price discrimination can emerge in a competitive environment see Dana (1998) and Varian (2000).

Problem set

 Assume that all consumers are of the same type—they all have the same demand function. Prove that profit is maximized when the price of the variable input is set at marginal cost and the price of the machine required to turn this input into the desired product is set equal to the resulting consumer surplus.

2 TWO-PERSON EXCHANGE

A brother *B* and a sister *S* have jointly inherited a house. Each has a family, so they are not willing to share the house, which they suspect is worth more to *B* than to *S*. If it were worth, say, \$100,000 to *B* and only \$20,000 to *S* it doesn't seem fair for *B* to pay his sister only \$25,000 for her share. Moreover, *S* would not have an incentive to reveal her reservation value truthfully if the price will be 125% of the seller's value. However, the arbitration rule that has *B* paying half of what the house is worth to him gives *B* a strong incentive to understate his reservation

260 Hidden Characteristics

value. But if the parties do not reveal their true reservation values we have no guarantee that the house will be used by the family that values it most—that is, gets the most benefit from it. The same problem arises when two business partners *B* and *S* have decided that they cannot continue working together. One of them will take over the business by buying the other's share. They have different abilities and different expectations about the future profitability of the enterprise. How should they dissolve the partnership?

2.1 Dominant strategy equilibrium

Formally, there is a single asset, owned by individual *S*, and a potential buyer *B*. The asset is worth a minimum of V_S to *S* and a maximum of V_B to *B*. These are the respective reservation values. We seek a recipe for deciding when the asset should be transferred from *S* to *B* and at what price. An *exchange mechanism* is a decision rule under which each party reports its reservation value and then determines whether *B* gets the asset. The exchange mechanism also specifies a price paid by the buyer and an amount of money received by the seller.

DEFINITION: Exchange mechanism

The buyer's reservation value V_B is the maximum that B would be willing to pay for the asset. The seller's reservation value V_S is the minimum that S would accept to relinquish the asset. An exchange mechanism requires them both to report their reservation values and determines when the asset changes hands, as a function of the reservation value R_B reported by B and the reservation value R_S reported by S. If it is exchanged then the mechanism specifies the amount $P(R_B, R_S)$ paid by the buyer and the amount $Q(R_B, R_S)$ received by the seller.

We want to employ a mechanism that has three properties: *incentive compatibility*, which means that truthful revelation is a dominant strategy for each party; *asset efficiency*, which means that *B* gets the asset if it is worth more to *B* than to *S*, otherwise *S* keeps the asset; and the *participation constraint*, which means that neither *B* nor *S* winds up with lower utility than he or she started with.

DEFINITION: Incentive compatibility, asset efficiency, and the participation constraint

An exchange mechanism is incentive compatible if for each pair (V_B , V_S) (i) no reported value R_B gives the buyer B a higher payoff than reporting the true value V_B and (ii) no reported value R_S gives the seller S a higher payoff than reporting the true value V_S . The mechanism is asset efficient if B gets the asset when $V_B > V_S$ and S keeps the asset if $V_B \le V_S$. The mechanism satisfies the participation constraint if neither S nor B pays anything when no trade takes place.

2. Two-Person Exchange



We prove that the *only* mechanism with these three properties requires the buyer to make a payment equal to the seller's reservation value, and the seller to receive an amount equal to the buyer's reservation value. When the seller's reservation value is at least as high as the buyer's then the asset stays with the seller, and no one pays any money or receives any money. This is called the *Groves bargaining mechanism* (GBM).

DEFINITION: Groves bargaining mechanism (GBM)

When $V_B > V_S$ the asset is transferred from the seller to the buyer, with $P(V_B, V_S) = V_S$ and $Q(V_B, V_S) = V_B$. When $V_B \le V_S$ the seller keeps the asset and $P(V_B, V_S) = 0 = Q(V_B, V_S)$.

Notice that the definition anticipates our proof that both agents will report their true reservation values: V_B for the buyer and V_S for the seller.

Before proving that the GBM is the only one that satisfies asset efficiency, the participation constraint, and incentive compatibility, we demonstrate that the GBM actually does have our three properties. GBM satisfies asset efficiency and the participation constraint by definition. Now, consider incentive compatibility. We begin with the case $V_B < V_S$ (Figure 5.1). Under truthful revelation there is no trade and no payment by B. Would it ever be to B's advantage to misrepresent B's reservation value? What would happen if B reported a reservation value R_B in region H, where R_B is higher than V_S ? B would get the asset and would be required to pay V_S , which exceeds the true worth V_B of the asset to B, resulting in a loss to B. Truthful revelation would have resulted in no gain or loss to B, and hence would be preferred by B to reporting R_B in region H. However, suppose Bwere to report a reservation value R_B in region L or M where $R_B < V_S$. This yields the same outcome as V_B because both are below V_S . Therefore, when $V_B < V_S$ the buyer B can never profit from misrepresenting the true reservation value, but can lose by doing so. In other words, truthful revelation is a best response by *B* to any V_S for which $V_B < V_S$.

We continue to assume that $V_B < V_S$, as in Figure 5.1, and now consider the incentive of the seller *S*. Suppose that the seller reports a reservation value R_S in region *L* below V_B . In that case trade will take place, and the seller will receive V_B , which is less than the seller's true value V_S , resulting in a loss of $V_S - V_B$ to the seller. But had the seller revealed her reservation value V_S truthfully there would have been no trade and hence no loss. If the seller were to report a value R_S in the region *M* or *H* above V_B then the buyer's reservation value will still be below the reported reservation value of the seller, and there will be no trade. That's the same outcome that results from truthful revelation. Hence, when V_B



Figure 5.2

is less than V_S the GBM induces truthful revelation by both the buyer and the seller. That is, neither can benefit from deviating from the truth, but either can be hurt by doing so.

Now, we consider the more interesting case of potential gains from trade: $V_B > V_S$ (Figure 5.2). Can the buyer ever gain by misrepresenting *B*'s reservation value? What would happen if *B* reported a reservation value R_B in region *L* where R_B is below V_S ? *B* would not get the asset and would forgo the profit of $V_B - V_S$ that would have resulted from truthful revelation. If *B* were to report any reservation value R_B in region *M* or *H* then we would still have $R_B > V_S$, and *B* would acquire the asset, just as in the case of truthful revelation. Moreover, the payment that *B* would have to make is the same for any R_B in *M* or *H*, because that payment equals V_S , which is independent of R_B , for any $R_B > V_S$. Therefore, the buyer cannot profit by deviating from truthful revelation when $V_B > V_S$ and would be hurt by any deviation in region *L*.

On one hand, suppose that $V_B > V_S$ and the *seller* reports a reservation value R_S in region H above V_B . Then no trade will take place, in which case the seller forfeits the profit of $V_B - V_S$ that would have accompanied truthful revelation. On the other hand, if the seller were to report a value R_S in the region L or M below V_B then trade will still take place and the seller will receive the same profit $V_S - V_B$ that results from truthful revelation. That's because when R_S is less than V_B , the seller receives a payment V_B that is independent of R_S . Hence, when V_B is greater than V_S the GBM induces truthful revelation by both the buyer and the seller—truthful revelation is a dominant strategy for each.

Proving that the GBM is the only mechanism with our three properties is a little more demanding.

Uniqueness of the GBM

The GBM is the only incentive-compatible mechanism that satisfies asset efficiency and the participation constraint.

To prove this we have to begin by examining the price and revenue functions of a mechanism that we know almost nothing about. All we know is that it satisfies asset efficiency, incentive compatibility, and the participation constraint. We show that the three properties imply that the price and revenue functions are precisely those of the GBM. We let $P(V_B, V_S)$ denote the price that our mystery mechanism requires the buyer to pay when the asset changes hands, and we let $Q(V_B, V_S)$ denote the amount received by the seller. If we can show that $P(V_B, V_S) = V_S$ and $Q(V_B, V_S) = V_B$ then we have proved that this mechanism must be the GBM.

2. Two-Person Exchange

Before presenting the proof, we give an informal demonstration that our three properties force the price and revenue functions to be identical to the ones specified by GBM. Suppose that asset is worth more to the buyer than the seller. Consider two different possible buyer reservation values V_B^1 and V_B^2 , both of which are greater than V_S . Asset efficiency implies that when *B* reports V_B^1 then *B* gets the asset for a price of $P(V_B^1, V_S)$, and when *B* reports V_B^2 then *B* gets the asset and pays $P(V_B^2, V_S)$. If $P(V_B^1, V_S) > P(V_B^2, V_S)$ then when *B*'s true reservation value is V_B^1 the buyer is better off reporting V_B^2 and paying the smaller amount $P(V_B^2, V_S)$. Similarly, $P(V_B^1, V_S) < P(V_B^2, V_S)$ leads to a violation of incentive compatibility. Therefore, $P(V_B^1, V_S) = P(V_B^2, V_S)$ must hold for any two reservation values V_B^1 and V_B^2 that exceed V_S .

Why does the price have to equal V_S ? The price can't be below V_S if $V_B > V_S$. Otherwise, there would be a situation in which the asset is worth more to the seller than the buyer, but the buyer could misrepresent and report V_B and get the asset for a price below V_S and below B's true reservation value. Therefore, $P(V_B, V_S) \ge V_S$ if $V_B > V_S$. However, if $P(V_B, V_S) > V_S$ we can replace V_B with any V_B^1 above V_S . The price will not change by the argument of the previous paragraph. Now bring V_B^1 closer and closer to V_S , but keep it above V_S . The price will still not change, but it can't be above V_B^1 or else the participation constraint will be violated. And it can't exceed V_S . This rules out every possibility except $P(V_B, V_S) = V_S$. Similarly, one can show that $Q(V_B, V_S^1)$ must equal $Q(V_B, V_S^2)$ for any two seller reservation values V_S^1 and V_S^2 that are below V_B and then use that fact to establish $Q(V_B, V_S) = V_B$. Now let's turn to the formal proof.

We want to prove that if the participation constraint is satisfied then incentive compatibility and asset efficiency imply $P(V_B, V_S) = V_S$ and $Q(V_B, V_S) = V_B$. We begin by showing that the price paid by the buyer can never exceed the buyer's reported reservation value. That is, $P(V_B, V_S) \leq V_B$ must hold for every combination of V_B and V_S such that $V_B > V_S$. If we actually had $P(V_B, V_S) > V_B$ then truthful revelation would not be a dominant strategy for the buyer, who could report a reservation value of zero, resulting in no trade and no payment by the buyer. But when $P(V_B, V_S) > V_B$, truthful revelation would result in the buyer acquiring the asset worth V_B and having to pay a larger amount $P(V_B, V_S)$ for it. Therefore, incentive compatibility implies $P(V_B, V_S) \leq V_B$. Similarly, if $Q(V_B, V_S) < V_S < V_B$ then the asset would change hands under truthful revelation (because $V_B > V_S$), but the seller would be paid less than the minimum V_S that S would be willing to accept to part with the asset, resulting in a loss of $V_S - Q(V_B, V_S)$. The seller could avoid that loss by reporting a reservation value of $2 \times V_B$, in which case no exchange would take place. Therefore, if truthful revelation is a dominant strategy for the seller for all possible values of V_S we must have $Q(V_B, V_S) \ge V_S$ for arbitrary V_S and V_B such that $V_S < V_B$.

Now, suppose that $P(V_B, V_S) < V_S$ as in Figure 5.3. In that case when the buyer's *true* reservation value is T_B between $P(V_B, V_S)$ and V_S then asset efficiency requires that no trade take place, leaving no profit for the buyer with reservation value T_B . But if that buyer were to report a reservation value of V_B then trade

$$P(V_B, V_S)$$
 T_B V_S V_B

Figure 5.3

would take place and the buyer would pay $P(V_B, V_S)$, which is less than T_B , leaving a positive profit of $T_B - P(V_B, V_S)$. Therefore, incentive compatibility requires $P(V_B, V_S) \ge V_S$ for every combination of V_B and V_S such that $V_B > V_S$.

To prove that $P(V_B, V_S)$ cannot actually be larger than V_S , we suppose the contrary and show that one or more of the required properties must be violated as a result. Suppose, then, that $P(V_B, V_S) > V_S$ as in Figure 5.4. When the buyer's true reservation value is V_B , B can report a value R_B between V_S and $P(V_B, V_S)$. Agent B will still get the asset because $R_B > V_S$ but will pay some price $P(R_B, V_S) \leq R_B$. This price $P(R_B, V_S)$ must be less than $P(V_B, V_S)$ because $P(R_B, V_S)$ cannot exceed R_B , as we have already discovered. The resulting profit to the buyer will be at least $V_B - R_B$, which is greater than the profit of $V_B - P(V_B, V_S)$ that results from truthful revelation. Therefore, incentive compatibility requires $P(V_B, V_S) \leq V_S$. Because we have also established $P(V_B, V_S) \geq V_S$, we have proved that $P(V_B, V_S) = V_S$ for all reservation values such that $V_B > V_S$. It remains to prove that $Q(V_B, V_S) = V_B$ when $V_B > V_S$.

Suppose that $Q(V_B, V_S) > V_B$ as in Figure 5.5. In that case when the seller's *true* reservation value is T_S between V_B and $Q(V_B, V_S)$ then asset efficiency requires that no trade take place. But if the seller were to misrepresent and report V_S , then trade would take place and the seller would receive $Q(V_B, V_S)$, which is more than T_S , yielding a positive profit for the seller of $Q(V_B, V_S) - T_S$. Therefore, incentive compatibility requires $Q(V_B, V_S) \le V_B$ for every combination of V_B and V_S such that $V_B > V_S$.

Finally, we show that $Q(V_B, V_S) < V_B$ cannot hold for any combination of V_B and V_S such that $V_B > V_S$ as shown in Figure 5.6. If we did have $Q(V_B, V_S) < V_B$, then when the seller's true reservation value is V_S S can report a value R_S between $Q(V_B, V_S)$ and V_B . There will still be a sale, but S will receive at least R_S , which is higher than $Q(V_B, V_S)$. (Recall that our three conditions require that the seller receives at least as much as his or her own reservation value.) Therefore, incentive compatibility requires $Q(V_B, V_S) \ge V_B$. But we have already ruled out $Q(V_B, V_S) > V_B$, so we have proved that $Q(V_B, V_S) = V_B$ holds for all reservation values such that $V_B > V_S$. Because we know that $P(V_B, V_S) = V_S$ also holds we have proved that the mechanism must be the GBM if it satisfies asset efficiency, incentive compatibility, and the participation constraint.

Applications

In the case of two heirs splitting an indivisible asset (such as a house) we begin by assuming that heir *B* will buy out the other heir, individual *S*. Then V_B is

$$V_S$$
 R_B $P(V_B, V_S)$ V_B

Figure 5.4

2. Two-Person Exchange

$$V_S$$
 V_B T_S $Q(V_B, V_S)$

Figure 5.5

the value to *B* of *S*'s share—specifically, the difference between the value to *B* of outright ownership of the asset and the value if it is shared with *S*. Then V_S is the value of the asset to *S* when ownership is shared. Assuming that the GBM is employed, if $V_B > V_S$ then *B* gets the asset and pays V_S with person *S* receiving V_B .

If players *S* and *B* are business partners dissolving their firm, with partner *B* buying out partner *S*—perhaps because the business would be more profitable under the stewardship of *B*—then V_B is the difference between the value to *B* of outright ownership of the firm and the value of continuing the partnership, and V_S is the value to *S* continuing the partnership.

The budget imbalance problem

When V_B is larger than V_S the GBM requires *B* to pay V_S for the asset but *S* receives the larger amount V_B . Is it possible that a third party would supply the difference $V_B - V_S$? Instead of using a mechanism such as the GBM, the two parties could battle each other in court and perhaps dissipate 50% of the value of the asset that is in dispute. Because that often happens, they should be willing to pay a modest fee to an arbiter—if they could be sure that the arbiter would settle the matter efficiently and fairly. That would certainly happen if the arbiter used the GBM. The arbiter could charge a fee that yielded a positive annual profit, although there would be a loss on some cases that could be covered by a profit from other cases. The fee could be set high enough so that the arbiter could supply the difference between $Q(V_B, V_S)$ and $P(V_B, V_S)$ in each case. The two parties would be willing to pay the fee because the alternative would be a costly legal battle.

This scheme has a fatal flaw: The two parties *B* and *S* would have a strong incentive to collude and have the buyer submit a very high (untruthful) V_B so that they could split $Q(V_B, V_S) - P(V_B, V_S)$.

Remark on the participation constraint

Suppose that the two parties did each pay a fee *F* to an arbiter who then applied the GBM. That would be equivalent to the following mechanism: $P(V_B, V_S) = V_S + F$ and $Q(V_B, V_S) = V_B - F$ when $V_B > V_S$, with $P(V_B, V_S) = F$ and $Q(V_B, V_S) = -F$ when $V_B \le V_S$. In words, each pays a fee *F*, whatever happens, and then the GBM is applied. Truthful revelation *is* a dominant strategy for this mechanism: Consider the buyer. Given V_S , the *difference* between $P(V_B^1, V_S)$ and $P(V_B^2, V_S)$ for any two reported reservation values V_B^1 and V_B^2 above V_S is

$$V_S \qquad Q(V_B, V_S) \quad R_S \qquad V_B$$

Figure 5.6

266 Hidden Characteristics

the same for this mechanism as for the GBM. We've just added a constant to each of the GBM prices. Similarly, given V_B , the difference between $Q(V_B, V_S^1)$ and $Q(V_B, V_S^2)$ is the same for this mechanism as for the GBM, for any V_S^1 and V_S^2 below V_B . Of course this new mechanism does not satisfy the participation constraint but, problems of collusion aside, one can imagine individuals being willing to participate anyway.

$\partial 2.2$ Nash equilibrium

We now constrain the exchange mechanism by requiring the payment received by the seller to equal the amount paid by the buyer. Also, we now suppose that the buyer is uncertain about the value of the asset. We also relax the incentive compatibility requirement and merely require the existence of a Nash equilibrium that is efficient. Because of the buyer's uncertainty, *B*'s payoffs will be evaluated in terms of expected utility (EU).

A risk-neutral buyer and a risk-neutral seller must agree on the price at which the seller is to deliver a single asset to the buyer. The two agree that the asset is worth more to the buyer than the seller, but the buyer does not know the actual value of the asset. The value is known to the seller, though. We have a hidden characteristic problem. Even though the seller knows the value of the asset to himself before negotiation takes place, that value is a random variable from the buyer's perspective. The asset may be a firm that the seller owns. The buyer is a better manager than the seller, so the firm would generate more profit if it were managed by the buyer. But the actual profit depends upon a technological innovation for which the seller is seeking a patent, and the seller knows much more about the discovery than the buyer.

To be specific, let v denote the value of the asset to the seller. The random variable v is assumed to be uniformly distributed on the interval 0 to 1, inclusive. (This distribution is introduced in Section 6.5 of Chapter 2.) The value to the buyer is assumed to be 1.5v. Although both people know that the value to the buyer is 50% higher than its value to the seller, the buyer does not know v itself until the sale is complete and the asset is in her hands. The seller knows v before negotiations take place, though.

Because $\int x \, dx = 0.5x^2$, the expected value of v is

$$\int_0^1 v \, dv = \frac{1}{2}(1^2 - 0) = \frac{1}{2}$$

Suppose that the following simple bargaining scheme is adopted. The buyer submits a bid for the asset, which the seller either accepts or rejects. If the buyer's bid of *b* is accepted by the seller then the asset changes hands at that price. The seller's payoff is the selling price minus the seller's value *v*, and the buyer's payoff is the difference between the value to the buyer and the price paid. However, the buyer's bid has to be determined before the asset changes hands and before the value is known to her. Therefore, the payoff used by the buyer to determine her optimal bid is the *expected value* of the difference between the value to her and the price that she pays, *conditional on acceptance of the offer*

2. Two-Person Exchange

by the seller. If the buyer's offer is rejected there is no exchange and no further negotiation.

DEFINITION: A take-it-or-leave-it offer by the buyer

The buyer bids *b*. If the seller accepts, his payoff is b - v, and the buyer's payoff is the expected value of 1.5v - b, *conditional on acceptance*. If the offer is rejected there is no further negotiation, in which case the seller's payoff is zero and the buyer's is zero.

The seller will accept *b* only if $v \le b$. If v > b then his payoff is higher if he keeps the asset. The buyer knows this, although she doesn't know *v* itself. Let's determine the expected value to the buyer resulting from a bid of *b*. If v > b the buyer's payoff is 0. If $v \le b$ the asset will change hands and the buyer's profit is 1.5v - b. Because $\int 1.5x \, dx = 0.75x^2$, for any bid *b* the *expected* payoff to the buyer is

$$\int_0^b (1.5v - b)dv = 0.75b^2 - b^2 = -0.25b^2$$

Note that we have integrated over the subinterval [0, b]. That's because the seller keeps the asset when v > b, and hence the buyer's payoff is 0.

Whatever the buyer bids, the expect value will be negative. Even though the asset has more value to the buyer than the seller, and both know that, there is no price that the buyer would be willing to pay that the seller would accept. Because the asset stays with the individual who values it least, the outcome is inefficient.

No trade will ever take place if the buyer makes a take-it-or-leave-it offer.

The inefficiency of this mechanism is a direct consequence of asymmetric information. A bid of *b* is accepted by the seller only when *v* is below *b*, which means that the value to the buyer is 1.5b at most. However, the buyer pays *b* dollars for certain when an offer of *b* is accepted. She pays \$b for something worth less on average—a losing proposition. Therefore, the buyer will bid 0 and the sale will never take place even though both parties are aware that the asset is worth 50% more to the buyer than the seller. This is surely inefficient. If *both* parties knew *v* they could split the profit—that is, trade at the price 1.25v.

Consider another bargaining mechanism: The seller makes an offer *s*, which the buyer can either accept or reject. If the buyer accepts the seller's bid *s* then the asset changes hands at that price, in which case the seller's payoff is s - v and the buyer's payoff is the expected difference between the value to the buyer and the price paid. If the offer is rejected there is no exchange and no further negotiation.

DEFINITION: A take-it-or-leave-it offer by the seller

The seller bids *s*, and if the buyer accepts she pays *s* to the seller in return for the asset. In that case the seller's payoff is s - v and the buyer's payoff is the expected value of 1.5v - s. If the offer is rejected there is no further negotiation, in which case the seller's payoff is 0 and the buyer's is 0.

Let's work out the equilibrium value of *s*. The seller knows *v* so he would not set *s* below *v*. (The seller's payoff is higher when he keeps the asset than when he gives it up for s < v.) We know that $s \ge v$, and so does the buyer, although the buyer does not know *v*. If v < s < 1.5v then both parties gain by a sale at *s* dollars because the asset is worth 1.5v to the buyer. The buyer does not know *v* but she knows that the seller knows that the asset is worth 1.5v to the buyer. The buyer does not know *v* but she knows that the seller knows that the asset is worth 1.5v to the buyer. The buyer does not know *v* but she knows that the seller knows that the seller will set *s* between *v* and 1.5v. But this cannot be an equilibrium strategy. The seller has an incentive to charge a high *s* even when *v* is very low, relying on the buyer to *assume* that v > s/1.5.

The expected value of the asset to the buyer is

$$\int_0^1 1.5v \, dv = 0.75 \times (1^2 - 0^2) = 0.75.$$

Therefore, the expected payoff to the buyer when the seller charges *s* is simply 0.75 - s.

There is no Nash equilibrium: The seller will set *s* above 1 and the buyer will only accept *s* below 0.75.

No trade will ever take place if the seller makes a take-it-or-leave-it offer.

Is there any bargaining mechanism that will permit the realization of the gains from trade, which both parties know to be positive for each? No! The most favorable scheme from the buyer's standpoint is the one where she makes a final offer that the seller has no authority to modify and can only accept or reject it. As we have seen, even that fails to leave the buyer with a positive expected profit.

Sources

Section 2.1 is based on Danilov and Sotskov (2002). Section 2.2 is based on Samuelson (1984, 1985).

Links

See Farrell (1987) and Maskin (1994) on laissez-faire and efficiency.

Problem set

1. Section 2.1 doesn't acknowledge the possibility of a tie, in the sense that $V_B = V_S$. Show that truthful revelation is a dominant strategy for the GBM, whatever the tie-breaking rule—the asset goes to *B* in the case of a tie, or the

∂3. The Used-Car Market

asset goes to *S* in case of a tie, or a coin is flipped, and so forth—assuming that P(V, V) = V = Q(V, V) for all *V*.

- **2.** Demonstrate how profitable it is for *B* and *S* to collude if they pay a fee *F* to an arbiter, who collects V_S from the buyer and pays V_B to the seller.
- **3.** Prove the claim of the last paragraph of Section 2.1: Truthful revelation is a dominant strategy for the mechanism that requires each agent to pay a fee *F* and then applies the GBM.
- **4.** For the model of Section 2.1, prove that truthful revelation is a dominant strategy for any mechanism for which V_S is the *difference* between what the buyer pays when the asset changes hands and when it doesn't, and V_B is the *difference* between what the seller receives when the asset changes hands and when it doesn't.
- **5.** Show that the GBM (of Section 2.1) is not the pivotal mechanism (of Section 2 of Chapter 8), although it is a Groves mechanism (of Section 3 of Chapter 8).

∂3 THE USED-CAR MARKET

The used-car market is one of the hidden characteristic problems for which the market system has not developed a completely satisfactory solution. Many used cars on the market are "lemons"—cars that frequently require expensive repairs. Individuals who purchase new cars often try to sell them when they are discovered to be lemons, and hence the used-car market contains a disproportionately high number of low-quality cars. This depresses the price of used cars because the buyer can't tell which are lemons. There is asymmetric information. Many car owners who would otherwise put their good cars up for sale find that the selling price of their cars is too low. They are better off continuing to drive their high-quality automobiles than selling them for a low price that reflects the low *average* quality in the used-car market. This further lowers the average quality of used cars at equilibrium, resulting in an even lower equilibrium price. And so on. In terms of the economist's jargon, many car owners find that the *reservation value* of their cars is higher than the price that the car will fetch on the market.

DEFINITION: Reservation value

The car owner's reservation value is the minimum that he of she would be willing to accept to part with the car. The buyer's reservation value is the maximum that he or she would be willing to pay.

Sellers' utility will increase if and only if they sell their cars for more than their reservation values. Buyers' utility will increase if and only if they buy their cars for less than their reservation values. These observations follow from the definition of "willing." (Read the definition of *reservation value* again.)

270 Hidden Characteristics

The used-car market exhibits a degree of market failure: There are owners of high-quality automobiles who would be willing to sell their cars at prices that buyers would be prepared to pay *if* they could be certain of the quality. However, one cannot distinguish high-quality cars from lemons *before* purchasing, so the price of high-quality used cars reflects the large fraction of lemons in the market. Consequently, there are buyers and sellers of the high-quality cars who are not able to strike a deal. The highest price that the seller could obtain is often below the seller's reservation price. The outcome is not efficient. To drive this point home—pun intended—consider what happens the day after you accept delivery of your new car. The car's value on the used-car market is already well below the price you paid on the previous day and is thus below your reservation value. (Why has the car's market price fallen so much in one day? This question has already been answered.)

The difference between the job-market example (Section 6) and the present model of the used-car market is that *signaling* occurs in the former and this can ensure that high-quality goods or services are credibly identified. (The outcome is not fully efficient in the job-market scenario because signaling consumes resources.) When it is possible for high-quality sellers to signal at a relatively low cost, the market can force low-quality sellers to reveal themselves.

We conclude with a numerical illustration of market failure when there is no signaling. Assume that there are many more buyers in the used-car market than sellers; competition among the latter will result in all sellers charging the same price if there is no possibility of signaling (no warranties, etc.).

Example 3.1: The car is worth 50% more to the buyer than to the seller

For a given quality level *q* the seller's reservation value is *q* and the buyers' reservation value is 1.5*q*. The buyers are risk neutral and they do not know *q*, and they do not expect owners of low-quality cars to truthfully reveal *q*. Quality is uniformly distributed over the interval from 0 to 1, as explained in Section 6.5 of Chapter 2. Because $\int x \, dx = 0.5x^2$, the expected value of a car from the buyer's perspective is

$$\int_0^1 q \, dq = \frac{1}{2} \left(1^2 - 0 \right) = \frac{1}{2}.$$

In the absence of signaling all cars sell for the same price *p*. Therefore, if q < p a seller will put her car on the market, receiving a payoff of *p*, which is higher than the payoff of *q* from keeping the car. If q > p the seller will not offer the car for sale. This enables us to determine the average quality of used cars on the market when buyers observe the price *p*: To do so we first compute the density function for the distribution of cars *q* satisfying $0 \le q \le p$. The density function for $0 \le q \le 1$ is f(q) = 1. Therefore, the density function for $0 \le q \le p$ is f(q) divided by the probability that *q* falls between 0 and *p*, and that is (1/p) f(q) = 1/p. (See the following box.)

∂3. The Used-Car Market

The average quality of cars on the market at a price of *p* is

$$\frac{1}{p} \int_0^p q \, dq = \frac{1}{p} \left(\frac{1}{2} p^2 - 0 \right) = \frac{1}{2} p$$

The risk-neutral buyer maximizes his expected payoff, and hence there will be no sale at any positive price p, because the expected payoff to a buyer would be (1/2) p - p, a negative number. There are no trades, even though every agent knows that mutually beneficial trades are possible in principle: The value of any owner's car is two-thirds of what it is worth to any buyer. *If* quality could be costlessly discerned then competition among buyers would bid up the price of a car of quality q to just about q and trade would take place.

The quality *q* is uniformly distributed on the interval [0, 1]. But only cars with $q \le p$ are on the market, so the used cars on the market are uniformly distributed on the interval [0, *p*], which has length *p*. The probability of a car on the market being somewhere in that interval must be 1. Therefore the density *d* must solve $d \times p = 1$. Note that

$$\frac{1}{p} \int_0^p f(q) \, dq = \frac{1}{p} \int_0^p dq = \frac{1}{p} (p-0) = 1$$

We have calculated the density function for $0 \le q \le p$ in Example 3.1 correctly.

As Example 3.1 demonstrates, complete collapse of a market is a theoretical possibility when one side of the market has information that is hidden from the other side. In fact, the market will be active, but will function with less than perfect efficiency. For instance, used-car dealers offer limited warranties on good-quality cars, and that mitigates the asymmetric information problem to some extent. But there will still be trades that could increase the utility of buyer and seller but that will not take place because of the hidden information problem. For instance, someone who buys a new car one week and then finds the next week that he or she has to move two thousand miles away would sell the new car and buy another in the new locale if a potential buyer could verify that the car is not being sold because it is a lemon. As it is, the individual will spend the time and money required to drive it to the new home, because net of those expenses the reservation value of the car exceeds the price for which it could be sold. One of the many applications of Example 3.1 is the market for cars that are only a few weeks or months old but are owned by people who would like to sell, but not because they have discovered their cars to be of low quality.

Source

Akerlof (1970), a seminal contribution to the theory of asymmetric information, is the basis of this section. In separate contributions, George Akerlof, Michael Spence, and Joseph Stiglitz showed that the presence of asymmetric information in real-world markets required a new way of modeling economic exchange. They were awarded the Nobel Prize in Economics for 2001.

272 Hidden Characteristics

Links

Molho (1997) has an extensive discussion of the lemons problem. Hendel and Lizzeri (1999) analyze a model that incorporates interaction between new and used-car markets over time and manufacturers whose products differ with respect to reliability. That article includes many references to the literature.

Problem set

Rework Example 3.1 with the buyer's reservation value set at λ*q* instead of 1.5*q*. (λ is some constant larger than 1.)

4 CREDIT RATIONING

The central paradigm of economic analysis is the notion that in any market operating under competitive conditions, the price will adjust until demand equals supply. This entails two principles: There will be a price P^* at which demand equals supply, and market forces will drive the price to P^* over time. How quickly the prevailing price moves close to its equilibrium value depends on the nature of the market, but in the case of financial markets we would expect fairly quick convergence to equilibrium.

In fact, credit markets are an important exception to the central paradigm because of a significant asymmetric information problem: The borrower knows considerably more about the riskiness of the project for which funding is being sought than the lender does. (Borrowers also know a lot more about their willingness to work hard to bring the project to successful completion.) This can prevent the lender from raising interest rates when the demand for loans exceeds the supply: An increase in the interest rate can induce an increase in the riskiness of the pool of applicants for loans, thereby *reducing* the lender's profit. If interest rates don't rise, and hence demand continues to exceed supply, the lender will screen applicants by investigating their background and examining in detail the business venture that will be financed by the loan. (The firm that sells me a stove doesn't care about my background.)

Asymmetry information by itself would not create problems were it not for the asymmetry in the return to the lender. On one hand, if the project flops then the lender will not be repaid at all or will only be paid a fraction of the amount borrowed. On the other hand, when the project is very successful, the lender's payoff is not proportionally high—it can never be more than the amount of the loan plus interest charges. Limited liability constrains the amount that the borrower repays when the project is a failure, but there is also a restriction on the amount to be repaid when the project is successful. Both limits are in the borrower's favor.

The asymmetry in the lender's payoff forces the lender to worry about the probability of default. By the same token, the possibility of default changes the pool of loan applications when the interest rate changes. There tend to be more very risky projects seeking funding when the interest rate is high, and hence a larger fraction of the projects would fail if all were to be funded. Assuming for the minute that that is true, it follows that when the demand for loanable funds exceeds the supply, an increase in the interest rate—which is possible, because of

4. Credit Rationing





the excess demand at the current rate—will not necessarily increase the lender's profit. The higher interest rate yields the lender a higher payoff when a project is successful, but the higher interest rate also raises the number of loans that default. Therefore, beyond a certain point, the lender will stop raising the interest rate, even though there is excess demand, and will instead devote resources to investigating the project for which an applicant is seeking funding to try and weed out the very risky ones. We refer to this as *credit rationing*. (In the market for home loans, the lender can often require the borrower to put up collateral, but there is limited scope for collateral in business loans.)

DEFINITION: Credit rationing

Credit rationing occurs at equilibrium if some borrowers' loan applications are turned down, even if they are willing to pay the market interest rate and fulfill all other requirements of the loan contract—putting up collateral, for instance.

As Figure 5.7 illustrates, credit rationing causes the supply of credit to diminish as the interest rate r increases beyond a certain level. Because the supply curve bends back at interest rates above r^0 , the demand exceeds the supply at every interest rate. Consequently, the market rate of interest will settle at a level r^* at which demand exceeds supply. Even at equilibrium the total amount of money for which borrowers apply is greater than the amount that lenders are willing to part with. The lenders then have to ration, and they typically use some measure of the degree of riskiness to screen out applications that they consider undesirable. But note that there is scope for pernicious screening devices, such as racial preferences.

The next subsection shows why there *will* be proportionally more risky projects seeking funding when the interest rate is high.

274 Hidden Characteristics

Table 5.1	1
-----------	---

	Tyler's project		Samantha's	project
Return	Probability	Payoff	Probability	Payoff
High Low	¹ / ₂ ¹ / ₂	110 100	¹ / ₂ ¹ / ₂	120 90

4.1 The borrower's point of view

We explore the credit market by means of simple examples. It is assumed throughout that both borrowers and lenders are risk neutral. That doesn't mean that lenders don't worry about risk. An increase in the riskiness of loans can reduce the lender's expected monetary value.

Example 4.1: Two borrowers

Two individuals, Tyler and Samantha, each seek financing for their projects, which are specified by Table 5.1. Assume that in both cases the project requires \$100 of capital to initiate, and each entrepreneur has applied for a \$100 loan. Tyler is hoping to finance a project that will pay \$110 with probability $1/_2$ and will return only \$100 with probability $1/_2$. In Samantha's case, the project will yield more than Tyler's high payoff if it is successful but will yield less than Tyler's low payoff in case of failure. The return in the payoff column is the profit from a project, net of all economic costs except the cost of borrowing the necessary funds. Assume for simplicity that the money is only borrowed for one year—and that the project lasts only one year. (Alternatively, we could assume that the numbers are discounted to the end of the first year.)

Note that the two projects yield the same return on average. Confirm this by calculating the EMV (expected monetary value):

EMV for Tyler = $\frac{1}{2} \times 110 + \frac{1}{2} \times 100 = 105$. EMV for Sam = $\frac{1}{2} \times 120 + \frac{1}{2} \times 90 = 105$.

But Samantha's project is riskier, because the spread between the high and low payoffs is greater.

DEFINITION: Degree of risk

If projects *A* and *B* have the same EMV and the same probability of success, we say that *A* is riskier than *B* if the high payoff from *A* is larger than the high payoff from *B*, but the low payoff from *A* is smaller than the low payoff from *B*.
4. Credit Rationing

Suppose that the current interest rate is 8%. Would Tyler be willing to borrow at that rate? The answer is "yes" if and only if the profit net of all costs, including borrowing costs, is positive. Let's calculate Tyler's EP (expected profit):

EP at 8% for Tyler =
$$\frac{1}{2} \times (110 - 100 - 100 \times 0.08) + \frac{1}{2} \times (100 - 100) = 1.$$

The borrowing cost is the principal (\$100) plus the interest ($$100 \times 8\%$), but when the project is unsuccessful Tyler can only pay back the principal. The entrepreneur's profit can never be lower than zero. The expected profit is positive, so Tyler would be one of the loan applicants at an interest rate of 8%. What if the rate were to rise to 11%? Even the high return of 110 is insufficient to cover the principal and the \$11 interest charge on the loan. Tyler's payoff would be zero whether the project succeeded or failed. Tyler would not seek funding for his project at an interest rate of 11%. (Tyler's return would actually be negative if this project were one part of an ongoing concern and the interest had to be covered by profit from the firm's other activities.)

Now let's see what Samantha would decide at the two interest rates:

EP at 8% for Sam =
$$\frac{1}{2} \times (120 - 100 - 100 \times 0.08) + \frac{1}{2} \times (90 - 90) = 6.$$

When Samantha's project is unsuccessful she can only repay 90% of the principal and none of the interest. The expected profit is positive, so Samantha would apply for a loan at an interest rate of 8%.

EP at 11% for Sam =
$$\frac{1}{2} \times (120 - 100 - 100 \times 0.11) + \frac{1}{2} \times (90 - 90) = +4.5.$$

Profit is again positive, so Samantha would seek funding for her project at an interest rate of 11%. But Tyler would not, at that higher rate, illustrating our point that the riskiness of the loan applicant pool increases when the interest rate increases. Why does Samantha stay in the hunt for funding but Tyler does not when the interest rate rises? Because, even though the payoff *to the project* is lower for Samantha than for Tyler when their projects turn sour, the two *entrepreneurs* get the same payoff in that state—zero. But Samantha gets a higher payoff than Tyler when the projects succeed. Therefore, Samantha can make a profit for herself at a higher interest rate than Tyler can.

Let's find the watershed interest rate, above which Tyler will not apply for a loan: Let *r* denote the interest rate expressed as a decimal fraction. (When the interest rate is 7% we have r = 0.07.) Then

EP at 100 r% for Tyler =
$$\frac{1}{2} \times (110 - 100 - 100 r) + \frac{1}{2} \times (100 - 100)$$

= $\frac{1}{2} \times (10 - 100 r).$

This will be positive if and only if 10 - 100 r > 0, which is equivalent to 100 r < 10. At interest rates above 10% it will not be profitable for Tyler to

undertake his project. Similarly,

EP at 100 r% for Sam =
$$\frac{1}{2} \times (120 - 100 - 100 r) + \frac{1}{2} \times (90 - 90)$$

= $\frac{1}{2} \times (20 - 100 r).$

Table 5	5.2
---------	-----

Return	Probability	Payoff
High Low	¹ / ₂ ¹ / ₂	$105 + \alpha$ $105 - \alpha$

The EP for Samantha will be positive if and only if 20—100 r < 0, which is equivalent to saying that the interest rate is less than 20%. Therefore, at interest rates between 10 and 20%, Samantha will apply for a loan but Tyler will not. In that range, only the riskier of the two projects will attempt to get funding.

Now we consider a general version of our example. Again, let r denote the interest rate expressed as a decimal fraction.

Example 4.2: A generic version of Example 4.1

Each value of α identifies a different project (see Table 5.2). Assume again that a project requires \$100 of capital, whatever the value of α . Project α pays \$105 + α with probability $\frac{1}{2}$ and only \$105 - α with probability $\frac{1}{2}$. As before, the return is net of all economic costs except the cost of borrowing the necessary funds.

Note that all projects yield the same return on average:

EMV of project
$$\alpha = \frac{1}{2} \times (105 + \alpha) + \frac{1}{2} \times (105 - \alpha) = 105.$$

However, as α increases the risk of the project increases, because the spread between the high and low payoffs increases with α . To determine if an entrepreneur who is seeking funding for project α will actually apply for a loan when the interest rate is *r*, we have to consider two cases.

Case 1: Even when the project fails it yields enough to repay the loan and cover all the interest charges. In that case

EP for project
$$\alpha = \frac{1}{2} \times (105 + \alpha - 100 - 100 r)$$

 $+ \frac{1}{2} \times (105 - \alpha - 100 - 100 r)$
 $= 5 - 100 r.$

As we expect, the borrower's profit falls as the interest rate rises. Why is profit independent of α in this case? Because the entrepreneur (borrower) gains α with probability $\frac{1}{2}$ but also loses α with probability $\frac{1}{2}$.

Things are different when the borrower has to default on the loan if the project turns sour, because in that case the higher α imposes costs on the *lender*, who gets a smaller faction of the principal repaid on average as α increases. But when

4. Credit Rationing

the project is successful, the higher the α the more the borrower gets to keep. That's why higher interest rates cause a reduction in the fraction of low-risk loans in the applicant pool.

Case 2: The borrower only pays back $105 - \alpha$ when the project fails, because $105 - \alpha$ is less than 100 + 100 r:

EP for project
$$\alpha$$
 at $r = \frac{1}{2} \times (105 + \alpha - 100 - 100 r)$
 $+ \frac{1}{2} \times [105 - \alpha - (105 - \alpha)]$
 $= \frac{1}{2} \times (5 + \alpha - 100 r).$

In this case, when a project fails, the entrepreneur gets zero—but no less.

In either case the entrepreneur's expected profit is

$$\frac{1}{2} \times (5 + \alpha - 100 r) + \frac{1}{2} \times \max\{(5 - \alpha - 100 r), 0\}.$$

If $5 - \alpha - 100 r$ is negative, then expected profit is $1/2 \times (5 + \alpha - 100 r) + 1/2 \times 0$, and if $5 - \alpha - 100 r$ is positive then expected profit is $1/2 \times (5 + \alpha - 100 r) + 1/2 \times (5 - \alpha - 100 r)$. Note that expected profit is positive if and only if $5 + \alpha - 100 r$ is positive. Therefore, the entrepreneur behind project α will apply for a loan if and only if

$$5+\alpha-100 r>0.$$

(If $5 + \alpha - 100 r$ is negative then $5 - \alpha - 100 r$ will certainly be negative.) If 100r < 5 then $5 + \alpha - 100 r$ is positive for all α , so everyone will apply for a loan when the interest rate is less than 5%. When $100 r \ge 5$ then α will apply for a loan if and only if $5 + \alpha - 100 r > 0$ or $\alpha > 100 r - 5$. The threshold value of α , which is 100 r - 5, increases with the interest rate. Projects for which α is below 100 r - 5 will not apply for a loan, but projects for which α is above 100 r - 5 will seek funding. In short, the range of α values for which a loan is sought shrinks as the interest rate increases, with the safer projects ($\alpha < 100 r - 5$) withdrawing their loan applications. In other words, the riskier the project the more likely it is that the entrepreneur will apply for a loan at a given interest rate.

4.2 The lender's point of view

At a given interest rate *r*, a risky project will always be less profitable for the lender than a safer one for Example 4.2. That is because the lender's profit is 100 *r* when the project is successful, whatever the value of α . But in the case of failure, the lender's loss increases with the riskiness of the project. The lender has advanced \$100 but only gets back $105 - \alpha$ when $100 + 100 r > 105 - \alpha$, and $105 - \alpha$ is smaller the larger is α . Therefore, the lender has an interest in screening projects to estimate the riskiness of each application when $\alpha > 5 - 100 r$.

There is more to the story. We have learned that the entrepreneurs backing less risky projects will not apply for a loan if the interest rate is sufficiently high. Therefore, the lender can't simply raise the interest rate and then accept only the safe applications: There may not be any. For Example 4.1, both Tyler and

Samantha will apply for a loan when the interest rate is 8%, but only Samantha will apply at 11%.

Suppose that there are two applicants, Tyler and Samantha of Example 4.1, when the interest rate is 8%. What is the lender's profit? When a project succeeds the lender gets the principal back and an interest payment of \$8 from each borrower. When Tyler's project fails the lender gets the principal back but there is no interest payment. When Samantha's project fails then the lender only gets \$90. There is a loss of \$10. Assuming that the two projects are statistically independent, the lender's expected profit is

EP of lender at
$$8\% = \frac{1}{2} \times 8 + \frac{1}{2} \times 0 + \frac{1}{2} \times 8 + \frac{1}{2} \times -10 = +3$$
.

There is a profit of \$3.

At an interest rate of 6% we have

EP of lender at
$$6\% = \frac{1}{2} \times 6 + \frac{1}{2} \times 0 + \frac{1}{2} \times 6 + \frac{1}{2} \times -10 = +1$$
,

which is lower than the lender's profit at 8%. (Confirm that both Tyler and Samantha would apply for a loan when the interest rate is 6%.) Therefore, we know that there will be a range of interest rates at which profit increases as the interest rate increases. In other words, there is a range of interest rates in which the lender has an incentive to raise the rate if demand for credit exceeds supply. Hence, there will be an upward sloping segment of the supply of credit curve—the piece below r^0 in Figure 5.7.

Now consider the situation when the interest rate is 11%. Tyler will not apply for a loan! If Samantha's project were to succeed the lender gets the principal back and an interest payment of \$11, but when it flops the lender loses \$10. The lender's expected profit is

EP of lender at
$$11\% = \frac{1}{2} \times 11 + \frac{1}{2} \times -10 = +0.5$$
.

The lender's profit is lower at 11% than at 8% because of the change in the set of projects seeking credit. We refer to this as *adverse selection*. This accounts for the downward sloping part of the supply curve—that is, the piece above r^0 in Figure 5.7.

Return to Example 4.2. Assume temporarily that when the project fails there is not enough money to cover both the principal and the interest on the loan. When the interest rate is 100 r% and project α succeeds the lender gets the principal back along with an interest payment of 100 r. But when project α fails the lender only gets $105 - \alpha$. There is a loss of $100 - (105 - \alpha) = \alpha - 5$.

EP of lender from
$$\alpha = \frac{1}{2} \times 100 r + \frac{1}{2} \times (5 - \alpha) = \frac{1}{2} \times (100 r + 5 - \alpha).$$

This will be positive if and only if $100 r + 5 - \alpha > 0$. That is equivalent to

$$\alpha < 100 r + 5.$$

In other words, given the interest rate *r*, only projects whose risk parameter α is below 100 *r* + 5 will be profitable for the lender, who will want to screen out projects for which α exceeds 100 *r* + 5.

4. Credit Rationing

But what about cases where the borrower can repay both principal and interest, even when the low payoff is realized? Could any of these projects have a risk parameter α that violated the inequality $\alpha < 100 r + 5$? No. If $5 - \alpha - 100 r \ge 0$ then $\alpha \le 5 - 100 r$, which certainly implies $\alpha < 5 + 100 r$. Therefore, for *all* projects the entrepreneur backing the project will apply for a loan if and only if $\alpha < 100 r + 5$. The higher the interest rate the higher the proportion of risky projects in the loan application pool.

Source

Stiglitz and Weiss (1981) worked out the theory that is sketched in this section.

Links

Jaffee and Stiglitz (1990) and Chapter 1 of Freixas and Rochet (1997) provide comprehensive discussions of credit rationing.

Problem set

- 1. This question involves a set of entrepreneurs, each of whom is seeking funding for a project. Each project requires an initial \$100 investment. To keep the calculations simple we assume that each project will be 100% loan financed *if* the entrepreneur decides to carry it out. Each entrepreneur is identified with a number α : Entrepreneur α 's project will return 100 + α with probability $\frac{2}{3}$ and will return 100 - 0.5 α with probability $\frac{1}{3}$.
 - **A.** For each combination of interest rate *r* (expressed as a decimal fraction) and α represented in Table 5.3, determine if entrepreneur α will apply for a loan to fund the project at the specified interest rate.

Table 5.3

$\alpha = 5$ and $r = 0.04$	$\alpha = 5$ and $r = 0.08$	$\alpha = 5$ and $r = 0.10$
$\alpha = 9$ and $r = 0.04$	$\alpha = 9$ and $r = 0.08$	$\alpha = 9$ and $r = 0.10$
$\alpha = 12$ and $r = 0.04$	$\alpha = 12$ and $r = 0.08$	$\alpha = 12$ and $r = 0.10$

- **B.** As a function of α , what is the threshold rate of interest, above which project α will not seek funding?
- **2.** This question concerns a set of projects that return \$120 when successful and \$60 otherwise. What makes one project riskier than another in this case is the fact that the probability π of success is lower for riskier projects. Each project requires an initial \$100 bank loan. The bank charges an interest rate of 100 r% on loans. *As a function of r*, what is the value of π below which the bank will not make a profit by funding the project?
- **3.** What is the economic rationale for limited liability, which protects the borrower but not the lender?

- **4.** Why doesn't the lender insist on an equity stake in the project that would allow the lender to get a proportionally higher payoff when the project is proportionally more successful?
- **5.** Why does the lender usually insist that the borrowers put some of their own money into their projects?

∂5 BUNDLING AND PRODUCT QUALITY

The manufacturer of an appliance or a car knows that some consumers have a high willingness to pay for a luxury model but would buy the economy version if the sticker price on the luxury model is too high. Because many consumers aren't prepared to buy a luxury model at any price that would be profitable for the manufacturer, the firm may have to supply two models—luxury and economy—to maximize profit. However, the existence of the economy version puts a constraint on the sticker price of the luxury model. This section demonstrates how to solve for the profit-maximizing menu of models and associated prices.

If the individual's willingness to pay for each model were actually known to the manufacturer, the firm could charge each consumer the maximum the consumer would be willing to pay for the model—the one that is most profitable for the firm to sell to that individual—given the consumer's willingness to pay for each model. As it is, willingness to pay is hidden from the supplier.

Assume that quality can be measured. We let *x* denote the amount of quality embedded in the model. Then higher values of *x* represent higher quality. Of course quality is really multidimensional, particularly in the case of a sophisticated product such as a car. But even a one-dimensional quality parameter gives us a framework from which we can draw much insight.

DEFINITION: Quality

We let *x* denote the level of quality in a particular unit of a good. A *package* (*x*, *C*) consists of a model embodying *x* units of quality that sells for *C*.

It is also possible to interpret *x* as the *quantity* of some good. We begin with that interpretation, and when the analysis is complete we reinterpret our findings in terms of quality choice by the firm.

A monopoly is attempting to price discriminate by offering its output in the form of sealed packages with fixed prices. A package containing more output bears a higher price tag, but the price is not a linear function of quantity. If package *B* has twice as much output as package *A* its price will be more than double that of *A* if the larger package is targeted for consumers who get more benefit from the good and are willing to pay proportionally more. This will generate more profit than a linear pricing schedule. (A linear schedule can be represented by a single number—the price. If *Q* units cost *Q* times as much as one unit let *P* denote the cost of one unit. For arbitrary *Q* the total cost is $P \times Q$. With linear pricing the seller merely has to announced *P*.)

There are two potential problems for a monopolist attempting to impose a nonlinear, price-discriminating schedule. We illustrate the first with a simple example.

Example 5.1: Unbundling

Suppose package *A* contains ten units of output and sells for \$10, and *B* contains twenty units of output and is priced at \$30. Even if the manufacturer won't let a particular consumer buy two *A* packages, an arbitrageur can purchase two *A* packages and sell them for a total of \$25 to someone who otherwise would have to buy a *B* package from the monopolist. Similarly, if there is a quantity discount, with *A* selling for \$10 and *B* selling for \$15, an entrepreneur might be able to buy a *B* package and divide it into two ten-unit packages and sell each for \$9 to individuals who might otherwise have to pay \$10 to the monopolist.

Note that *quality* cannot be unbundled. A garage cannot take apart a \$50,000 automobile and use the parts to make two cars that sell for \$30,000 each. When we apply the analysis of this section to *quantity* bundling we must confine our attention to goods that cannot be resold. (Or goods such as airline travel for which resale can be blocked by the seller, in this case by checking the traveler's identity.) The case of a public utility producing electricity will provide the motivation. It is possible to store electricity for future resale but it is very costly to do so. We can assume that our public utility monopoly does not have to worry about resale.

The second difficulty arises from the fact that the monopolist cannot identify the consumers who are willing to pay more for electricity. Those individuals cannot be expected to voluntarily disclose their identity, knowing that they will be charged more when they do. In fact a consumer for whom the product provides a high level of benefit and for whom the *B* package is targeted can buy two "A packages." The monopolist can rule this out simply by offering this consumer an all-or-nothing proposition: "Either you buy one B package or we will not sell you anything." But there is a hidden characteristic problem. The firm cannot directly identify the individuals who derive a high level of benefit from the product. The best that the monopolist can do is to design the packages so that high-benefit consumers will not want to buy package A even though it costs less per unit than *B*. They will want to purchase *B* at the proportionally higher price because it provides more of the good. Of course they will not want B if the cost per unit is too high. Designing the packages and choosing the price tags is not a simple task. The trick is to design the packages and select the prices so that a high-benefit customer will choose to buy the package the monopolist designed for that person.

Of course, our analysis applies only to cases for which the high-benefit consumer is unable to buy multiple *A* packages, and thereby get the same quantity as in a *B* package, but at a lower total cost. For instance, the monopoly public utility can control delivery and will offer a consumer an all-or-nothing choice between the two packages. When "quantity" actually refers to the "amount of quality" our assumption is satisfied because two low-quality appliances do not amount

to the same thing as a single high-quality appliance. We begin the analysis by specifying production costs and consumer preferences.

5.1 The Model

The model has two commodities, *X* and *Y*. The former is the one that we're interested in, and we can interpret it as an automobile or electricity and so forth. In fact, *X* is anything that can't be unbundled. *Y* denotes money, or generalized purchasing power. That is, *Y* is a composite commodity, representing expenditure on goods other than commodity *X*.

Customer *i*'s utility function is $U_i = B_i(x_i) + y_i$ where x_i and y_i are the amounts of *X* and *Y*, respectively, consumed by individual *i*. The benefit that *i* derives from x_i units of *X* is $B_i(x_i)$. That is, B_i is a benefit function. We assume that the marginal benefit of *X* is positive for all x_i but that marginal benefit decreases and as x_i increases. Suppose that each individual is endowed with (begins with) ω_i units of *Y*. If *i* pays a total of C_i dollars for x_i units of commodity than *X* then *i*'s utility will be $U_i = B_i(x_i) + \omega_i - C_i$. Because ω_i is constant we need only compute the change in utility, which is $\Delta U_i = B_i(x_i) - C_i$ assuming that $B_i(0) = 0$. If ΔU_i is positive then *i* will purchase the package (x_i, C_i) or some more attractive package if one is available, but if $\Delta U_i < 0$ then *i* will not purchase (x_i, C_i) because it would cause utility to decline. This takes us to the participation constraint.

DEFINITION: The participation constraint

If the monopolist's profit-maximizing strategy involves consumer *i* purchasing package (x_i, C_i) then $B_i(x_i) - C_i \ge 0$ must hold.

Finally, we choose units so that one unit of *X* costs \$1 to produce. Then if there are *n* individuals, and each individual *i*, purchases x_i units of *X*, the producer's cost will be $x_1 + x_2 + \cdots + x_n$. Its revenue is the total amount paid: $C_1 + C_2 + \cdots + C_n$. The profit-maximizing menu may contain many packages for which $x_i = 0 = C_i$.

DEFINITION: The firm's profit

There are *n* consumers. If the monopolist offers the menu of packages (x_1, C_1) , (x_2, C_2) , ..., (x_n, C_n) and each consumer buys exactly one of the packages, then profit is

 $C_1+C_2+\cdots+C_n-x_1-x_2-\cdots-x_n$

because one unit of *X* costs \$1 to produce.

There are *n* consumers, but there will typically be fewer than *n* types. For instance, consumers 1, 2, ..., *m* may be of one type and consumers m+1, m+2, ..., *n* another.

∂5.2 Full information equilibrium

To give us a point of comparison, begin with the full information assumption that the monopolist knows each person's benefit function B_i . What package should be offered to person i? Let x_i^0 be the value of x_i that maximizes $B_i(x_i) - x_i$, which is the consumer's benefit less the monopolist's cost of producing x_i . That is, x_i^0 solves $B'_i(x_i) = 1$. (Note that diminishing marginal benefit means that the second derivative of the function that we are maximizing is negative, and hence the first-order condition is sufficient for a global maximum.) Suppose the monopolist offers i the package $(x_i^0, B_i(x_i^0) - \epsilon)$. That is, the total cost to iof x_i^0 units of X is $C_i^0 = B_i(x_i^0) - \epsilon$. Here ε is a *very* small positive number, so the charge is just slightly less than the total benefit. Person i faces a take-itor-leave-it proposition. Because $\Delta U_i = B_i(x_i^0) - C_i = B_i(x_i^0) - [B_i(x_i^0) - \epsilon] = \epsilon$, which is positive, the monopolist's offer will be accepted. This is the profitmaximizing strategy under the full information assumption that there is no hidden characteristic problem.

Why is it profit maximizing under full information? If $\Delta U_i < 0$ then *i* will not buy the package and the monopolist will receive zero profit from *i*. Therefore, the monopolist must respect the participation constraint $B_i(x_i) - C_i \ge 0$. As long as C_i is substantially below $B_i(x_i)$ the monopolist can raise C_i without violating $\Delta_i U_i > 0$ and sell the same x_i units at a higher price. Therefore, profit maximization requires C_i almost equal to $B_i(x_i)$. Let's approximate and set C_i exactly equal to $B_i(x_i)$. But we do not know what x_i is. The monopolist wants to maximize $C_i - x_i$ but $C_i = B_i(x_i)$ under profit maximization. Because x_i^0 denotes the level of x_i that maximizes $B_i(x_i) - x_i$ we have indeed found the profit-maximizing set of take-it-or-leave-it offers.

Surprisingly, we have an efficient outcome even though the monopolist has succeeded in extracting all the benefit from each consumer. (Set $\epsilon = 0$ for convenience.) $\Delta U_i = B_i(x_i^0) - C_i^0 = B_i(x_i^0) - B_i(x_i^0) = 0$. Therefore, each consumer pays a charge equal to the benefit the *consumer* derives from the *X* received and there is no net gain in utility. Nevertheless, the outcome is efficient if all the profits are returned to the community. (The company's shareholders are members of the community.) To prove this, we show that any outcome satisfying $x_i = x_i^0$ for all *i* actually maximizes total utility, $\sum_i U_i$, as long as $\sum y_i$, the total amount of *Y* consumed, equals the total amount left over after the required $\sum x_i^0$ units are used in the production of *X*. That is, as long as $\sum y_i = \sum \omega_i - \sum x_i^0$ holds. (\sum denotes summation over all individuals.)

$$\sum U_i = \sum [B_i(x_i) + y_i] = \sum B_i(x_i) + \sum y_i$$
$$= \sum B_i(x_i) + \sum \omega_i - \sum x_i,$$

and this is maximized by setting $B'_i(x_i) - 1 = 0$ for arbitrary *i*. We know that x_i^0 is the unique solution to this equation.

Now set $\sum U_i = \sum [B_i(x_i^0) + y_i]$. Then $\sum U_i = \sum B_i(x_i^0) + \sum y_i$, and this total is preserved if we redistribute commodity *Y* among the consumers as long as the total $\sum y_i$ is unchanged. Therefore, any outcome maximizes total utility if $x_i = x_i^0$ for each *i* and $\sum y_i = \sum \omega_i - \sum x_i^0$. Therefore, any such outcome is efficient: If

we could make one person's utility higher without lowering anyone else's we could make the sum higher, which is impossible. (See Section 5.1 of Chapter 2 on this point.)

In general, any outcome that has one of the agents extracting all of the surplus from the other agents is efficient. Once we specify formally what we mean by "extracting all the surplus" it is easy to prove efficiency. Suppose that each agent *i* has some initial level of utility μ_i . Agent 1 extracts all of the surplus from each of the other agents if, for all i > 1, the *final* level of utility is equal to μ_i , the starting level. In symbols, agent 1 chooses the outcome so as to maximize U_1 subject to the constraint $U_i > \mu_i$ for all i > 1. The solution s^* to this problem must be efficient: If s^* is *not* efficient then there is a feasible outcome t^* that gives one person more utility than s* and gives everyone at least as much utility as s*. If $U_1(t^*) > U_1(s^*)$ we contradict the fact that s^* maximizes U_1 subject to $U_i \ge \mu_i$ for all i > 1, because $U_i(t^*) \ge U_i(s^*) \ge \mu_i$ for i > 1. If $U_i(t^*) > U_i(s^*)$ for some i > 1 then we can extract a tiny amount of commodity Y from person *j* and still have $U_i > U_i(s^*)$, provided that we are careful to confiscate a sufficiently small amount of Y. If we then give this tiny amount of Y to person 1 we will have $U_1 > U_1(s^*)$, again contradicting the fact that s^* solves the constrained maximization problem, because $U_i(t^*) \ge U_i(s^*)$ for i > 1.

Before determining the equilibrium outcome in the real-world interpretation of our model, with asymmetric information, we present the simple example that is used to illustrate the theory.

Example 5.1: Two preference types

There are exactly two types of consumers, *H* (high benefit) and *L* (low benefit). Suppose $U_H = 4\sqrt{x_H} + y_H$ and $U_L = 2\sqrt{x_L} + y_L$. At the full information profit-maximizing equilibrium we have $C_i = B_i$ for i = H and i = L. Therefore, a firm chooses x_H to maximize $4\sqrt{x_H} - x_H$. The first derivative of this function is $(2/\sqrt{x_H}) - 1$, and the second derivative is negative. Therefore, we solve $(2/\sqrt{x_H}) - 1 = 0$ to get $x_H^* = 4$. Similarly, x_L^* maximizes $2\sqrt{x_L} - x_L$. The first derivative is $(1/\sqrt{x_L}) - 1$, and when we set that equal to zero we get $x_L^* = 1$. Then $C_H^* = B_H(x_H^*) = 4\sqrt{4} = 8$ and $C_L^* = B_L(x_L^*) = 2\sqrt{1} = 2$. The full information equilibrium has the monopolist selling 4 units of *X* to each *H* type, on a take-it-or-leave-it basis, at a price of \$8 for all 4 units, and selling 1 unit of *X* to each *L* type at a price of \$2, also on a take-it-or-leave-it basis. If the number of *H* types and *L* types is n_H and n_L , respectively, then the firm's profit is

$$n_H \times (8-4) + n_L \times (2-1) = 4n_H + n_L$$

Verify that $n_H \times U_H + n_L \times U_L$ (total utility) is maximized however the total profit is divided between the individuals. (If $n_H = 1 = n_L$ then profit is 5.)

Example 5.1 will be the subject of our inquiry for the rest of Section 5.

∂5.3 Asymmetric information equilibrium

We continue our investigation of Example 5.1, but we now drop the full information assumption that the individual with utility function $B_H(x) + y$ can be identified. We do assume, however, that the monopolist knows the functional forms B_H and B_L . Of course, it does not know which person has which function. We also assume (for convenience) that there are exactly as many H types as Ltypes. Consequently, total profit is maximized when the firm has maximized the profit from a pair of individuals consisting of one H and one L. When we use the word profit from now on we will be referring to the profit from sales to one H-Lpair. If it helps, you can assume that the market consists of one H person and one L person.

An offer (α , β) is a specification of the amount α of *X* in the package and the price β of the package. What is the profit-maximizing menu of offers? If the monopolist simply offers each person a choice of (1, 2) and (4, 8), the profit-maximizing strategy under full information, the *H* types will choose the former:

 $\Delta U_H(1,2) = 4\sqrt{1} - 2 = 2$ but $\Delta U_H(4,8) = 4\sqrt{4} - 8 = 0.$

The L types will choose (1, 2) also:

$$\Delta U_L(1,2) = 2\sqrt{1} - 2 = 0$$
 but $\Delta U_L(4,8) = 2\sqrt{4} - 8 = -4$.

The monopolist's profit will be 2 + 2 - (1 + 1) = 2, which is *not* a maximum even under the assumption that B_H and B_L cannot be identified by the monopolist. The monopolist can continue to offer (1, 2), the contract that extracts all the surplus from the *L* types, but design a contract (x_H , C_H) such that the *H* types will not prefer (1, 2), and it will otherwise extract as much surplus as possible. The first consideration requires

$$4\sqrt{x_H} - C_H \ge 4\sqrt{1} - 2.$$

This is called a *self-selection*, or *incentive compatibility*, constraint.

DEFINITION: Self-selection constraints

If the producer wants the *H* type to choose x_H and the *L* type to choose x_L then the respective package costs C_H and C_L must satisfy

$$\Delta U_H(x_H, C_H) \geq \Delta U_H(x_L, C_L)$$

and

$$\Delta U_L(x_L, C_L) \geq \Delta U_L(x_H, C_H)$$

We justify this later, but let's assume that in computing the profit-maximizing strategy, we don't have to worry about the *L* types buying the package designed for the *H* types. ("If they want to buy the upscale package, let 'em.") The monopolist will maximize $C_H - x_H$ subject to $4\sqrt{x_H} - C_H \ge 2$. If $4\sqrt{x_H} - C_H > 2$ then C_H can be increased without violating the self-selection constraint and without

changing x_H . This will increase profit, so profit maximization requires $4\sqrt{x_H} - C_H = 2$, or $C_H = 4\sqrt{x_H} - 2$. Then the monopolist maximizes $4\sqrt{x_H} - 2 - x_H$. Setting the first derivative equal to zero gives us $2/\sqrt{x_H} = 1$, and thus $x_H = 4$. (This does yield a maximum because the second derivative is negative.) The same level of service is provided (this is a consequence of our simple utility functions) but this time at a charge of $C_H = 4\sqrt{x_H} - 2 = 6$. The monopolist's profit is 6 + 2 - (4 + 1) = 3, which is lower than under the full information assumption of voluntary disclosure of one's type (5) but higher than 2, which is the profit that results when the "full information" solutions are packaged and the consumers are allowed to choose between them.

The strategy of placing packages (1, 2) and (4, 6) on the market and letting the individual choose results in an efficient outcome: We have already shown that $x_L = 1$ and $x_H = 4$ maximizes $U_L + U_H$. We still haven't reached the maximum profit, however. Suppose the monopolist offers only one package, (4, 8). Four units of *X* at a total cost of \$8, take it or leave it:

$$\Delta U_L(4, 8) = 2\sqrt{4} - 8 = -4$$
 and $\Delta U_H(4, 8) = 4\sqrt{4} - 8 = 0.$

Consumer *L* will be better off not buying the package and would not buy even if the cost were reduced slightly. Consumer *H* would buy the package because it satisfies *H*'s participation constraint—with no room to spare: $\Delta U_H(4, 8) = 0$. The firm sells one package—to consumer *H*—and its profit is 8—4 = 4. This is the highest profit yet, apart from the full information solution. (In fact, to induce the *H* type to buy, the cost of \$8 would be reduced slightly, and the resulting profit would be slightly less than \$4, say \$3.99.)

To show that the strategy of offering only one package, (4, 8), actually maximizes profit, subject to the participation and self-selection constraints, we begin at the beginning: Maximize $C_L + C_H - x_L - x_H$ subject to *H*'s self-selection constraint $4\sqrt{x_H} - C_H \ge 4\sqrt{x_L} - C_L$ and the participation constraints $4\sqrt{x_H} - C_H \ge 0$ and $2\sqrt{x_L} - C_L \ge 0$.

Why don't we have to worry about *L*'s self-selection constraint? Because $2\sqrt{x_L} - C_L \ge 0$ by the participation constraint, and thus $2\sqrt{x_L} - C_L \ge 2\sqrt{x_H} - C_H$ is automatically satisfied if $2\sqrt{x_H} - C_H \le 0$. But suppose that $C_H < 2\sqrt{x_H}$. Then $C_H < 4\sqrt{x_H}$ and $2\sqrt{x_L} - C_L \ge 2\sqrt{x_H} - C_H > 0$, and thus $C_L < 2\sqrt{x_L}$. Because $C_H < 4\sqrt{x_H}$ and $C_L < 2\sqrt{x_L}$ we can increase *both* C_L and C_H by the same amount (keeping x_L and x_H fixed) without violating the participation constraints, provided that the increase is sufficiently small. Moreover, the *H* and *L* self-selection constraints will still be satisfied because the left-hand and right-hand sides will fall by the same amount in each one. Profit will have increased without violating any of the constraints. Therefore, we will not have $C_H < 2\sqrt{x_H}$ at the profit-maximizing outcome. Consequently, in searching for the profit-maximizing decision we can ignore *L*'s self-selection constraint.

The next step is to show that profit maximization implies $C_L = 2\sqrt{x_L}$. If $2\sqrt{x_L} - C_L > 0$ then we can increase C_L without violating $2\sqrt{x_L} - C_L \ge 0$. If we don't change any of the other variables we will have increased profit, and *H*'s self-selection constraint will still hold as a consequence of increasing C_L alone.

Therefore, profit maximization requires $2\sqrt{x_L} - C_L = 0$, and hence $C_L = 2\sqrt{x_L}$. In that case, *H*'s self-selection constraint reduces to

$$4\sqrt{x_H} - C_H \ge 4\sqrt{x_L} - 2\sqrt{x_L} = 2\sqrt{x_L}.$$

It follows from $4\sqrt{x_H} - C_H \ge 2\sqrt{x_L}$ that *H*'s self-selection constraint must hold as an equality, and hence $C_H = 4\sqrt{x_H} - 2\sqrt{x_L}$. If, to the contrary, $4\sqrt{x_H} - C_H > 2\sqrt{x_L}$ holds then profit could be increased by increasing C_H without changing x_H or x_L , and this could be done without violating *H*'s self-selection constraint. (Note that *H*'s participation constraint $4\sqrt{x_H} - C_H \ge 0$ will automatically hold if $2\sqrt{x_L} - C_L \ge 0$ and *H*'s self-selection constraint is satisfied.) Now that we have $C_H = 4\sqrt{x_H} - 2\sqrt{x_L}$ we can express profit as

$$C_H + C_L - x_H - x_L = 4\sqrt{x_H} - 2\sqrt{x_L} + 2\sqrt{x_L} - x_H - x_L$$

= $4\sqrt{x_H} - x_H - x_L$.

All of the constraints have been embodied in the objective function $4\sqrt{x_H} - x_H - x_L$. Obviously, maximization of this expression requires $x_L = 0$, and hence the participation constraint implies $C_L = 0$. Then we want to maximize $4\sqrt{x_H} - x_H$. The first derivative is $2/\sqrt{x_H} - 1$. When we set that equal to zero we get $x_H = 4$. Then $C_H = 4\sqrt{x_H} - 2\sqrt{x_L} = 8$.

The asymmetric information equilibrium of the market of Example 5.1

The equilibrium has $x_H = 4$, $C_H = 8$, and $x_L = 0 = C_L$. Profit is 8 - 4 = 4.

Interpret x_i as the model designed for the consumers in market group *i*, with the model identified by the "amount" of quality that it provides. We can see why

Volkswagen stopped producing the Beetle for export to the United States and Canada in the 1970s at the height of its popularity. The decision came in the wake of legislation in both countries that introduced strict safety and emission standards that would have substantially increased the cost of producing the Beetle. The fourth problem set question asks you to show that an increase in the cost of production can result in the manufacturer canceling a popular product line, retaining only the more expensive version. profit-maximizing firms sometimes discontinue production of a popular model if it is at the low quality end of the spectrum. Doing so relaxes the constraint on the price of the luxury model. This gives a new interpretation to the automobile manufacturer's boast that features that used to be optional are now standard. (In this case, low quality does not mean unreliable; it simply means less luxurious.)

Does the material in this section shed any light on why publishers of textbooks stop selling the first edition of a book after the second edition appears, even when they have a stock of first editions that could be sold at a discounted price?

The profit-maximizing solution is not efficient. We don't know the utility level of each individual because we don't know the share of profit received by each. But we can compute the change in utility for each as a result of an increase in the production of *X* by one unit, if that unit is delivered to person *L* and at

the same time *L*'s consumption of *Y* is reduced by one unit. If we do not change x_H or y_H we have a feasible outcome. Certainly, *H*'s utility will not change. The change in *L*'s utility is $2\sqrt{1} - 1 = 1$, so *L* is better off and *H*'s utility is unchanged. Note that profit is unchanged: The unit of *Y* needed to finance the production of the additional unit of *X* is contributed by *L*, who also receives the extra unit of *X*. Because profit is unchanged, there is no change in the income of the firm's shareholders, and hence no change in the utility of anyone in society, other than *L*.

This is a special case of a general phenomenon. The strategy that maximizes profit subject to self-selection constraints typically results in inefficiency. If we ignore the self-selection constraints and simply maximize profit subject to the participation constraints then we get an efficient outcome because we are maximizing one agent's payoff subject to preventing the payoff of everyone else from falling below a given level. (See the argument just prior to Example 5.1.) However, in the real (asymmetric information) world, we must add self-selection constraints, and hence the resulting profit-maximizing solution is constrained away from an efficient one.

There is a defect in our argument that the asymmetric information equilibrium can be modified to increase one person's utility without diminishing anyone else's utility. We assumed that the government or some central agency could identify the consumer with utility function U_L . But if this is possible then the firm can do so as well, and it will impose the full information profitmaximizing outcome which *is* efficient. (The identity of *H* is known *after* the individual choices are made, but if *H* knows in advance that this disclosure will be used to modify the outcome then *H* would have behaved differently in the first place.)

Can we find a way of giving individuals more utility than they enjoy at the asymmetric information equilibrium without employing the full information assumption? Let's try: Let *R* be the share of the profit received by *H* at the asymmetric information equilibrium. The total profit is 4, so the share of the profit that goes to *L* is 4 - R. The utility levels at the asymmetric information equilibrium are

$$U_L = 2\sqrt{0} + \omega_L + 4 - R = \omega_L + 4 - R$$

and

$$U_H = 4\sqrt{4} + \omega_H + R - 8 = \omega_H + R.$$

(Note that *R* could be negative.) If we set $x_H = 4$ and $x_L = 1$ can we satisfy the self-selection constraints $4\sqrt{4} - C_H \ge 4\sqrt{1} - C_L$ and $2\sqrt{4} - C_H \le 2\sqrt{1} - C_L$? Both will hold if and only if we have $2 \le C_H - C_L \le 4$. Set $C_L = 1$ and see what happens. Then $C_H = 4$ will satisfy the constraints, and we have each person paying for what he or she consumes.

The government doesn't have to identify the individuals; it just has to make the two packages available, and in their self-interest consumer *H* will choose the package with 4 units of *X* and a price tag of \$4 and *L* will choose the package with

∂5. Bundling and Product Quality

1 unit of *X* and a price tag of \$1. (A profit-seeking firm wouldn't do this because it yields a profit of 0.) The resulting utility levels will be

$$U_L = 2\sqrt{1 + \omega_L - 1} = 1 + \omega_L$$

and

$$U_H = 4\sqrt{4} + \omega_H - 4 = 4 + \omega_H.$$

Do we have $1 + \omega_L > \omega_L + 4 - R$ and $4 + \omega_H > \omega_H + R$? If so, the utility of both *L* and *H* has increased.

The inequalities hold if 3 < R < 4, but how can *R* be determined? How can the government determine what share of the profit goes to the individual with utility function U_H under the asymmetric information assumption? Of course, after the individuals are observed to make their choices the government will know who's who, and then it can transfer some *Y* between the consumers to ensure that both have a higher level of utility than under the asymmetric information equilibrium. That is, the government can choose y_L and y_H so that

$$U_L = 2\sqrt{1} + y_L \text{ and } U_H = 4\sqrt{4} + y_H,$$

$$y_L + y_H = \omega_L + \omega_H - 5,$$

$$2 + y_L > \omega_L + 4 - R \text{ and } 8 + y_H > \omega_H + R$$

If we set $y_L = -y_H + \omega_L + \omega_H - 5$ then the desired inequalities are

$$2 - y_H + \omega_L + \omega_H - 5 > \omega_L + 4 - R$$

and

$$8 + \gamma_H > \omega_H + R.$$

Set $y_H = \omega_H + R - 7.5$. Now we have $U_H > \omega_H + R$. (Assume $\omega_H \ge 7.5$ to ensure $y_H \ge 0$.) Also, $y_L = -y_H + \omega_L + \omega_H - 5 = \omega_L + 2.5 - R$, and hence

$$U_L = 2 + y_L = \omega_L + 4.5 - R > \omega_L + 4 - R.$$

So, we can find transfers that do the job. However, if the consumers know that the transfers are part of the utility-enhancing change, and they know that the size and direction of the transfers depend on their choices, the incentives for both to choose the package targeted for each of them are undermined.

Asymmetric information induces an equilibrium that is not efficient, but we're not sure how to modify the rules of the game to guide the community to an efficient outcome that leaves everyone better off than under the profitmaximizing asymmetric information equilibrium or even if it is possible to do so.

Source

The example of this section is based on Arrow (1984).

Problem set

1. Let $U_H = 8\sqrt{x_H} + y_H$ and $U_L = 6\sqrt{x_L} + y_L$. The monopolist offers two packages on a take-it-or-leave-it basis. Package M_L is designed for consumer L

and offers 9 units of X at a total cost of \$18. Package M_H is designed so that consumer H will not prefer M_L , and M_H maximizes profit given that selection constraint. Derive M_H .

- **2.** Electricity is provided by a monopoly. There are two consumers, H and L, with utility functions $U_H = 6 \ln(x_H + 1) + y_H$ and $U_L = 4 \ln(x_L + 1) + y_L$. Each unit of *X* costs \$1 to produce.
 - **A.** Compute the full information equilibrium. Determine the monopolist's profit.
 - **B.** Assume that the monopolist continues to offer the package designed for consumer *L* in your solution to *A*. What is then the profitmaximizing package associated with consumer *H*, assuming that the monopolist cannot determine who *H* is or who *L* is? What is the associated profit?
 - **C.** Determine the asymmetric information equilibrium and the monopolist's profit.
 - **D.** Rank the three profit levels.
- **3.** Rework question 2 when each unit of *X* costs \$2 to produce, leaving all other features of the model unchanged.
- **4.** There are two consumers, *H* and *L*, with utility functions $U_H = 20 \ln(x_H + 5) + y_H$ and $U_L = 15 \ln(x_L + 5) + y_L$. It costs *x* dollars to produce *x* units of commodity *X*. Show that $x_L > 0$ at the asymmetric information equilibrium but $x_L = 0$ when the cost of production doubles. Is the framework of this section appropriate if commodity *X* is an automobile?

6 JOB-MARKET SIGNALING

When the hidden characteristic is a quality variable, and quality can be either good or bad, producers of the high-quality version have an incentive to signal their quality. But would the signal be credible? Only if the low-quality supplier cannot gain by transmitting the same signal. When would this be possible? Typically signaling consumes resources, and it is substantially more costly for the supplier of the low-quality commodity to transmit the same signal as the supplier of the high-quality commodity. Under the right conditions, the additional cost to those who have only low-quality items to sell motivates them to provide a weaker signal and hence reveal their type. However, because the signal imposes real costs on the individual and on society, truthful revelation comes at a price: The resources consumed in signaling do not provide any direct utility. We show that there is a range of signals consistent with equilibrium, and often the same result could have been obtained with a lower investment in signaling. There are even cases in which everyone invests in signaling but the signaling doesn't distinguish the high-quality from the low-quality producers.

6. Job-Market Signaling

We use the labor market to illustrate. The signal is the amount of training that an individual has undergone. Training is costly and individuals who know themselves to be innately intelligent and hard working (relatively speaking) are more likely to graduate and be certified. Therefore, the population of graduates contains a disproportionately high number of individuals who are innately productive—that is, intelligent and hard working. In other words, training can be used to sort workers into H types, who have relatively high ability and productivity, and L types, who have relatively low ability and productivity. A simple model makes the point.

6.1 To make a long story short

High-quality (*H* type) workers generate substantially more profit for the firm than low-quality workers (*L* types) because *H* types are more productive. Although individuals (and their parents) make choices during the formative years that help to determine a person's type, at the time an employer makes a hiring decision the worker's type has been determined. Therefore, the employer faces a hidden characteristic problem. To attract *H* types a firm can offer a higher wage: $W_H > W_L$, where W_H is the wage paid to *H* types and W_L is the wage paid to *L* types. (We refer to the payment to the worker as a wage, but it is in fact the present value of the expected compensation—including benefits—over the lifetime of the job.) But the *L* types cannot be expected to truthfully identify themselves, claiming the lower wage W_L . If we suppose that production takes place in teams in a setting that makes it impossible to identify the contribution of individuals in the short run, then it is not possible for a manager to directly separate the *L* types from the *H* types when they are hired.

Suppose, however, it costs H types C_H dollars to graduate from college and it costs L types a higher amount, C_L . The L types may take one or two extra semesters to graduate. They will also have to work much harder in high school to get admitted to a good college. Then it is possible to induce the L types to reveal themselves in spite of their preference for anonymity. The firm simply pays a salary W_H to anyone who has a graduation certificate and a salary W_L to workers without a certificate. If $W_H - C_L < W_L$ the L types will not pay the cost C_L necessary to obtain a certificate; it is more advantageous to obtain the lower salary W_L without the additional education required to qualify for the higher wage. (To simplify, we initially assume that training is not productive; it serves only to sort the two types.) If $W_H - C_H > W_L$ then H types will incur the cost C_H of obtaining a graduate certificate, obtaining the higher net salary $W_H - C_H$. Note that both conditions hold if $C_H < W_H - W_L < C_L$. In particular, C_H must be less than C_L . (All monetary amounts are discounted present values.) If $W_H = 1000$ and $W_L = 600$ then $C_H < 400 < C_L$ is required for signaling to reveal a worker's characteristic. This gives a range of equilibria, many of which will be inefficient because the signaling could be done at lower cost to society. This phenomenon will be encountered in the more sophisticated model to follow.

A firm will pay a huge sum for a celebrity to advertise its product, even when consumers know that the testimony has been paid for and probably does not represent the celebrity's true opinion. The point is to signal the firm's confidence in the quality of its product. The producer cannot recover its advertising expenditure if consumers find that its product is inferior and hence stop buying it. Introductory offers at very low prices have the same effect, but in that case there is a danger that the consumer will interpret the low price as evidence of low quality. We have shown that workers can pay for and receive training in equilibrium, even when that training does not enhance productivity. What's missing from the story is a discussion of employer profit maximization, which will depend in part on employer beliefs about the relationship between the amount of training attained and the worker's ability. Therefore, we now consider a more elaborate model with an explicit role for firms. The amount of training (measured in years) can have more than two values, and training can contribute directly to productivity. The richer model will also exhibit different kinds of equilibria, and equilibria with different levels of educational attainment.

6.2 A general model

There are two types of workers, *H* types and *L* types. It is common knowledge that the fraction ρ of the population is type *H*. Worker *i*(= *H* or *L*) has the utility function

$$U_i(x, y) = B_i(x) + y$$

where *x* is leisure consumption and *y* the total market value of all other goods and services. Although *x* and *y* will typically be different for *H* and *L*, we do not often use subscripts on *x* or *y*. The identity of the worker will be clear from the context. We assume that the marginal utility of leisure consumption is positive for all *x*, but if x'' > x' then the marginal utility of leisure is lower at x'' than at x'. (You can follow this section without knowing any calculus, but if you do know calculus a few shortcuts are available. We assume $B'_i(x) > 0$ and $B''_i(x) < 0$ for all *x*.)

We let *T* denote the individual's time endowment. For instance, if the basic period is a day and we measure time in hours, then T = 24. If *e* is the amount of time it takes for the individual to acquire an education, then x = T - e. This means that U_i will fall as *e* increases at an increasing rate because of the diminishing marginal utility of leisure assumption. If e'' > e' then the marginal utility of leisure is higher at e'' than at e' because leisure consumption is lower at e''. That is, the marginal cost of education is positive, and the marginal cost increases as education increases. The cost of acquiring an education will play a central role, so we simplify and write

$$U_i = w(e) - c_i(e)$$

where *w* is income, as a function of years of education *e*, and $c_i(e)$ is the amount of leisure sacrificed when *e* years of education are attained. The function w(e) is the compensation schedule posted by firms.

Example 6.1: A simple utility function

Let $U = 3\sqrt{x} + y$. That is, $B(x) = 3\sqrt{x}$. We have B(0) = 0 and B(4) = 6. Then $\Delta B = 6$, and $\Delta B/\Delta x = 6/4 = 1.5$. And B(9) = 9, so when we increase *x* from 4 to 9 we have $\Delta B/\Delta x = 3/5 = 0.6$. The marginal utility of leisure is lower at higher values of *x*. If T = 24 then c(8) = B(24) - B(16) = 14.7 - 12 = 2.7. And c(10) = B(24) - B(14) = 3.47. Consequently, $\Delta c/\Delta e = 0.76/2 = 0.38$. However, when e = 12 we have c(12) = B(24) - B(12) = 4.30. When *e* increases from 10 to 12 we have $\Delta c/\Delta e = 0.83/2 = 0.415$. The marginal cost of acquiring an education has increased.

A key assumption is that it is more costly for an L type to achieve a given education level because an L type has to put in more hours studying (over more semesters, perhaps) than an H type. Let $m_i(e)$ be the value of the marginal product of individual *i*. We can have m_i increasing with *e*, to reflect the fact that productivity increases with education. By definition, for any level of education *e*, the high-ability (H) type has a higher marginal product than the low-ability (L) type.

DEFINITION: The basic model

Type *i*'s utility function is $U_i = w(e) - c_i(e)$ where w(e) is the pay offered by the employer as a function of *e*, the level of education attained, and $c_i(e)$ is the cost of acquiring *e*. We assume that for any education level *e* we have $c_H(e) < c_L(e)$ and $m_H(e) > m_L(e)$, where $m_i(e)$ is the value of the marginal product of type *i*.

Example 6.2: Two simple cost functions

We use the following cost functions for the rest of Section 6:

$$c_H(e) = \frac{1}{2}e^2$$
 and $c_L(e) = \frac{3}{4}e^2$.

We can derive these cost functions from the function B(x). For instance, suppose that $B_H = xT - \frac{1}{2}x^2$, where *T* is the time endowment. Then

$$c_H(e) = B(T) - B(T-e) = T^2 - \frac{1}{2}T^2 - [(T-e)T - \frac{1}{2}(T-e)^2]$$

= $\frac{1}{2}e^2$.

Of course, education is more than a signal; it also enhances productivity. However, to get some quick insight we temporarily assume that each type's value of the marginal product is independent of the highest level of education reached. This assumption is dropped in Section 6.3.

Example 6.3: Education is not productive

Let $m_L = m$ and $m_H = 2m$, where *m* is some positive constant. What is the equilibrium? There are many equilibria; in fact there are two *types* of equilibria, pooling and separating, as we are about to see.

DEFINITION: The two kinds of equilibria

In a *pooling* equilibrium the two types of workers get the same amount of education, and each receives the same pay, namely the weighted average marginal product, weighted by the proportion of each type in the work force. In a *separating* equilibrium, the *H* types get more education than the *L* types, and the firms pay more to the workers with the higher level of education.

Pooling equilibria

All workers obtain the same number of years of schooling and are paid the same wage. What values of the wage and of *e* are consistent with worker utility maximization and firm profit maximization?

Recall that the proportion ρ of the population is *H* type, so the expected (or mean) productivity is $\rho \times 2m + (1 - \rho) \times m = m\rho + m$. At this point we import a classical result from the theory of competitive labor markets: The workers are paid the value of their marginal product. Therefore, in this model $m\rho + m$ is paid to each worker at equilibrium *if* the same wage is paid to all.

Pooling equilibrium wage schedule offered by each firm, based on a given critical level *g* of education:

- If *e* < *g* assume that the worker is *L* type and pay him or her *m*.
- If *e* ≥ *g*, assume that the worker is *H* type with probability *ρ* and *L* type with probability 1 − *ρ* and offer the wage *mρ* + *m*.

What we have here is an equilibrium system of *beliefs* in addition to the usual market clearance property of equilibrium. The employers' demand schedules for workers are functions of their beliefs. At equilibrium, employers' beliefs must be confirmed by observation—of the amount of output produced, which in turn is a function of the wage schedule, via worker's decisions about how much education to acquire. At equilibrium, we have a completed circle.

What decision will an *H*-type individual make when confronted with this wage schedule? There is no point in choosing e > g because that would increase training costs without bringing any increase in pay. Therefore, *H* and *L* will each set e = g if the following conditions hold for *H* and *L* respectively:

$$m\rho + m - \frac{1}{2}g^2 > m$$
 and $m\rho + m - \frac{3}{4}g^2 > m$.

Obviously, the first inequality will hold if the second does.

Example 6.4: A third of the labor force is *H* type

We continue with the basic example but explicitly examine the case $\rho = \frac{1}{3}$. Then *L* types will acquire *g* units of education if $\frac{1}{3}m + m - \frac{3}{4}g^2 > m$, or

$$g^2 < \frac{4}{9}m$$
, or $g < \frac{2}{3}\sqrt{m}$.

H types will also set e = g if this inequality holds.

Suppose that m = 9. Then we have a pooling equilibrium for any value of g satisfying $0 \le g < 2$, with a wage of 12 to each worker. Each worker will be optimizing by setting e = g and taking a wage of 12 instead of 9. (Verify: Because g < 2, U_H is greater than $12 - \frac{1}{2}(2 \times 2) = 10 > 9$ when $e_H = g$, and similarly U_L is greater than $12 - \frac{3}{4}(2 \times 2) = 9$ when $e_L = g$.) Under competitive conditions each firm pays a wage equal to 12, the expected value of the per capita marginal product. The firm's subsequent observations confirm this expectation. Everyone obtains g years of schooling, so the productivity of the group with g years of schooling will be observed to be 12 per capita. We are at equilibrium. This corresponds to a cohort of individuals with the same undergraduate degree earning the same pay in spite of their different abilities.

With full information, every worker would be required to set e = 0 for efficiency, if education makes no contribution to productivity. With asymmetric information we can have an equilibrium with each worker spending a substantial amount of time in higher education. (If average productivity per worker is 12, then even g = 1 is a substantial investment in education.) But we're not finished with the analysis. Continue to assume that education is not productive.

Separating equilibria

Suppose *H* types obtain more education in equilibrium than *L* types. Let e_H and e_L denote the equilibrium education levels of *H* types and *L* types, respectively.

Separating equilibrium wage schedule offered by each firm, based on a given critical level g of education

- If e < g assume that the worker is *L* type with probability 1 and pay *m*.
- If $e \ge g$, assume that the worker is *H* type with probability 1 and pay 2m.

What is the worker's response? Setting e > g just increases the worker's costs without any reward in terms of higher salary, so no worker will choose more than g years of schooling. Similarly, if 0 < e < g then the worker will receive the same wage as someone who sets e = 0. Therefore, regardless of type, the worker will set e = 0 or e = g. Table 5.4 displays the ingredients of a worker's decision.

A separating equilibrium must satisfy the *self-selection constraints*: Each type must find it advantageous to send a signal that is different from the one transmitted by the other type.

Table	5.4
-------	-----

е	U_H	UL
0 g	$m 2m - \frac{1}{2}g^2$	m 2 $m-{}^{3}\!/_{4}g^{2}$

DEFINITION: Self-selection (or incentive compatibility) constraints

At a separating equilibrium the outcome obtained by the H types is not preferred by the L types to the outcome they receive. Similarly, the outcome obtained by the L types is not preferred by the H types to the outcome they receive.

A separating equilibrium then implies the following incentive compatibility constraints:

$$2m - \frac{1}{2}g^2 > m$$
 and $m > 2m - \frac{3}{4}g^2$.

To have a separating equilibrium, an *H* type must get higher utility with e = g than with e = 0, and the *L* type must get higher utility with e = 0 than with e = g. The two inequalities reduce to

$$\frac{3}{4}g^2 > m > \frac{1}{2}g^2.$$

Example 6.5: The L type's value of marginal product is nine

If m = 9 then the last inequality becomes $\sqrt{12} < g < \sqrt{18}$. If g is between 3.464 and 4.243 then *H* types will set e = g, and the *L* types will set e = 0. Each *H* is paid 18 and each *L* is paid 9.

The critical *g* has to be large, to discourage *L* types from setting e = g, but not so large as to induce *H* types to forego higher education. But if ${}^{3}/_{4} g^{2} > m > {}^{1}/_{2} g^{2}$ then only *H* types will obtain higher education (i.e., will set e = g) and firms' expectations will be confirmed. Again we have a range of equilibria in which there is investment in education even though education does not enhance productivity. In a model in which education *does* contribute to productivity, one would expect to find investment beyond the point justified by considerations of productive efficiency. This is what we encounter in the next section.

6.3 When education is productive

For both the *H* types and the *L* types value added increases with *e*, but an additional unit of higher education adds more to the productivity of an *H* type than it does to the productivity of an *L* type.

Example 6.6: Marginal product is a function of education

We work out the equilibria for the case $m_H(e) = 6e$ and $m_L(e) = 3e$. We continue to assume that the cost functions are $c_H(e) = \frac{1}{2}e^2$ and $c_L(e) = \frac{3}{4}e^2$.

To establish a benchmark case, assume temporarily that the worker's type (ability) is directly observable—the full information case.

Full information equilibrium

Each worker *i* will be confronted with the wage schedule $m_i(e)$, because the worker's type is known and competition ensures that the wage will equal the value of the *individual's* marginal product. If workers' utility is not maximized at *e* then we can't be at equilibrium. Another firm will offer the worker a contract that requires the utility-maximizing value of *e*. This can be done in a way that increases the worker's utility and also the profit of that firm attempting to attract the worker.

The *H*-type worker's utility-maximizing level of *e* is obtained by solving

maximize
$$w - c = 6e - \frac{1}{2}e^2$$

The gap between the average earnings of high school and college graduates almost doubled between 1979 and 1991 (Mishel and Bernstein, 1992). Much of this increase in the rate of return to education is attributable to training in the use of computers. The proliferation of computers accounts for at least onethird, and perhaps as much as one-half, of the increase in the rate of return to education (Krueger, 1993).

То maximize this quadratic we set $e = 6/(2 \times \frac{1}{2}) = 6$. (The first derivative is 6 - e.) Because the second derivative is negative, utility maximization requires e = 6.) Then w_H , the *H* type's pay, is 36 because $m_H(6) = 36$. Similarly, at equilibrium the L-type worker's education level will maximize $w - c = 3e - \frac{3}{4}e^2$ and hence will set $e = 3/(2 \times 3/4) = 2$. (The first derivative of the objective function is 3 - 1.5e, so e = 2 at the maximum.) Then $w_L = 6$ because $m_L(2) = 3 \times 2$. All of this is displayed in Table 5.5.

Note that if we drop the full information assumption, *L* types would masquerade as *H* types because they prefer a wage of 36, even though it would $\cot \frac{3}{4} \times 6 \times 6 = 27$ to obtain the 6 years of education necessary to pass as *H* types (36 - 27 = 9), which is greater than $6 - \frac{3}{4} \times 2 \times 2 = 3)$. Therefore, the full infor-

mation outcome is not an equilibrium in an asymmetric information world. The firms could not pay a wage of 36 to *everyone* with 6 years of higher

e 5.5

e	Pay	U_H	U_L
6	36	18	9
2	6	4	3

education because, when $\rho = \frac{1}{3}$, the average value of marginal product is only $\frac{1}{3} \times 36 + \frac{2}{3} \times 18 = 24$. Therefore, we need to work out the asymmetric information equilibria.

Pooling equilibria

Both *H* and *L* types choose gyears of education, with *g* to be determined. Because education is the only observable variable that depends on ability, each worker is paid the same wage, the expected value of marginal product when everyone sets e = g. We need to work out the implications of worker utility maximization and firm profit maximization. We use the formula for maximizing a quadratic repeatedly (Section 1 of Chapter 2).

The fraction ρ of the entire population is *H* type, so the expected productivity at e = g is

$$\rho m_H(g) + (1 - \rho) m_L(g) = \frac{1}{3}(6g) + \frac{2}{3}(3g) = 4g.$$

When the wage is same for each worker then each is paid 4g at equilibrium as a consequence of competition among firms for workers.

Pooling equilibrium wage schedule offered by each firm, based on a given critical level g of education

If e < g assume that the worker is *L* type, and pay 3e.

If $e \ge g$ assume that the worker is *H* type with probability ρ , and pay 4*g*.

Will the *H* type choose the wage 4g or the wage 3*e* for some e < g? Because $4e - \frac{1}{2}e^2$ is maximized at e = 4, and hence the graph of $4e - \frac{1}{2}e^2$ is a hill with peak at 4, if $e < g \le 4$ then

$$3e - \frac{1}{2}e^2 < 4e - \frac{1}{2}e^2 < 4g - \frac{1}{2}g^2.$$

Therefore, the *H* type sets e = g and receives a wage of 4g when $g \le 4$.

Consider *L*'s decision: $4e - {}^{3}/_4e^2$ is maximized at $e = 4/(2 \times {}^{3}/_4) = 2^2/_3$ (Figure 5.8). Hence $e < g \le 2^2/_3$ implies

$$3e - \frac{3}{4}e^2 < 4e - \frac{3}{4}e^2 < 4g - \frac{3}{4}g^2$$
.

Therefore, *L* sets e = g and receives a wage of 4g when $g \le 2^{2}/_{3}$.

What about *L*'s decision at $g > 2^{2}/_{3}$? The function $3e - {}^{3}/_{4}e^{2}$ is maximized at e = 2, where $U_{L} = 3$. Consequently, $g > 2^{2}/_{3}$ and $4g - {}^{3}/_{4}g^{2} > 3$ implies that *L* will set e = g. Now, $4e - {}^{3}/_{4}e^{2} - 3 = 0$ implies g = 0.9 or 4.43 (Figure 5.8 again). Because the graph of $4e - {}^{3}/_{4}e^{2} - 3$ is a hill that reaches its peak at $g = 2^{2}/_{3}$, if $0.9 \le g \le 4.43$ we have $U_{L} \ge 3$, and hence *L* sets e = g and receives a wage of 4g. Therefore, if $g > 2({}^{2}/_{3})$ we can have a pooling equilibrium only if $g \le 4.43$.

Will *H* set e = g if $g \le 4.43$? We have already demonstrated that *H* sets e = g if $g \le 4$. Can we have $3e - \frac{1}{2}e^2 > 4g - \frac{1}{2}g^2$ if e < g and $4 \le g \le 4.43$? The maximum value of $3e - \frac{1}{2}e^2$ is 4.5, which occurs at e = 3. But for g = 4.43 we



have $4g - \frac{1}{2}g^2 = 7.9$, and that's the lowest value of $4g - \frac{1}{2}g^2$ over all g satisfying $4 \le g \le 4.43$. (See Figure 5.9: Note that $4g - \frac{1}{2}g^2 = 4.5$ if g = 1.35 or g = 6.65.) Therefore, *H* will choose g and receive a wage of 4g if $g \le 4.43$.

There is a pooling equilibrium for each value of $g \le 4.43$.

Whatever value of *g* emerges at equilibrium, we don't have the full information choices $e_L = 2$ and $e_H = 6$ that would be mandated by efficiency considerations alone.

Separating equilibria

Suppose *H* types obtain more education in equilibrium than *L* types. Could we have $e_H = 6$ and $e_L = 2$ (the full information equilibrium choices from Table 5.5) at equilibrium? No. An *L* type would prefer setting e = 6 and a having a wage of 36, on the one hand, to e = 2 with a wage of 6, on the other hand. Everyone would choose e = 6. This cannot be sustained as an equilibrium because the



average product is ${}^{1}\!/_{3} \times 36 + {}^{2}\!/_{3} \times 3 \times 6 = 24$: Cost per worker (to the firm) is 36 and revenue per worker is 24. Then what are the possible equilibrium values of e_{H} and e_{L} ?

Separation requires that the workers' choices reveal their types. Hence, a worker choosing e_L will receive a wage of $3e_L$, and a worker choosing e_H will receive a wage of $6e_H$. Both are consequences of competition among producers for workers. The equilibrium will then have to satisfy the self-selection constraints—also called the incentive compatibility constraints. The following are the self-selection conditions for *H* and *L* respectively:

$$6e_H - \frac{1}{2}e_H^2 \ge 3e_L - \frac{1}{2}e_L^2,$$
[1]

$$3e_L - {}^3\!/_4 e_L^2 \ge 6e_H - {}^3\!/_4 e_H^2.$$
^[2]

Statement [1] says that an *H* type prefers obtaining e_H years of training and a wage of $6e_H$ to a wage of $3e_L$ with e_L years of training. And [2] says that *L* prefers obtaining e_L years of training and a subsequent wage of $3e_L$ to e_H years of training with a wage of $6e_H$. In short, if [1] and [2] hold then no worker will have an incentive to conceal his or her true type. We already know that the maximum value of the left-hand side of [2] is 3, occurring when $e_L = 2$.

If *L*-types do not get individual utility of at least 3 at equilibrium they could set up their own firm. They could each set $e_L = 2$ in this new firm and pay themselves a wage of 6, yielding $U_L = 3$. If they received applications from *H*-types wanting to work for a wage of 6 the *L*-type owners of the firm would gladly welcome them aboard, realizing a profit of $6 \times 2 - 6 = 6$ per *H*-type worker. (This argument will not work for *H* types. If they form their own firm they will face the same problem as existing employers: *L* types will attempt to masquerade as high-productivity workers. In that case, a wage that would be viable if the firm were staffed by *H* types alone would not be viable if *L* types joined the firm and received the same wage.) Therefore, $U_L \ge 3$ at a separating equilibrium. The only way that an employer could provide $U_L \ge 3$ would be to offer a wage of 6 and insist on 2 years of training. (That's because $3e_L - {}^{3}_{I_A} e_L^2$ is maximized at $e_L = 2$.) Therefore, $e_L = 2$ at the equilibrium, which must satisfy

$$6e_H - \frac{1}{2}e_H^2 \ge 3 \times 2 - \frac{1}{2} \times 2 \times 2 = 4$$
 and $3 \ge 6e_H - \frac{3}{4}e_H^2$. [3]

Now, maximize U_H subject to [3]. Consider the second part of [3]. The function $6e_H - {}^{3}\!/_4 e_H^2$ is maximized at $e_H = 4$. Starting at $e_H = 4$, the value of the function decreases as e_H increases *or* decreases (Figure 5.10). Now, $e_H = 0.54$ and $e_H = 7.46$ are the solutions to the equation $6e_H - {}^{3}\!/_4 e_H^2 = 3$. Therefore, the second part of [3] will be violated if $0.54 < e_H < 7.46$, and it will be satisfied otherwise. Therefore, [3] can be replaced by

$$6e_H - \frac{1}{2}e_H^2 \ge 4$$
 and either $e_H \le 0.54$ or $e_H \ge 7.46$. [4]

The function $6e_H - \frac{1}{2}e_H^2$ is maximized at $e_H = 6$, and so U_H falls as e_H increases beyond 7.46 or falls below 0.54. Therefore, if we maximize U_H subject to [4] we will





have either $e_H = 0.54$ or $e_H = 7.46$. It is easy to see that 7.46 gives the higher value of U_H . (Besides, $0.54 < e_L = 2$.) We get $U_H = 6 \times 7.46 - \frac{1}{2} \times (7.46)^2 = 16.9$, so [4] is satisfied. We have a separating equilibrium.

At a separating equilibrium we have $e_H = 7.46$ and $e_L = 2$. Firms pay a wage of 3e if e < 7.46 and pay 6e if $e \ge 7.46$. *H* types receive a wage of 44.76, and *L* types receive a wage of 6.

Confirm that we have an equilibrium by computing the worker's response: Consider an *L* type's decision. Because $6e - {}^{3}\!/_4 e^2$ is maximized at $e = 6/(2 \times {}^{3}\!/_4) = 4$, when e > 4 the value of the function declines as *e* increases. Therefore, an *L* type would never set e > 7.46. Similarly, $6e - {}^{1}\!/_2 e^2$ is maximized at $e = 6/(2 \times {}^{1}\!/_2) = 6$, so that function decreases as *e* increases, if e > 6 initially. Consequently, an *H* type would not set e > 7.46. Therefore, an *L* type's decision reduces to a choice between $e_L = 2$ (which provides the maximum utility available with the wage schedule 3e) and $e_L = 7.46$. Now, $U_L(7.46) = 6 \times 7.46 - {}^{3}\!/_4(7.46)^2 = 3$, which does not exceed the utility of 3 realized by *L* when $e_L = 2$ and the wage is 3×2 .

An *H* type will set $e_H = 7.46$ because $U_H = 6 \times 7.46 - \frac{1}{2}(7.46)^2 = 16.9$, which is larger than $3 \times 3 - \frac{1}{2}(3)^2 = 4.5$, the highest level of utility attainable by *H* with the schedule 3e.

We have confirmed that each *H* type maximizes utility by setting $e_H = 7.46$, and each *L* type maximizes utility by setting $e_L = 2$. If no firm wants to depart

from the wage schedule "pay 3*e* if e < 7.46 and 6*e* if $e \ge 7.46$ " then we are indeed at equilibrium. Each firm expects that a worker presenting a certificate for 2 years of training is an *L* type and that each worker with a certificate for 7.46 years of training is an *H* type. If the firm hires n_L of the former and n_H of the latter it will expect its output to be $3 \times 2 \times n_L + 6 \times 7.46 \times n_H$ and of course that is exactly what it will be. Employers' expectations are confirmed, and they have no reason to modify the wage schedule.

Note that *L* types make the same investment in education as in the full information efficient outcome, but *H* types invest more than they would in full information. Surprisingly, the proportion ρ of *H* types plays no role in the computation of a separating equilibrium. Asymmetric information results in more investment in education than can be justified by considerations of the return to society from enhanced productivity, even if the fraction of *L* types is small. (Given that workers in industrialized economies do not have a uniform educational background we can conclude that the separating equilibrium is the applicable one.)

Sources

Spence (1973) was the first to show how signaling could emerge as a solution to the asymmetric information (hidden characteristic) problem introduced into the literature by Akerlof (1970). In separate contributions, George Akerlof, Michael Spence, and Joseph Stiglitz showed that the presence of asymmetric information in real-world markets required a new way of modeling economic exchange. They were awarded the Nobel Prize in Economics for the year 2001.

Links

Cameron and Heckman (1993) find that workers who enter the labor market with a high school equivalency degree are paid 10% less on average than workers who enter with a conventional high school diploma. Riley (2000) and Chapter 13 of McAfee (2002) give an overview of the economics of signaling. The latter is decidedly nontechnical, and the former is intended for readers with a good economics background. Riley (1989) fits between the two.

Problem set

- **1.** Use the technique of Example 6.2 to derive the cost function $c_L(e) = \frac{3}{4}e^2$ from a utility function of the form $U_L = B_L(x) + y$.
- **2.** For the model of Section 6.3, show explicitly that there is a pooling equilibrium for g = 1.5 and also for g = 3.5.
- **3.** There are two types of workers, *H* and *L*. The value of the marginal product of an *H* type is 30*e* and the value of the marginal product of an *L* type is 12*e*, where *e* is the level of education attained. One-third of the workers are *H* types, but an employer cannot directly distinguish an *H* from an *L*. The cost

7. Competitive Insurance Markets

to an *H* type of *e* units of education is $\frac{1}{2}e^2$, and the cost to an *L* type of *e* units of education is e^2 .

- A. Find a separating equilibrium, and characterize it.
- **B.** For what range of values of g would a pooling equilibrium exist? Describe a pooling equilibrium.
- **4.** There are two types of workers, *H* and *L*. The value of the marginal product of an *H* type is 25*e*, and the value of the marginal product of an *L* type is 10*e*. Half of the workers are *H* types, but an employer cannot directly distinguish an *H* from an *L*. The cost to an *H* type of *e* units of education is $0.6e^2$, and the cost to an *L* type of *e* units of education is $0.8e^2$.
 - A. Find a separating equilibrium, and characterize it.
 - **B.** For what range of values of *g* would a pooling equilibrium exist? Describe a pooling equilibrium.
- **5.** When will the producer of a high-quality product be able to use a warranty to signal that its output is superior to that of its low-quality rival even though consumers cannot directly determine quality?

7 COMPETITIVE INSURANCE MARKETS

This section highlights the difficulties of eliciting hidden information when agents differ with respect to their information about the likelihood of events. The hidden characteristic in this case is the probability that an individual will suffer a mishap—have a car accident, be burglarized, be hospitalized, and so forth. We assume that the individual knows the probability of this happening, but no one else does. Moreover, this person cannot be expected to willingly disclose this hidden characteristic, especially if those who report a higher probability of accident are charged higher insurance premiums. We embed these facts in a model of a competitive insurance market in a mature capitalist economy. We see that there exists no competitive equilibrium for some values of the parameters. When an equilibrium *does* exist, individuals will reveal their accident probability by their choice of insurance contract. Nevertheless, the competitive equilibrium may not be efficient when it does exist, although it is not easy to determine when government regulation can improve on the market outcome.

7.1 The model

Christopher knows more about the likelihood of his having an accident in a certain situation than others do. This means that some of the information about the probability of an accident is hidden from the company offering him an insurance contract. In the case of automobile insurance, there is a lot of information about our driving habits that is available to insurance companies. Young men are more likely to be risky drivers than young women, and this observation is used by companies in determining rates. The correlation is far from perfect, however, but to the extent that a riskier driver is identified and charged a higher premium

we can say that prices reflect costs to society—part of the cost in this case is the possibility of you or I being the dangerous driver's victim. Information about

I was driving home one day, correcting the manuscript for this section, trying to think of a good example of private information in this context. Then I realized that my practice of correcting manuscripts while driving is the perfect illustration. My insurance company would love to know about this dangerous habit. speeding tickets and prior accidents are also used in determining automobile insurance premiums. This is called experience rating.

Suppose that an insurance company has already used all available information to categorize drivers by risk and has determined that Christopher and Laurie are in the same risk category. There is additional private information that Christopher has about his driving habits and Laurie has about hers that is hidden from the insurance provider. Given the evi-

dence available to insurance companies, there will still be differences in risk that cannot be directly observed, and it is these additional, private characteristics that are the subject of this section.

To abstract from most of the other issues, assume that there is only one basic commodity, which we call wealth. Uncertainty concerns the status of an individual's wealth, which is either partly destroyed—by accident or fire, say—or remains intact. We let *a* represent the value of an individual's wealth after an accident when no insurance is purchased and let *z* represent the value of the same individual's wealth without insurance and also without an accident. The actual wealth level may be different from both *a* and *z*. If the individual buys insurance but there is no accident, then his or her wealth will be lower than *z* by the amount of the premium. And if the individual has an accident after buying insurance then his or her wealth will be higher than *a* because he or she will receive a claim check from the insurance company to partly compensate for the loss z - a.

Let *x* denote the amount of wealth available to finance consumption when the individual suffers an accident, and let *y* denote the amount of wealth available for consumption when there is no accident. An insurance contract or policy, *P*, requires the purchaser to pay a stipulated fee (the premium) of *f* dollars before the resolution of uncertainty. If the person does not have an accident then no further exchange takes place, but if he or she is involved in an accident then the insurance company pays *c* dollars net. The amount of the claim check is actually c + f, but the net claim is *c* because the individual must pay the annual premium whether he or she has an accident or not. Therefore, if the individual purchases a policy charging a premium *f* and paying the net claim *c* we have

$$x = a + c$$
 and $y = z - f$.

We allow f = 0 = c, which represents the case when no insurance is purchased.

The number π is the probability that the individual has an accident (or a fire in his or her house, etc.). To make the model really simple, we assume there are only two possible values of π and hence only two types of individuals: low risk, *L*, and high risk, *H*, so we write π_L and π_H , respectively. The hidden characteristic

7. Competitive Insurance Markets

is the value of π . Of course, $\pi_H > \pi_L$, $1 - \pi_L$ is the probability that an *L* type does not have an accident, and $1 - \pi_H$ is the probability that an *H* type does not have an accident. (The event "Christopher has an accident" is statistically independent of the event "Laurie has an accident." Don't jump to the conclusion that they have accidents simultaneously.)

For analytical convenience all individuals are assumed to have the same utility-of-wealth function U(w). They are risk averse: The marginal utility of wealth is positive, but it diminishes as wealth increases. (In terms of calculus, U'(w) > 0 and U''(w) < 0, for all w > 0.) Each individual has the same endowment allocation: *a* if there is an accident, and *z* if there is no accident. And z > a because an accident destroys wealth. We are assuming that individuals are identical, except for the value of π . There are n_L low-risk individuals and n_H high-risk individuals. Although U(w) is the same for each individual, it is not the case that expected utility is the same for an *L* type and an *H* type because the probabilities differ. Expected utility for *L* and *H* respectively is

$$u_L(x, y) = \pi_L U(x) + (1 - \pi_L)U(y),$$

$$u_H(x, y) = \pi_H U(x) + (1 - \pi_H)U(y).$$

We assume that *U* is monotonic: If w > w' then U(w) > U(w'). Therefore, u_L will increase if *x* increases and *y* does not decrease, or if *y* increases and *x* does not decrease. The same can be said of u_H . (The utility derived from leisure is not needed for this model because we assume that there is no opportunity for the individual to affect the probability of an accident by sacrificing leisure to devote effort to preventive care.)

Risk aversion implies diminishing marginal rate of substitution (MRS) for the indifference curve "expected utility = a constant"

To confirm that the indifference curve has the shape of the one in Figure 5.11, note that the MRS at (x, y) for type i (i = L or H) is

$$\frac{\pi_i U'(x)}{(1-\pi_i)U'(y)}$$

where U'(x) and U'(y) denote the marginal utility of wealth at *x* and *y* respectively. As we move down the indifference curve, increasing *x* and decreasing *y*, U'(x) will fall and U'(y) will increase because marginal utility diminishes with wealth as a consequence of risk aversion. Therefore, we have diminishing MRS.

The crucial assumption is that insurance companies cannot identify *L* types and *H* types directly. They can distinguish them only by observing their choices, and then only if the *H* types have no incentive to purchase the contract designed for the *L* types and vice versa. Consequently, even within our very abstract framework, in which individuals are identical in all but one respect and only wealth available for consumption affects well-being, there are potential difficulties in terms of the satisfactory performance of competitive markets.



Figure 5.11

DEFINITION: The individual variables and parameters

Without insurance, an individual's wealth is *a* if he or she has an accident and *z* otherwise. With insurance, the individual's wealth is *x* if he or she has an accident and *y* otherwise. If the insurance premium is *f* then y = z - f, and x = a - f + v, where *v* is the value of the claim check. The number of *L* types is n_L , and the number of *H* types is n_H . The probabilities of an accident for *L* types and *H* types are, respectively, π_L and π_H .

Assume that n_L and n_H are both large in absolute value, although one number might be small relative to the other. If n_L is large, then the total number of accidents suffered by *L* types will be close to the expected number, $\pi_L n_L$.

The *law of large numbers*: If *n* statistically independent experiments are performed, and π is the probability of success in a single experiment, then for any positive number ϵ however small, there is a value of *n* sufficiently large so that the probability is greater than $1 - \epsilon$ that the average number of successes in *n* trials will be arbitrarily close to π . If you haven't encountered this fundamental law, convince yourself of its truth by tossing a coin. With a large number of tosses the fraction of heads will be very close to the expected number, one-half. I can safely make this claim, because the probability of it being contradicted is *very* small if the number of tosses is *very* large.

7. Competitive Insurance Markets

We assume that the total number of accidents is always exactly equal to the expected number so we don't have to keep saying that the results hold only approximately but with high probability.

DEFINITION: The number of accidents

We assume that n_H is large, and use the law of large numbers to assume that the total number of accidents suffered by H types is exactly $\pi_H n_H$. Similarly, we assume that n_L is large, and that the total number of accidents suffered by L types is exactly $\pi_L n_L$.

Let's also assume, for simplicity, that administration costs are constant and relatively small. This permits us to assume zero administrative costs without affecting the results. In that case, competition among insurance companies will ensure that the value of premiums taken in at equilibrium equals the value of gross claims paid out at equilibrium. (This is derived with care in Section 7.4 of Chapter 2.)

7.2 The number of contracts in equilibrium

H types are identical to each other, so they will make the same choices. Suppose not, and some *H* types choose policy P^1 and some choose P^2 , different from P^1 . If one yields a higher value of u_H than the other, we can't be at equilibrium. Some *H* types will switch from the low-utility policy to the high-utility policy. Therefore $u_H(P^1) = u_H(P^2)$ at equilibrium. Further, P^1 and P^2 yield the same expected profit to the insurance company. If, say, P^1 provided higher expected profit per policy than P^2 then insurance companies would all offer P^1 in preference to P^2 . Their customers wouldn't object because $u_H(P^1) = u_H(P^2)$.

To summarize, if there are two different policies P^1 and P^2 purchased by H types in equilibrium then they yield the same utility and the same profit. Suppose $P^1 \neq P^2$, and they charge premiums of f^1 and f^2 and pay net claims of c^1 and c^2 , respectively. The insurance company's per capita profit from P^1 is $f^1 - \pi_H \times (c^1 + f^1)$. (The policy brings in a premium revenue of f^1 with certainty and disburses a claim check of $c^1 + f^1$ with probability π_H .) Similarly, the per capita profit from P^2 is $f^2 - \pi_H \times (c^2 + f^2)$. Then $f^1 - \pi_H \times (c^1 + f^1) = f^2 - \pi_H \times (c^2 + f^2)$. Consider policy P^3 constructed by averaging P^1 and P^2 . That is, P^3 charges a premium $f^3 = \frac{1}{2}f^1 + \frac{1}{2}f^2$ and pays a net claim of $c^3 = \frac{1}{2}c^1 + \frac{1}{2}f^2 - \frac{1}{2}\pi_H(c^1 + f^1) - \frac{1}{2}\pi_H(c^2 + f^2) = \frac{1}{2}[f^1 - \pi_H(c^1 + f^1)] + \frac{1}{2}[f^2 - \pi_H(c^2 + f^2)]$, the average of the expected profits from P^1 and P^2 .

At equilibrium, P^1 and P^2 yield the same per capita expected profit. Then P^3 must generate the same profit as P^1 and P^2 because the profit from P^3 is the average of the profit from P^1 and P^2 . Then all three policies yield the same per capita profit. But P^3 affords higher expected utility! A glance at Figure 5.11 makes this evident. P^3 is on the straight line between P^1 and P^2 , which are on the same indifference curve, and hence P^3 is on a higher indifference curve. Now, P^3 will yield the same profit as P^1 and P^2 but will afford more utility. Therefore, the

insurance company could modify P^3 slightly, raising the premium and providing the same claim, c^3 . This new policy will certainly bring in more profit than P^1 or P^2 and H types will prefer it to either P^1 or P^2 —as long as the increase in premium is not too large. Both insurance companies and their clients will prefer this new outcome to the one in which only P^1 and P^2 were available, so the original situation cannot be an equilibrium. We have proved that there is only one contract offered to the H types in equilibrium. Obviously, the same argument will establish that the L types will all choose the same contract at equilibrium.

At equilibrium, a contract will have n_H , n_L , or $n_H + n_L$ buyers.

Now, let π denote the probability that an individual randomly selected from the entire population has an accident. Then π is the expected number of accidents divided by the total population. That is,

$$\pi = \frac{\pi_H \times n_H + \pi_L \times n_L}{n_H + n_L}$$

Recall that x = a + c and y = z - f when an individual purchases policy *P* with premium *f* and net claim *c*. Because the individual ultimately cares only about *x* and *y*, we think of an insurance contract as a specification of *x* and *y*. If we need to, we can recover the premium and net claim by setting f = z - y and c = x - a.

DEFINITION: An insurance policy

An insurance policy specifies the values of *x* and *y*. Given those values we can recover the actual premium *f* and net claim *c*, because f = z - y and c = x - a.

The condition that all money taken in from a contract is paid out in the form of claim checks is

$$n(z-y) = \rho n(x-a+z-y)$$
^[5]

where $n = n_H$ (in which case $\rho = \pi_H$), or $n = n_L$ and $\rho = \pi_L$, or $n = n_H + n_L$ and $\rho = \pi$. The term on the left-hand side of equation [5] is the total amount of money collected in premiums from policyholders, and the term on the right is the amount of the claim check (the net claim *x*—*a* plus the premium *z*—*y*) sent to each policyholder having an accident multiplied by ρn , the number of accidents. We can divide both sides of [5] by *n* and rewrite the zero-profit conditions as

$$\pi_L x + (1 - \pi_L) y = \pi_L a + (1 - \pi_L) z,$$
[6]

$$\pi_H x + (1 - \pi_H) y = \pi_H a + (1 - \pi_H) z,$$
[7]

$$\pi x + (1 - \pi)y = \pi a + (1 - \pi)z.$$
 [8]





Equation [6] is the zero-profit condition for a group composed exclusively of *L* types. It equates the expected market value (EMV) of wealth with insurance to the EMV of wealth without insurance. Why bother buying insurance then? Because (x, y) delivers higher expected utility than (a, z). Equation [7] is the zero-profit condition for a group composed exclusively of *H* types, and [8] is the zero-profit condition when the two types buy the same policy. Consider Figure 5.12. If the *L* types had a contract that left them above line [6] then it would not generate enough premium income to pay all the claims. If they had a contract below line [6] then it would yield a positive profit because it takes in more premium income than required to honor the claims, and that is not consistent with equilibrium. The analogous statement holds for [7] with respect to *H*, of course.

If $\rho x + (1 - \rho)y > \rho a + (1 - \rho)z$ then (x, y) is not feasible. If $\rho x + (1 - \rho)y < \rho a + (1 - \rho)z$ then (x, y) is not consistent with equilibrium. Therefore, $\rho x + (1 - \rho)y = \rho a + (1 - \rho)z$ at equilibrium $(n = n_H \text{ and } \rho = \pi_H, \text{ or } n = n_L \text{ and } \rho = \pi_L$, or $n = n_H + n_L$ and $\rho = \pi$).

∂7.3 Full information equilibrium

To establish a benchmark case, suppose that insurance companies *are* able to distinguish H types from L types. Assume, for instance, that all can

be relied on to answer truthfully when asked to which risk group they belong.

There will be no cross-subsidization at equilibrium: All claims made by H types are paid out of premiums contributed by H types, and all claims made by L types are paid out of premiums contributed by L types. Here's why: We can't have L above [6] and H above [7] in Figure 5.12 because neither policies would collect enough revenue to pay the claims presented.

What about a cross subsidy? Could the L types wind up below [6] with the Hs above [7]? This means that the policy obtained by L types collects more premium income than is required to finance claims by L types. The surplus is used to finance the deficit from the H policy. But this is inconsistent with equilibrium. A company could offer a policy to Ls that cut the surplus in half. This would be preferred by the Ls, and it would be profitable for the company offering it. (The company would not have to worry that H types would buy it as well. We are temporarily assuming that insurance companies know who the Hs are: They would not be allowed to purchase the contract designed for the Ls.) The original policy would quickly be driven off the market. A similar argument will show that the H types will not subsidize the L types at equilibrium. Therefore, there is no cross subsidy at equilibrium.

Because there is no cross-subsidization, H types and L types will buy different contracts at equilibrium because equation [6] must hold for P^L , the contract obtained by L types at equilibrium, and [7] must hold for the H types' contract P^H . If the left-hand side of equation [6] exceeds the right-side, the contract is not feasible because the value of gross claims paid out will exceed the value of premiums collected. If the right-hand side exceeds the left-hand side, then insurance companies are earning excess profits and competition will force premiums to fall, and hence the original state was not in equilibrium. Similarly for [7]. Therefore, we can apply the complete insurance theorem of Section 7.4 of Chapter 2 to each risk type: The policy P^L obtained by the L types solves [6] and x = y. The H types will get the policy P^H that solves [7] and x = y, as shown in Figure 5.12.

At the full information equilibrium the *L* types obtain $x = y = \pi_L a + (1 - \pi_L)z$, and the *H* types get $x = y = \pi_H a + (1 - \pi_H)z$.

The full information equilibrium is efficient: As we demonstrated in Section 7.5 of Chapter 2, the outcome $x = y = \pi_L a + (1 - \pi_L)z$ would be efficient if the economy consisted only of *L* types. And the outcome $x = y = \pi_H a + (1 - \pi_H)z$ would be efficient if the economy consisted only of *H* types. Therefore, we can't increase anyone's expected utility without harming someone else if we are limited to rearrangements within the *L* group or within the *H* group, or both. If we shifted some wealth from one group to another we would obviously have to reduce the expected utility of someone within the former group. Hence, it is not possible to increase one person's expected utility without diminishing someone else's.
Example 7.1: Full information equilibrium in a special case

The probability that *L* has an accident is ${}^{1}\!/_{4}$, and ${}^{1}\!/_{2}$ is the probability that *H* has an accident. Then ${}^{3}\!/_{4}$ is the probability that an *L* type does not have an accident, and ${}^{1}\!/_{2}$ is the probability that an *H* type does not have an accident. Each individual has the utility-of-wealth function

$$U(w) = \ln(w+1).$$

Then $U'(w) = (w+1)^{-1}$, and hence $U''(w) = -(w+1)^{-2} < 0$ for all $w \ge 0$, so the individuals are risk averse. Each person's endowment is a = 4 if there is an accident and z = 12 when there is no accident. The relevant expected utility functions are

$$u_L(x, y) = \frac{1}{4} \ln (x+1) + \frac{3}{4} \ln (y+1),$$

$$u_H(x, y) = \frac{1}{2} \ln (x+1) + \frac{1}{2} \ln (y+1).$$

Let's determine the basket $P^L = (x, y)$ that would be chosen by *L* types if the amount of money they paid into insurance companies in premiums were paid out in claims. The premium per capita is 12 - y and the claim check per capita is x - 4 + 12 - y, the net claim plus the premium. Assuming that n_L is large, the total number of accidents suffered by *L* types will be close to the expected number, $\frac{1}{4}n_L$. For convenience, assume that it is exactly equal to $\frac{1}{4}n_L$. Then we have

$$n_L \times (12 - y) = \frac{1}{4}n_L \times (x - 4 + 12 - y), \text{ or } \frac{1}{4}x + \frac{3}{4}y = 10,$$
 [9]

after dividing both sides by n_L and rearranging. Note that equation [9] says that the EMV of consumption with insurance equals the EMV of wealth without insurance, namely ${}^{1}\!/_{4} \times 4 + {}^{3}\!/_{4} \times 12$. That is what we would expect to see if all premium money received by insurance carriers were paid out as claims. Now, to find P^L we maximize u_L subject to [9]. (If you prefer a shortcut, use the complete insurance theorem of Section 7.4 of Chapter 2.) We have y = 40/3 - x/3 from [9], and substituting this into u_L yields

$$V(x) = \frac{1}{4}\ln(x+1) + \frac{3}{4}\ln(40/3 - x/3 + 1),$$

which we want to maximize. $V'(x) = \frac{1}{4}(x+1)^{-1} + \frac{3}{4}(y+1)^{-1} \times -\frac{1}{3} = \frac{1}{4}(x+1)^{-1} - \frac{1}{4}(y+1)^{-1}$. Note that V'' < 0 at all points. Therefore, if V'(x) = 0 yields nonnegative values of *x* and *y* the equation V'(x) = 0 will characterize the solution to our problem. But V'(x) = 0 implies x = y, and substituting this into [9] yields x = y = 10. Therefore, $P^L = (10, 10)$, and $u_L(P^L) = 2.40$. This means that, subject to constraint [9], *L* would want a policy with a premium of 2 = 12 - 10 and a net claim of 6 = 10 - 4 in case of an accident. Note that the value of the claim check is \$8 when *L* suffers an accident, but *L* still has to pay the premium in a year when *L* makes a claim, so the net addition to consumption is \$6.

Similarly, to find P^H , the choice of H types when the amount of money paid in as premiums by H types is paid out in claims to H types, we maximize u_H subject to

$$\frac{1}{2}x + \frac{1}{2}y = \frac{1}{2} \times 4 + \frac{1}{2} \times 12 = 8.$$
 [10]

We know we will have x = y (the individuals are risk averse), in which case [10] implies x = y = 8. Therefore, $P^H = (8, 8)$, with $u_H(P^H) = 2.20$. The premium per capita is 12 - 8 = 4 and the net payment in case of accident is 8 - 4 = 4. Note that in the full information world, the *H* types pay a higher premium and get less coverage than *L* types.

Even though *H*'s net claim equals the premium, it would not be true to say that the insurance contract provides nothing of value. The expected utility of an *H* type without insurance is $u_H(4, 12) = 2.087$, and expected utility with insurance is $u_H(8, 8) = 2.20$, which is significantly higher.

Life insurance policies have a suicide clause that exempts the insurance company from its obligation to pay off if death is self-inflicted and occurs 365 days or less from the date of issue of the policy. (Some contracts have a two-year limit.) The suicide rate is lowest in the twelfth month of the life of the policy and highest in the thirteenth month-the twentyfourth and twenty-fifth months, respectively, for policies with a two-year limit (Milgrom and Roberts, 1992, p. 178). Evidently, some people insure their lives knowing that suicide is a serious possibility. (There is undoubtedly a moral hazard element as well. Some insured individuals commit suicide in the thirteenth month who would not end their lives at all if their heirs weren't going to collect on a life insurance policy.)

The competitive equilibrium is efficient if all individuals disclose their risk category truthfully. A problem arises only when there are two risk categories and the insurer does not know to which group a client belongs. This is the situation that insurers actually face. H types have no incentive to reveal their true characteristic because they prefer the policy P^L intended for L types to policy P^H . The former provides more of each good: The H types will have much more wealth if they masquerade as L types. Everyone will declare herself to be in risk category L and will purchase P^L . This outcome is not feasible, however, because P^L yields zero profit only when H types are excluded. If P^L is purchased by some H types, who file more claims per dollar of premium than L types, there will not be enough premium income to honor each claim.

∂7.4 Asymmetric information equilibrium

Assume from now on that high-risk individuals will not directly reveal their identity and only individual *i* knows *i*'s risk parameter π_i . There are only two possible equilibria, one in which everyone has the same policy and one in which the different types make different decisions.

DEFINITION: Pooling and separating equilibria

A pooling equilibrium is one in which the same contract is obtained by both risk categories.

7. Competitive Insurance Markets

A separating equilibrium is one in which the H types are separated out by providing the L types with a contract that less desirable to the H types than the contract designed for them.

The constraint that incorporates feasibility and the zero-profit condition for a pooling equilibrium is [8] from Section 7.2. Recall that π , the probability that an individual randomly selected from the entire population does have an accident, is given by

$$\pi = \frac{\pi_H \times n_H + \pi_L \times n_L}{n_H + n_L}.$$
[11]

Let MRS_H and MRS_L denote the MRS of an H type and an L type, respectively:

$$MRS_{H} = \frac{\pi_{H}U'(x)}{(1 - \pi_{H})U'(y)},$$
$$MRS_{L} = \frac{\pi_{L}U'(x)}{(1 - \pi_{L})U'(y)}.$$

Now, $MRS_L < MRS_H$ at any point because $\pi_L < \pi_H$ and $1 - \pi_L > 1 - \pi_H$ (Figure 5.13). (We're computing the MRS at the same values of *x* and *y* for the two individuals.) Intuitively, the high-risk types are willing to sacrifice more *y* (consumption in the "no accident" state) to get an additional unit of *x* (consumption in case of an accident) because they have a higher probability of an accident.



Figure 5.13

314 Hidden Characteristics

Because $MRS_L < MRS_H$, at the exchange rate $\xi = \frac{1}{2}MRS_L + \frac{1}{2}MRS_H$ there is a number δ small enough in absolute value so that

 $u_L(x - \delta, y + \xi \delta) > u_L(x, y)$ and $u_H(x - \delta, y + \xi \delta) < u_H(x, y)$.

Suppose that we claim to have a *pooling equilibrium* with each consumer, regardless of type, obtaining the basket (x, y). If an insurance company offered a different contract *P* giving rise to $(x - \delta, y + \xi \delta)$ it would be preferred to (x, y) by *L* types but not by *H* types (Figure 5.13). The insurance company could offer *P* and be sure that *L* types would purchase it in preference to (x, y) but that *H* types would not. Even though the company would not be able to distinguish an *L*-type individual from an *H* type, by judicious contract design a company could rely on the *H* types to reveal themselves by their choice.

An insurance company that catered exclusively to *L*-type risks would have to pay claims to the fraction π_L of its policyholders. Before *P* was available, both types purchased (*x*, *y*) and the fraction π of policyholders filed claims. If $\delta > 0$ is small, then (*x*, *y*) and *P* are almost the same, but $\pi_L < \pi_H$, and a company offering *P* will pay a smaller fraction of its premium receipts in claims, though the premium and net claim per person will be almost the same as for (*x*, *y*). Therefore, *P* will yield more profit to the insurance companies offering it, and this means that the original situation in which each person purchased (*x*, *y*) is not in equilibrium. Companies would have incentive to offer a new contract *P*, and it would be more profitable if the individuals who preferred it in preference to (*x*, *y*) purchased it. (We still won't have equilibrium when *P* is introduced because the viability of (*x*, *y*) depends, through equation [8], on its being purchased by *L* types as well as by *H* types, but the former will defect to *P* as soon as it is offered. The companies that continue to offer (*x*, *y*) will take a loss, and that is not consistent with equilibrium.)

What we have discovered is that an equilibrium must separate *H* types from *L* types by offering different contracts such that

- · neither risk type would want to buy the contract intended for the other,
- the contract designed for *L* types satisfies equation [6], and
- the contract designed for *H* types satisfies equation [7].

There must be separate contracts at equilibrium. And individuals of the same risk type will buy the same contract. Therefore, only two contracts will be offered at equilibrium because there are two risk categories.

If the contract B^L available to L types is below the line representing equation [6], then the L types are subsidizing the H types. But this is inconsistent with equilibrium, because a new firm could enter and obtain a positive profit by offering a contract close to B^L but below the line representing [6]. This could be done in such a way that L types prefer the new contract and H types still prefer their original choice. Therefore, equation [6] must be satisfied at equilibrium by the contract designed for L types. (The new contract might not be more profitable than B^L , but it would provide a positive economic profit and hence would be offered by a new entrant to the industry.)





If *H* types wind up below the line representing [7] an entrepreneur could enter the market and offer a contract that yields higher *x* and higher *y*, but still yielding the entrant a positive profit—that is, the new value of *x* and *y* would also be below [7]. It would be profitable even if it were purchased only by *H* types. Therefore, at equilibrium, the contract purchased by *H* types must satisfy [7] and, as the previous paragraph established, the contract purchased by *L* types must satisfy [6]. And each type must have an *incentive* to purchase the contract designed for it by the insurance company. This *incentive compatibility*, or *self-selection* condition, is a constraint on a firm's profit-maximization calculation: If the firm makes high profit when the *H* types purchase policy B^H but the *H* types get higher expected utility from the policy B^L intended for the *L* types, then the firm has not maximized profit because the *H* types won't buy B^H .

DEFINITION: Self-selection constraints

If *L* types purchase B^L at equilibrium, and *H* types purchase B^H , then $u_L(B^L) \ge u_L(B^H)$ and $u_H(B^H) \ge u_H(B^L)$.

Conditions [6] and [7] are represented geometrically as straight lines in Figure 5.14. Because (a, z) satisfies both equations, it is on both lines. Because the

316 Hidden Characteristics

individuals are risk averse, we need not be concerned with any (x, y) for which x < a. (Why?) Is there any reason why the *H* types can't have their most-preferred contract subject to [7]? No. If any other contract B^H were offered, a company would have incentive to offer the most-preferred contract P^H consistent with [7], and this would be purchased by *H* types in preference to B^H . Both would give rise to the same profit (zero), but because P_H is preferred, a company could actually raise the premium slightly and make more profit (even if purchased only by *H* types) than with B^H , and *H* types would still prefer the new contract to B^H . Therefore, P^H , depicted in Figure 5.14, is offered to *H* types and chosen by them at equilibrium. $B^H = P^H$, in short.

Now, P^H imposes a constraint on B^L , the contract offered to L types. B^L must satisfy

$$u_H(P^H) \ge u_H(B^L),\tag{12}$$

the self-selection constraint for an H type. At equilibrium B^L will maximize u_L subject to conditions [6] and [12]. This means that the L types will not be offered their most-preferred contract P^L subject to [6], because that would be preferred by H types to P^H , and P^L is feasible if and only if it is purchased exclusively by L types. Note that the equilibrium pair of contracts is determined independently of n_H and n_L , the number of H types and L types respectively.

The asymmetric information equilibrium provides the H types with the contract P^H that they would obtain in a full information world, but the L types get less utility than they would in the full information equilibrium. The L types get their most-preferred contract subject to the zero-profit condition and the constraint that it is not preferred to P^H by the H types.

The contract B^L offered to L types at equilibrium is depicted in Figure 5.14. The H types are exactly as well off as they would be if the L types did not exist, but the L types are worse off as result of the presence of individuals with a higher accident probability! Without the H types the L types would have P^L at equilibrium, but as it is, they wind up with B^L . The existence of a tiny group of H types can have a strong negative impact on the welfare of the L types, but the loss in welfare to the latter is not balanced by any gain in welfare to the former group.

Example 7.2: Asymmetric equilibrium in the market of Example 7.1

We know that there will be a separating equilibrium, and *H* types will get the bundle (8, 8) at equilibrium. To find B^L , the bundle obtained by *L* types at equilibrium, we solve [9] and $u_H(x, y) = u_H(8, 8)$. That is

 $\frac{1}{4}x + \frac{3}{4}y = 10$ and $\frac{1}{2}\ln(x+1) + \frac{1}{2}\ln(y+1) = \frac{1}{2}\ln(8+1) + \frac{1}{2}\ln(8+1)$.

The first equation yields x = 40 - 3y. Multiply both sides of the second equation by 2. We get $\ln(x + 1) + \ln(y + 1) = \ln(8 + 1) + \ln(8 + 1)$, and hence



Figure 5.15

 $\ln(x + 1)(y + 1) = \ln(9 \times 9)$ because the logarithm of a product is the sum of the logarithms. Then (x + 1)(y + 1) = 81. Now substitute 40 - 3y for *x*:

$$(40 - 3y + 1)(y + 1) = 81$$

Then $3y^2 - 38y + 40 = 0$ and hence $y = (38 \pm 31.048)/6$. The smaller value won't do (why?) so we must have y = 11.51, and thus x = 5.47. Then $B^L = (5.47, 11.51)$, which gives H slightly less utility than P^H . (We have rounded off.) And $u_L(B^L) = 2.36 < 2.40 = u_L(P^L)$. Finally, $u_H(B^L) = 2.19685 < 2.1972 = u_H(P^H)$. Therefore, H would choose P^H in preference to B^L .

The pair consisting of P^H and B^L of Figure 5.14 is the only candidate for equilibrium, but even this may not be an equilibrium. Suppose that n_H is relatively small. Then the line corresponding to equation [8] is close to the line depicting equation [6] as shown in Figure 5.15. Then there is a profitable contract *C* that the *H* types prefer to P^H and the *L* types prefer to B^L . (Contract *C* in Figure 5.15 is profitable because it provides the same net claim as C^0 but requires a larger premium than C^0 , which is on the zero-profit line.) Therefore, (P^H, B^L) is not an equilibrium. But we have just established that it is the only candidate for equilibrium. (We haven't got the *wrong* equilibrium; we've discovered that there is no equilibrium.) Therefore, we have proved the following.

If $\pi_H > \pi_L$ and n_H/n_L is sufficiently small, then there does not exist a competitive insurance market equilibrium.

318 Hidden Characteristics

How small is small? The next example shows that the ratio does not have to be tiny for equilibrium to be ruled out.

Example 7.3: Nonexistence of equilibrium for a special case of Example 7.2

We show that both *H* and *L* prefer the bundle C = (9.8, 9.8) to P^H and B^L , respectively. We then determine the values of n_H and n_L that would allow each person to have bundle *C*. Certainly, *H* prefers *C* to P^H because C = (9.8, 9.8) provides more of each good than $P^H = (8, 8)$. And $u_L(9.8, 9.8) = \frac{1}{4} \ln 10.8 + \frac{3}{4} \ln 10.8 = \ln 10.8 = 2.3795 > 2.3617 = u_L(B_L)$. If there is a competitive equilibrium, then *H* will get P^H and *L* will get B^L , but both prefer *C* to their own equilibrium bundle. If the outcome that gives each person the bundle *C* is *feasible* it will satisfy

$$\pi n \times (9.8 - 4 + 12 - 9.8) \le n \times (12 - 9.8),$$

which is equivalent to

$$\pi \le 0.275.$$
 [13]

(As usual π is the probability that an individual chosen at random from the entire population has an accident, and $n = n_H + n_L$ is the total population. Note that [13] is equivalent to $\pi x + (1 - \pi)y \le \pi 4 + (1 - \pi)12$ for x = 9.8 = y. In words, the expected consumption per individual cannot exceed an individual's expected wealth, $\pi 4 + (1 - \pi)12$, a condition that must hold if everyone winds up with the same bundle.) Recall that

$$\pi = \frac{1/2 \times n_H + 1/4 \times n_L}{n_H + n_L}.$$

Therefore, statement [13] becomes:

$$\frac{1}{2}n_H + \frac{1}{4}n_L \le 0.275(n_H + n_L), \quad \text{or} \quad 2n_H + n_L \le 1.1n_H + 1.1n_L, \quad \text{or} \quad n_H \le \frac{n_L}{9}.$$

A small group of *H* types (10% or less of the population in this case) can spoil the possibilities for a competitive equilibrium in the insurance market.

What condition would ensure the existence of equilibrium for Example 7.2? We need an *L*-type indifference curve through B_L that lies above the line [8], which is $\pi x + (1 - \pi)y = \pi 4 + (1 - \pi)12 = 12 - 8\pi$. This would mean that no feasible pooling contract is preferred by *L* types to B^L .

Example 7.4: Existence of equilibrium for a special case of Example 7.2

We find the bundle on $\pi x + (1 - \pi)y = 12 - 8\pi$ that maximizes u_L . Solving this equation for y yields $y = (12 - 8\pi)/(1 - \pi) - \pi x/(1 - \pi)$. Then

7. Competitive Insurance Markets

 $dy/dx = -\pi/(1-\pi)$. Set

$$V(x) = \frac{1}{4}\ln(x+1) + \frac{3}{4}\ln(y+1)$$

with *y* treated as a function of *x*.

We want to maximize V(x). If this yields less expected utility than $u_L(B_L)$ we will know that no bundle on [8] is preferred by *L* to B_L . We have

$$V'(x) = \frac{1}{4(x+1)} + \frac{3}{4(y+1)} \times \frac{dy}{dx}$$
$$= \frac{1}{4(x+1)} + \frac{3}{4(y+1)} \times -\frac{\pi}{1-\pi}$$

Confirm that V''(x) < 0 for all $x \ge 0$. If we set V'(x) = 0 we have

$$3\pi(x+1) = (y+1)(1-\pi).$$

Now substitute $y = (12 - 8\pi)/(1 - \pi) - \pi x/(1 - \pi)$ into this equation and solve for *x*. We get $x = 3.25/\pi - 3$, and hence $y = (12 - 8\pi)/(1 - \pi) - [3.25/\pi - 3] \times \pi/(1 - \pi) = (8.75 - 5\pi)/(1 - \pi)$. These two values are functions of π , so we can state

$$x(\pi) = \frac{3.25}{\pi} - 3$$
 and $y(\pi) = \frac{8.75 - 5\pi}{1 - \pi}$

Recall that $u_L(B_L) = 2.36169$. We want $\frac{1}{4} \ln[x(\pi) + 1] + \frac{3}{4} \ln[y(\pi) + 1] < 2.36169$. Try $\pi = 0.4$, which is close to $\pi_H = 0.5$. Confirm that

$$x(0.4) = 5.125$$
 and $y(0.4) = 11.25$

But $u_L(5.125, 11.25) = \frac{1}{4}\ln(6.125) + \frac{3}{4}\ln(12.25) = 2.3322$, which is less than $u_L(B^L)$, as desired. Now, $\pi \ge 0.4$ implies

$$\frac{2n_H + n_L}{4n_H + 4n_L} \ge 0.4$$

and hence $n_H \ge 1.5n_L$ is *sufficient* for existence of equilibrium. If, for example, $n_L = k$ and $n_H = 2k$, an equilibrium exists, and it will be the one that gives each H the bundle P^H and each L the bundle B^L .

When a competitive equilibrium exists is the assignment of P^H to H and B^L to L efficient? *Assuming* knowledge of each individual's accident probability it is not difficult to find a scheme that would make everyone better off. If L types consume P^L and H types continue to consume P^H then we have a feasible allocation that makes the former better off without affecting the utility of the latter. (Of course, we can modify this outcome slightly so that *everyone* is better off.) But how would a planner or government offer P^L to low-risk individuals without the high-risk individuals claiming to be low-risk and also lining up for P^L ? Even though (P^H, P^L) is feasible and each L prefers it to (P^H, B^L) and each H is indifferent, it would impossible to implement the former. Is there a superior feasible allocation that *could* be implemented? Such an allocation exists if n_H/n_L is not too large.





An example, giving S^H to the H types and S^L to the L types, is depicted in Figure 5.16. First, note that H types prefer S^H to P^H , but they also prefer S^H to S^L , so they would choose S^H if the government offered S^H and S^L . They would not choose to masquerade as low-risk individuals. Second, the low-risk individuals themselves prefer S^L to B^L and to S^H . Third, S^L yields a positive profit because it requires a higher premium than A, which pays the same net claim and yields a zero profit. Finally, S^H entails a loss because it requires a lower premium than Q while paying the same net claim as Q, which breaks even. But the government could use the profit from S^L to cover the loss from S^H as long as

 $n_L \times [$ value of y at A -value of y at $S^L] \ge n_H \times [$ value of y at $S^H -$ value of y at Q].

This will be possible if n_H/n_L is not too large. (Note that the value of *y* at *A* less the value of *y* at S^L will be very small, because S^L must be near B^L to ensure that $u_H(S^H) > u_H(S^L)$.

The pair (S^H, S^L) is not consistent with equilibrium in competitive markets because it involves cross-subsidization. But (S^H, S^L) could be implemented by the government if n_H/n_L is not too large. (Given n_L , the larger is n_H the more high-risk individuals there are to be subsidized by the low-risk group.) The plan (S^H, S^L) is feasible if the *H* types chose S^H and the *L* types chose S^L . As we have seen, the individuals do have an incentive to make those choices. Therefore, the competitive equilibrium is not efficient.

Policy	UL	U _H	
$S^L = (7, 10.8)$	2.3709	2.2738	
$S^H = (8, 10)$	2.3477	2.2976	

Table 5.6

If $\pi_H > \pi_L$ and n_H/n_L is sufficiently small, then the competitive insurance market equilibrium is not efficient even if one exists.

Example 7.5: Efficiency of equilibrium in the market of Example 7.2

Let's see if we can find conditions under which the competitive equilibrium, with $B^L = (5.47, 11.51)$ and $P^L = (8, 8)$, is inefficient. Set $S^L = (7, 10.8)$: The government offers a contract that results in x = 7 and y = 10.8. If we set x = 7 in equation [6] and then solve for y we get y = 11. That is, x = 7 and y = 11 satisfies the zero-profit condition for *L* types. Therefore, if *L* types consume x = 7and y = 10.8 they generate a surplus that can be used to subsidize the *H* types. The L types are better off with S^L than with B^L because $u_l(B^L) = 2.36169$ and $u_L(S^L) = 2.37$. What would it take to make the *H* types better off than they would be with P^{H} ? If we set $S^{H} = (8, 10)$ then we certainly have $u_{H}(P^{H}) < u_{H}(S^{H})$ because S^L provides the same amount of x as P^L and provides 2 more units of y. But that also means that S^{L} is above the line representing equation [7]. In other words, S^L operates at a loss, and that will have to be covered by the surplus from S^H . If n_H is sufficiently small relative to n_L then no matter how small the per capita surplus from S^L it will cover the deficit from S^H . Finally, Table 5.6 shows that the self-selection constraints are satisfied. We have $u_L(S^L) > u_L(S^H)$ and $u_{H}(S^{H}) > u_{H}(S^{L})$.

In determining whether the market outcome can be improved we have been careful to impose the same informational constraint on the government that private insurance companies face. What additional information would the government have to possess to verify that (S^H, S^L) is feasible? It would have to know n_H and n_L to be sure that the surplus collected from the *L* types is sufficient to cover the subsidy to the *H* types. But how could it know the number of *H* types without being able to identify the *H* types? One answer is that the *H* types reveal themselves by their choice of P^H at equilibrium. But suppose that insurance is provided initially by the government and not private insurance companies. The numbers n_H and n_L can actually be determined from data that is available to the government. Recall the definition of π from [11]. If we let *n* denote the total population, $n_H + n_L$, then the expected number of accidents for the population as a whole is πn . This will be very close to ω , the actual number of accidents. Then $\omega = \pi_H n_H + \pi_L n_L = \pi_H n_H + \pi_L (n - n_H)$. Because ω is known, if π_H and π_L

322 Hidden Characteristics

are known we can solve for n_H , which will then also give us n_L . However, U will also have to be known, to ensure that the H types will choose S^H in preference to S^L and to ensure that each group is better off than it would be under a competitive equilibrium. It's not clear that the government can obtain the relevant information, but it is also not clear that the market outcome is the best that can be achieved, given the hidden information problem.

Source

Rothschild and Stiglitz (1976) found some limitations in the Spence (1973) notion of asymmetric information equilibrium and used competitive insurance markets to illustrate. This section is based on the Rothschild-Stiglitz analysis. In separate contributions, George Akerlof, Michael Spence, and Joseph Stiglitz showed that the presence of asymmetric information in real-world markets required a new way of modeling economic exchange. They were awarded the Nobel Prize in Economics for the year 2001.

Links

Molho (1997) and Hirshleifer and Riley (1992) are similar to this section's treatment of competitive insurance markets, but different issues are highlighted. The latter is the more technical of the two.

Problem set

- **1.** In Section 7.2 we proved that a particular type will not purchase two distinct contracts in equilibrium. Prove that for any positive integer *K*, if each of the policies P^1, P^2, \ldots, P^K is purchased by at least one member of a particular risk group in equilibrium, then $P^1 = P^2 = \cdots = P^K$.
- **2.** Prove that if the expected profit per policyholder is the same for P^1 and P^2 then the expected profit per policy of $1/_2P^1 + 1/_2P^2$ is identical to the expected profit per policy of P^1 .
- **3.** Explain why $(\pi_H \times n_H + \pi_L \times n_L)/(n_H + n_L)$ is the probability that an individual randomly selected from the entire population will have an accident.
- **4.** In uncovering the properties of a competitive equilibrium, why didn't we have to worry about the possibility that the insurance contract would make an individual worse off, and hence the individual would not buy a policy at all? (*Hint*: Show that *L*'s expected utility increases as we move along [6] and away from the endowment point, and similarly for *H* and [7].)
- **5.** Show that at the separating equilibrium u_L increases as π_H falls.
- **6.** Using the definitions of $x(\pi)$ and $y(\pi)$ from Example 7.4, find a necessary and sufficient condition on π such that $U_L(B^L)$ is at least as high as $U_L(x, y)$ for any (x, y) on [8].
- **7.** For the setup of Example 7.5, find values of n_H and n_L such that an asymmetric information equilibrium exists, *and* there exist S^H and S^L such that $U_H(S^H) > U_H(P^H)$, $U_L(S^L) > U_L(B^L)$, *and* any loss from S^H is covered by a surplus from S^L .

7. Competitive Insurance Markets

- **8.** Find the point in the derivation of the competitive equilibrium at which we use the assumption that the two types have the same utility-of-wealth function. How would the argument be modified to handle the general case?
- **9.** There are two types of individuals, and there is the same number of each type. The probability that one type has an accident is 0.10, and the probability is 0.40 for the other type. Each individual has the utility-of-wealth function \sqrt{w} , where *w* is wealth. If an individual has an accident his or her wealth is 1000, but if there is no accident wealth is 2000. This is true of either type.
 - **A.** What is each type's expected utility function? Write down the competitive insurance market's zero-profit condition for a society consisting only of individuals with the probability of accident of 0.10. Write down the zero-profit condition for a society consisting only of individuals with the probability of accident of 0.40. Now, give the zero-profit condition for a competitive insurance market that offers everyone, regardless of type, the same contract.
 - **B.** Find the full information competitive equilibrium. (You may use the complete insurance theorem.) State the expected utility of each individual at equilibrium and the expected profit of the insurance companies.
 - **C.** Determine the pooling contract P^0 for which x = y. Now find a new contract P that would provide positive expected profit for any firm offering it if only P^0 were initially available. Show that the individuals purchasing P would have more utility than with P^0 . Calculate the expected profit for the firm selling P.
 - **D.** Identify the asymmetric information competitive equilibrium. Calculate the expected utility of an individual of each type, and calculate a firm's expected profit. Explain why the equilibrium really is an equilibrium. (*Hint*: Compare the expected utility at the endowment point with expected utility at equilibrium for the relevant type. Now, starting at the endowment point, show that this type's expected utility falls as *y* increases. If you have to solve an equation of the form $ay + b\sqrt{y} + c = 0$, set $q = \sqrt{y}$ and $q^2 = y$ and use the formula for solving a quadratic equation:

$$a = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

- **10.** Each of *n* individuals has utility-of-wealth function $U(w) = 50w w^2$. Let *x* represent wealth if there is an accident and let *y* denote wealth if there is no accident. If the individual buys no insurance then x = 10 and y = 20.
 - A. Is the individual risk averse? Explain.
 - **B.** Show that when the odds are fair and the individual is risk averse he or she will set x = y. You may use a general argument, or the utility-of-wealth function $U(w) = 50w w^2$.

324 Hidden Characteristics

- **C.** Find the competitive equilibrium insurance contract assuming that everyone has a probability of accident of 0.10.
- **D.** Find the competitive equilibrium insurance contract assuming that everyone has a probability of accident of 0.20.
- **E.** By means of a diagram, identify the full information competitive equilibrium and also the asymmetric information competitive equilibrium when some individuals have a probability of accident of 0.10 and some have a probability of accident of 0.20.

Auctions

1.	Intro	oduction	326
	1.1	Ten significant auctions	327
	1.2	Auctions and efficiency	329
		Problem set	334
2.	The	Vickrey Auction	334
	2.1	Equilibrium bids	335
	2.2	Social cost pricing	337
	2.3	Incentives, efficiency, and social cost pricing Problem set	341 347
3.	Four	Basic Auction Mechanisms	349
	3.1	Vickrey, English, Dutch, and first-price auctions	349
	3.2	Outcome equivalence	350
	3.3	Equilibrium bids in a first-price, sealed-bid auction	353
	∂ 3.4	The case of <i>n</i> bidders	356
		Problem set	357
4.	Reve	enue Equivalence	358
	4.1	Revenue equivalence for the four basic auctions	361
	∂ 4.2	Expected revenue is equal for the Vickrey and	
		first-price auctions	362
	4.3	Other probability distributions	363
	4.4	Equilibrium payoffs	363
	4.5	Proof of the revenue equivalence theorem	365
	∂ 4.6	Integral calculus proof of the revenue equivalence	
		theorem	371
_		Problem set	372
5.	Арр	lications of the Revenue Equivalence Theorem	374
	5.1	Multistage auctions	375
	5.2	Adoption of a standard	375
	5.3	Civil litigation	376

	5.4	Procurement	376
	5.5	Car sales	377
6. Interdependent Values		rdependent Values	377
	6.1	Revenue equivalence	379
	6.2	The winner's curse	380

Auctions have been used for more than 2500 years to allocate a single indivisible asset. They are also used to sell multiple units of some commodities, such as rare wine or a new crop of tulip bulbs. There are many different types of auctions in use, and far more that have never been tried but could be employed if we felt that they served some purpose. The aim of this chapter is to determine which type of auction should be used in a particular situation. Accordingly, we need to determine which bidder would get the asset that is up for sale and then how much would be paid for it.

1 INTRODUCTION

When the government sells things at auction-treasury bills and oil-drilling rights, for instance-the appropriate criterion for determining which type of auction should be used is the maximization of general consumer welfare. Because the bidders are usually firms, we recommend the auction type that would put the asset in the hands of the firm that would use it to produce the highest level of consumer welfare. Fortunately, this is correlated with the value of the asset to a bidder: The more valuable the asset is to consumers when it is used by firm X, the more profit X anticipates from owning the asset, and thus the higher the value that X itself places on the asset. The individual firm reservation values are hidden characteristics. On one hand, if the government simply asked the firms to report their reservation values it would get nothing resembling truthful revelation. Each firm would have a strong incentive to overstate its value to increase the probability of it being awarded the asset. On the other hand, if each firm is asked to report its reservation value and then the asset is awarded to the high-value firm at a price that is proportional to its reported value, then a firm can be expected to understate the maximum that it would be willing to pay for the asset. This chapter identifies the one

A sample of things that are auctioned: Estates, wine, art, jewelry, memorabilia, estate furniture, used cars to dealers, foreclosed houses, repossessed goods, import quotas in Australia and New Zealand, oil-drilling rights, assets from failed banks, confirmed seats on overbooked flights, contract jobs, government surplus goods, tulip bulbs, racehorses, and tobacco. auction mechanism that gives bidders an incentive to reveal their precise reservation values truthfully.

Private individuals and firms also use auctions to sell things, of course, and in those cases the seller's objective is to maximize its revenue. The private seller also has a hidden characteristic problem because the potential buyers have no incentive to truthfully reveal the maximum they would be willing to pay. Otherwise, the seller could simply select the buyer with the

1. Introduction

highest reservation value and charge that buyer a price a little below that value on a take-it-or-leave-it basis. We identify the auction formula that maximizes the seller's revenue.

1.1 Ten significant auctions

T-Rex skeleton: In October 1997, a Tyrannosaurus Rex skeleton that was 90% complete was sold for \$8.36 million at an auction conducted by Sotheby's. The winning bidder was the Field Museum of Natural History in Chicago. Paleon-tologists had worried about the sale, fearing that the winner would not make the skeleton available for research. (Perhaps someone with enormous inherited wealth would outbid the museums and then allow children to use the skeleton as a climbing apparatus.) However, it transpired that the highest 50% of the bids were from institutions. The Field Museum's supporters had raised more than \$7 million from private (anonymous) donors specifically for the T-Rex auction.

Radio spectrum: During the second half of the 1990s and first few years of the twenty-first century, previously unallocated portions of the radio spectrum were sold in a large number of auctions around the world. More than \$100 billion flowed into government treasuries. Academic economists played a leading role in designing these auctions, which were considered a big success, particularly in Britain and America. The initial U.S. auctions allocated narrowbands, used for pagers, and the later ones involved broadbands, for voice and data transmission. In 2000 the British government sold airwaves licenses for a total of \$34 billion or 2.5% of the British GNP. Similar sales were conducted in other European countries. The European licenses were for frequencies to be used by the third-generation mobile phones, which will allow high-speed access to the Internet. Other European countries also sold portions of the radio spectrum, with varying degrees of success. In terms of the money raised per capita, the Swiss auction realized only 20 Euros whereas the German and U.K. auctions yielded 615 and 650 Euros, respectively. Poor auction design accounts for the low yield in Switzerland and some other countries. Surprisingly, Spain and Sweden used the traditional "beauty contest" method to allocate licenses. This means that a jury of experts appointed by the government looked over the applications and selected the ones that they deemed best. Not only does this not solve the hidden information problem, it is susceptible to favoritism and corruption. (The vast sums that would have been paid had an auction been used are available to bribe the members of the selection committee.)

Pollution permits: Since 1990 the U.S. Environmental Protection Agency (EPA) has been auctioning permits for dumping sulphur dioxide (SO₂) into the air, resulting in a 50% reduction in the amount of SO₂ released into the air. This is significant because SO₂ is a prime ingredient in acid rain. The buyers of the pollution permits are firms that produce electricity by burning fossil fuel. An electric utility must surrender one permit to the EPA for each ton of SO₂ released. These permits have a high opportunity cost because they can be sold at auction to other electric utilities. This gives the firm an incentive to invest in cleaner production processes. By restricting the number of permits issued, the EPA can reduce the total amount of SO₂ released as a by-product of electricity generation. By allowing the permits to be traded, the EPA can achieve the reduction at lowest

cost: Firms that can find a low-cost way of modifying the production process to reduce SO_2 output will sell pollution permits to firms that can reduce SO_2 output only by switching to high-cost techniques. The additional permits allow the purchasing firm to escape some of the costly adjustments that would otherwise be required. The auction of pollution permits was initially run by the EPA, but private markets have taken over. (The EPA auction now handles only 3% of the transactions.) Auctioning pollution permits solves the hidden characteristic problem: A firm would not willing disclose the cost of reducing its SO_2 output if firms were to be required to adjust their production processes on the basis of that information. (See Section 2 of Chapter 3 for an extended discussion.)

Jobs: Because the managerial labor market is not well developed in China, the Chinese government has auctioned off top management jobs in some industries. Poland has also auctioned managerial jobs in some firms, as have other former communist countries. Jobs are auctioned in firms that are doing poorly. The bid consists of the promise of a bond, which the winner of the job must post and which is forfeited if the firm does not perform up to expectations. The bonds are about 5% of the firm's value at the time of the auction. The need for such an allocation scheme is due to a hidden characteristic problem. In transition economies, individuals often know more about their abilities as chief executives than the agency that chooses the new manager. Individuals with more confidence in their own abilities are likely to submit higher bids. Of course, this has a hidden action dimension: Having posted a bond, there is greater incentive to run the firm well.

Offshore oil: The U.S. federal government raised \$560 million in 1990 by auctioning licenses to drill for oil in the Gulf of Mexico.

Bank assets: In the 1980s and 1990s, the federal government auctioned off the assets of hundreds of failed banks and savings and loan institutions. These financial firms failed because the value of their assets was far below the value of their obligations to depositors. The assets were claimed by the government because it had to honor the deposit liabilities of the failed lending institutions. It could at least sell their assets to the private sector for whatever they would fetch. The auctions were not a great success because the government was too anxious and typically did not wait until more than a few bidders participated.

Kidneys: Each year about 100,000 people around the world are told that they will have to continue to wait for a kidney transplant. In the year 2002, 55,000 people were on the waiting list in the United States, and more than a quarter of them had been waiting for more than three years. Each year, about 6% of those on the waiting list will die, and almost 2% of the others will become too ill to qualify for a transplant. In 1999 a citizen of the United States attempted to auction one of his kidneys on the Internet. Such a transaction is illegal in the United States, and it was annulled by the firm operating the auction, but by that point the bidding had reached \$5.7 million.

Privatization: Since 1961 when the German government sold a majority ownership of Volkswagen to the public, removing it from state control, a large number of state-owned enterprises have been transferred to public ownership in Europe and Japan. In Britain, the value of state-owned enterprises decreased from about 10% of GDP to virtually zero in the 1980s. Transition

1. Introduction

economies—primarily, the former Soviet Union countries and Eastern European satellites—have privatized much of the production sector. In some of these transition economies, the assets have been sold at auction. In the case of the Czech Republic and Russia, some of these auctions involved the use of vouchers that were fairly evenly distributed to the public. The shares in each firm would have a voucher price, and each bidder would have to allocate a limited number of vouchers across the available shares. The voucher auctions typically did not lead to a high level of performance for the firms involved, primarily because the insiders managed to retain control of a firm's operations.

Electricity: For almost all of the twentieth century, the production of electricity in the United States was largely undertaken by local monopolies that were regulated by state governments. Most of the European producers were state enterprises. There was a wave of deregulation of electricity markets in the European Union and the United States at the end of the century, with Britain leading the way in 1990 when it substituted an electricity auction for state management. In country after country, the new industrial structure featured competition between private suppliers of electric power, with an auction mechanism used to allocate electricity among consumers of electric power. In Britain, France, the United States, and other countries the auction rules were designed by leading economists. They are revised when defects are detected.

Google's initial public offering: The term initial public offering (IPO) refers to the offer of shares to the general public by a firm owned by a handful of individuals-usually the founders-whose ownership shares were not previously traded on any stock exchange. The buyers become shareholders in the firm and the money they pay goes into the bank accounts of the original owners. An IPO is traditionally marketed by one of a handful of select investment banks, which charge a fee of 7% of the proceeds of the sale. In return for this substantial fee, the investment bank guarantees that the shares will be sold at the asking price. The fees and asking prices are not competitively determined the banks act like a cartel. Google, which runs one of the leading Internet search engines, broke tradition by offering its initial shares by auction over the Internet. A Dutch auction (see Section 3.1) collected more than 1.6 billion for the shares in August 2004. The advantage of the auction over the traditional method is that the latter is too vulnerable to manipulation. The investment bank handling the IPO can price the shares below their market value, in return for some form of (implicit) future compensation from the firms purchasing large blocks of the shares. Google used the online auction created by WR Hambrecht & Co.

1.2 Auctions and efficiency

When the government sells assets to the public its goal should not be to maximize its revenue. Its objective should be to see that the asset goes to the agent with the highest reservation value. Let's see why.

Suppose first that no production is involved. An antique of some sort—say a painting—is being allocated. Suppose also that individual preferences are quasi linear. Thus the individual's utility function has the form U(x, y) = B(x) + y, where commodity X is the good being auctioned and Y is generalized purchasing power—that is, dollars of expenditure on everything but X. Assume for

convenience that B(0) = 0. Then one unit of *X* obtained without cost would cause the individual's utility to increase from zero to B(1). If the individual

How to maximize government revenue: If the government auctioned the right to supply each commodity as a monopoly it would take in far more revenue than it could by any other fund-raising activity. Monopoly profits are very high, so bidders would pay lavishly for the right to be the sole supplier of a particular good. But then we'd have an economy full of monopolies, hardly the way to promote general consumer welfare. actually paid *P* for the unit of *X* then the change in utility would be $\Delta U = B(1) + \Delta y = B(1) - P$. If P < B(1) then ΔU is positive. The individual would be willing to pay any price *P* less than *B*(1) for one unit of *X* because that would increase utility. (A lower price is preferred to a higher price, of course.) But any price above *B*(1) would cause utility to fall. ($\Delta U = B(1) - P < 0$ when P > B(1).) Therefore, *B*(1) is the *maximum* that the individual would pay for one unit of *X*. That is, *B*(1) is the individual's reservation value for one unit of *X*.

DEFINITION: Reservation value

A bidder's reservation value is the maximum that the bidder would be willing to pay for the asset.

If the individual already has *x* units of *X* then the reservation value for the next unit is B(x + 1) - B(x). One of the factors influencing the reservation value is the degree to which close substitutes are available. The function *B* is different for different individuals, so we need one reservation value $B_i(1)$ for each individual *i*. To simplify the notation, we'll let V_i denote that value.

Now we show that efficiency requires that the asset be awarded to the individual with the highest reservation value. Suppose to the contrary that $V_i < V_j$ and *i* has the asset. But then U_i and U_j will both increase if *i* transfers the asset to *j* in return for $\frac{1}{2}V_i + \frac{1}{2}V_j$ dollars: The change in *i*'s utility is $\Delta U_i = -V_i + \frac{1}{2}V_i + \frac{1}{2}V_j - \frac{1}{2}V_i$, which is positive because $V_j > V_i$. And the change in *j*'s utility is $\Delta U_j = +V_j - (\frac{1}{2}V_i + \frac{1}{2}V_j) = \frac{1}{2}V_j - \frac{1}{2}V_i > 0$. We have increased the utility of both *i* and *j*, without affecting the utility of anyone else. Therefore, the original outcome was not efficient. (We have implicitly assumed that individual *j* has $\frac{1}{2}V_i + \frac{1}{2}V_j$ dollars.)

If V_H is the highest reservation value, and every individual $i \neq H$ has at least $\frac{1}{2}V_i + \frac{1}{2}V_H$ units of commodity *Y*, then efficiency requires that the asset be held by an individual whose reservation value is V_H .

Note that the sum of utilities is maximized when we give the asset to the individual with the highest reservation value, assuming that there is no change in the total consumption of *Y*. That's because there is a single indivisible asset, and hence the sum of utilities is

$$\alpha_1 V_1 + y_1 + \alpha_2 V_2 + y_2 + \alpha_3 V_3 + y_3 \dots + \alpha_n V_n + y_n$$

= $\alpha_1 V_1 + \alpha_2 V_2 + \alpha_3 V_3 + \dots + \alpha_n V_n + y_1 + y_2 + y_3 + \dots + y_n$

1. Introduction

where $\alpha_i = 1$ if individual *i* gets the asset and $\alpha_i = 0$ if *i* does not receive the asset. If $y_1 + y_2 + y_3 + \cdots + y_n$ does not change then this sum is obviously maximized by setting $\alpha_i = 1$ for the individual *i* with the highest V_i .

The outcome that assigns the asset to the individual with the highest V_i is efficient whether or not money changes hands, as long as $y_1 + y_2 + y_3 + \cdots + y_n$ is unaffected. That's because any outcome that maximizes total utility is efficient. (See Section 5.1 of Chapter 2.) In the interest of *fairness* we might require that the individual acquiring the asset make a payment that is some function of the reservation values of the individuals in the community. In Section 2.3, we demonstrate that because the individual reservation values are hidden information, *efficiency* considerations require that a payment be made by the individual receiving the asset. Moreover, we determine precisely how that payment must be related to the reservation values of the other members of the community.

We have assumed away the possibility of a trade after the auction. But does it really matter who gets the asset initially? If Rosie gets the asset and her reservation value is 600 but Soren's reservation value is 1000, can't they strike a mutually profitable trade, resulting in an efficient outcome? That assumes that both would disclose their reservation values willingly. Soren has an incentive to understate his, to keep the negotiated price down. But because he does not know Rosie's reservation value there is a possibility that he will claim that his value is, say, 500. But there is no price below \$500 at which Rosie is willing to trade. The negotiations might break down at this point. Note also that Rosie has an interest in *over*stating her reservation value. The efficient postauction trade might not take place.

We are back to the original hidden characteristic problem. Think of two heirs who squander the majority of a disputed legacy as they battle each other in court. A more common instance is that of a firm's owners and workers enduring a lengthy strike that does considerable harm to both, as management tries to convince workers that the owners' reservation value is too low to permit it accept their demands, and the workers try to convince management that their reservation value is too high to permit them to accept the owners' offer. Therefore, we must employ a mechanism to generate an efficient outcome when individuals are motivated by self-interest. We cannot rely on self-interest to lead to an efficient outcome without a framework of appropriate incentives.

Example 1.1: A bargaining breakdown

Individual *J* owns an asset that *J* wishes to sell to individual *K*. *J*'s reservation value is 2, but *J* does not know *K*'s reservation value. As far as *J* is concerned, *K*'s value is drawn from the uniform probability distribution on [0, 5], the interval from 0 to 5. By definition, this distribution is such that the probability that *K*'s reservation value is less than the number *P* is equal to the fraction of the interval [0, 5] that is covered by the subinterval [0, P]. (See Section 6.5 of Chapter 2.) In other words, the probability that the random value is less than *P* is *P*/5. Now, *J* offers to sell the asset to *K* at price *P*. This is a take-it-or-leave-it offer, so

332 Auctions

K will accept the offer if and only if *K*'s reservation value V_K is greater than *P*. The probability that $V_K > P$ is 1 minus the probability that $V_K < P$. Hence, the probability that $V_K > P$ is 1 - P/5. The expected payoff to the seller *J* is the probability of the offer's acceptance multiplied by the payoff to *J* in case of acceptance, which is $P - V_J$. Therefore, *J*'s expected payoff is

$$\left(1 - \frac{P}{5}\right) \times (P - 2) = \left(1 + \frac{2}{5}\right)P - \frac{1}{5}P^2 - 2$$

This is a quadratic, which we wish to maximize. The value of *P* that maximizes *J*'s expected payoff is

$$P^* = \frac{1+2/5}{2/5} = 3.5$$

Therefore, *J* will offer to sell the asset to *K* at a price of 3.5. However, if $V_K < 3.5$ the offer will be rejected by *K*. But if $V_K > 2 = V_J$ efficiency requires that the asset be held by *K*. Therefore, if *J* owns the asset and $2 < V_K < 3.5$, the outcome will not be efficient, and the inefficiency will not be corrected by a voluntary exchange between *J* and *K*.

Now, suppose that the asset is up for auction because it is used in a production process. The bidders are firms, and V_i is the firm *i*'s *economic* profit: If the firm were to use the asset in combination with other inputs it would be able to earn enough revenue to cover all the production costs, including a normal return on capital, and have V_i dollars left over. (Specifically, V_i is the present value of the stream of profits.) If firm *i* were to obtain the asset at any price *P* less than V_i it would still obtain a positive economic profit, and hence would be willing to pay any price less than V_i . (Of course, lower prices are more profitable than higher prices.) However, if it paid more than V_i , ownership of the asset would not yield enough revenue to cover all production costs and provide a normal return on capital. Therefore, V_i is the maximum that firm *i* would be prepared to pay for the asset, and hence is the firm's reservation value.

We establish that it is in consumers' interest to have the asset awarded to the firm with the highest reservation value by showing that V_i is the net benefit that firm *i* would provide to consumers by employing the asset in production. V_i is economic profit, which in turn equals revenue minus cost. Revenue is a measure of consumers' willingness to pay for the firm's output. Consumers wouldn't pay a lot for the good if it didn't deliver a corresponding high level of benefit. Therefore, the revenue that a firm takes in can be used as a measure of the gross benefit that consumers derive from the firm's activities. But a good may provide a high level of benefit only at a very high cost in terms of foregone output of other goods and services. A yacht, for example uses a lot of scarce resources—skilled labor and highly productive equipment—so the resources employed in producing the yacht could have been employed in producing other goods and services that generate a lot of consumer benefit. The more productive firm *i*'s inputs would be if employed somewhere else in the economy, the higher the demand for those inputs and hence the higher the market value of the inputs—as a result

1. Introduction

of competition by all firms for their use. Therefore, the cost of inputs used by firm *i* is a measure of the value of the goods and services that could be produced

Negotiations don't always break down, particularly when the difference in reservation values is extreme. Spectrum licenses were allocated by lottery in the United States from 1982 to 1993. In 1991 the lucky winner of a cellular telephone license subsequently sold it to Southwestern Bell for \$41.5 million (New York Times, May 30, 1991, p. A1). However, the lotteries spawned serious inefficiencies that were not quickly rectified by the market. The individual communications provider served a relatively small territory, significantly delaying the creation of a nationwide network that would allow cell phone users to "roam" (Milgrom, 2004, pp. 3, 20).

if the inputs were employed elsewhere. This means that the market value (i.e., cost) of the inputs used by firm *i* are a measure of the value of consumer goods and services lost to the economy by employing the inputs in firm *i*. The firm's cost is equal to the cost to consumers of the firm's activities. Therefore, "revenue minus cost" equals "gross benefit to consumers of firm *i*'s activities minus the cost to consumers of those activities." That is,

```
revenue -\cos t =  net benefit to consumers.
```

We want the asset to be awarded to the firm that delivers the highest net benefit to consumers. Therefore, we want to employ an auction mechanism that always allocates an asset to the firm with the highest reservation value, even when the firms bid strategically. We reached the same conclusion for assets that are

not involved in production and the bidders are households. We refer to this as asset efficiency.

DEFINITION: Asset efficiency

We say that the asset is allocated efficiently if it is assigned to the agent with the highest reservation value.

If the government simply asked each firm to report its reservation value, on the understanding that the asset would go to the firm with the highest value, we wouldn't get anything resembling truthful revelation. Every firm would have a strong incentive to vastly overstate its value, to increase its chance of obtaining the asset. But perhaps there is an auction that would give each firm an incentive to reveal its value truthfully. There is, and it is the subject of the next section.

Sources

The T-Rex auction is reported in *Science News*, December 13, 1997, vol. 152, pp. 382–3. The discussion of the European airwaves auctions is based on Binmore and Klemperer (2002) and Klemperer (2002b). For a discussion of the allocation of top managerial jobs in China see p. 217 in McMillan (1997). The data on the kidney transplant waiting list is from Roth, Sönmez, and Ünver (2004). The brief sketch of privatization is based on Megginson and Netter (2001). Example 1.1 is from Maskin (2003). Support for the claim that investment banks

334 Auctions

exploit their substantial market power can be found in *The Economist*, May 8, 2004, p. 14: "Acting like a cartel, these banks rarely compete on price." See also Nalebuff and Ayres (2003, p. 198): In effect they give gifts to "favored clients and executives whose business they are courting" in return for (implicit) future considerations.

Links

McMillan (1994, 2002) are good accounts of the auctioning of radio frequencies. Kirby, Santiesteban, and Whinston (2003) use the Vickrey auction in an experiment designed to determine if students who are more patient perform better. Demsetz (1968) suggested that the government auction the right to be the sole supplier of a particular good in the case of a natural monopoly. The winner would be the firm proposing the lowest output price. Laffont and Tirole (1987) extend this to the auctioning of the right to complete a government project. (Alternatively, see Chapter 7 of Laffont and Tirole, 1993.) Arrow (1979) and d'Aspremont and Gerard-Varet (1979) extend the analysis of resource allocation under uncertainty well beyond the single indivisible asset case.

Problem set

- 1. The proof that an outcome is efficient only if the asset has been awarded to the individual with the highest reservation value implicitly assumed that individual *j* has $\frac{1}{2}V_i + \frac{1}{2}V_j$ dollars. Show that the outcome in which the individual with the lowest reservation value has both the asset and all of the commodity *Y* is in fact efficient.
- **2.** Example 1.1 assumed that $V_J = 2$. Rework Example 1.1 with the individual *J*'s reservation value represented as a variable V_J , known to *J* of course. For what values of V_J and V_K will there be an inefficient outcome?

2 THE VICKREY AUCTION

Assume that a piece of physical capital—an asset—is to be sold, and there are several potential buyers. Each buyer attaches a different value to the asset because the bidders have different opportunities for combining it with other real assets that they own. This *reservation value* is the maximum sum of money that the individual or institution would be willing to pay for the asset. The reservation values are unknown to the seller. If they were known, the seller would simply sell the asset to the party with the highest reservation value for a price just under that reservation value. And because of that, buyers would not willingly and truthfully disclose their reservation values. The seller faces a hidden characteristic problem. Is there a scheme by which the seller could discover the individual reservation values and thereby sell the asset to the individual (or company) with the highest reservation value?

In the language of auction theory, we are assuming *private values*. At the other extreme is the *common values* case in which the asset has one specific value—its equilibrium market price—and every bidder accepts this, but they have different estimates of that market value.

DEFINITION: Private versus common values

Reservation values are private if each bidder's value is independent of the others'. In a common values auction, each bidder knows that the asset is worth the same to each bidder, but each has only a rough estimate of what that common value is.

Everyone is familiar with the oral auction with ascending bids. The auctioneer calls out a price until someone accepts that price, whereupon the auctioneer

Until very recently it was widely believed by economists that the second-price, sealed-bid auction was invented by William Vickrey in 1961. In fact, this auction has been used to sell stamps to collectors since at least 1893 (Lucking-Reiley, 2000). raises the price again. He then asks for a new bid—that is, acceptance of the new price—and so on until no one is willing to accept the price, at which point the article is sold to the bidder who accepted the last price, which will be the price actually paid by the winner. This is the standard *English auction*. However, we begin by investigating a close relative, the *second-price* auction, and show that it induces truthful rev-

elation of an individual's reservation value: The asset goes to the highest bidder who then pays a fee equal to the second-highest bid.

DEFINITION: The Vickrey or second-price auction

Each individual submits one bid, usually without knowing what anyone else has bid. The asset is awarded to the high bidder at a price equal to the secondhighest bid. If there are two or more individuals with the same high bid, the tie can be broken in any fashion, including randomly.

2.1 Equilibrium bids

If the Vickrey auction is used it is in a person's self-interest to enter a bid equal to his or her true reservation value. Let's prove this. First, consider a simple example.

Example 2.1: Four bidders

The reservation values of bidders *A*, *B*, *C*, and *D* are displayed in Table 6.1. What should individual *B* bid if the Vickrey auction is used? Will it depend on what the others bid? Suppose *B* bids 125. If that were the highest bid and the next highest bid is 100 then *B* would be awarded the asset at a price of 100. With any bid over \$100, *B* would wind up paying \$100 for something worth only \$70 to him. So, submitting a bid above one's reservation value can be very unprofitable. What if *B* bids below 70, and the highest bid is 100? From the standpoint of *B*, the

Table 6.1

Bidder	A	В	С	D
Reservation value	100	70	40	20

outcome would be the same as if *B* bid 70 (or anything below 100): The asset would go to someone else. The same reasoning will show that neither *C* nor *D* could benefit from submitting a bid different from their own reservation values but could be made worse off as a result. Now, consider *A*'s bid. If the highest bid submitted by anyone else is 70, then *A* gets the asset at a price of \$70 with a bid of \$100 or anything higher than \$70, leaving *A* with a profit of 100 - 70 = 30. If *A*'s bid is below 70, and someone else has submitted a bid of 70, then *A* will not be awarded the asset and will sacrifice the profit of 30. A bid different from *A*'s reservation value cannot benefit but could harm *A*.

Now, we prove that for any number of bidders, and any set of reservation values, no individual can profit from submitting a bid different from that person's reservation value if the asset is allocated by the Vickrey auction. In other words, bidding one's reservation value is a dominant strategy, regardless of what anyone else bids.

DEFINITION: Incentive compatibility

An auction mechanism is incentive compatible if for each participant, submitting a bid equal to the individual's reservation value is a dominant strategy.

Suppose that person *X* has a (true) reservation value of *V* and that *U* is the highest of all the bids except for *X*'s own bid. What should *X* bid? First, suppose that *U* is less than *V* (Figure 6.1). Under truthful revelation, *X* will bid *V*, will win the asset as the high bidder, and will pay *U* for it, because *U* would be the second highest bid. Can *X* benefit by submitting a bid other than *V*? There are three possibilities, illustrated in Figure 6.1: A bid such as *L* in the region below *U*, a bid *M* somewhere between *U* and *V*, and a bid *H* above *V*. With either *M* or *H* individual *X* will still be the high bidder, will still get the asset, and will still pay *U* for it because *U* would be the second-highest bid. Therefore, *M* and *H* have the same effect on *X*'s payoff as *V*. However, if *X* bids *L* below *U* then *X* will not be the high bidder and will not get the asset, thereby forfeiting the profit of V - U (the difference between the true value to *X* and the price paid) that

Figure 6.1

2. The Vickrey Auction



would result from a bid of V. (A bid of L = U by individual X would create a tie, which we suppose would be settled by the flip of a coin, in which case there is a positive probability that X would forfeit the profit of V - U.) In short, when V is higher than any other bid, deviating from a bid of V could never benefit X, but it can do harm.

Now let's consider the strategy that maximizes *X*'s payoff when there is another bid above *V*, the true reservation value of individual *X*. Let *U* denote the highest bid of everyone but *X*. (Here in our lab we know that *U* is in fact the highest bid of all.) Again, we need to consider three possibilities (Figure 6.2): The alternative bid *L* is in the region below *V*, or the alternative bid *M* is somewhere between *V* and *U*, or it is at *H* above *U*. With either *L* or *M* individual *X* will be outbid, just as with a bid of *V*, will not get the asset, and will not have to make a payment. But if *X* bids *H* above *U* then *X* will be the high bidder and will win the asset at the price *U*, the next highest bid. In that case *X* will have paid *U* for something worth only *V* to *X*, resulting in a loss of U - V. That loss would have been avoided by submitting a bid equal to *X*'s true reservation value. (A bid H = U by individual *X* would create a tie, which we suppose would be settled by the flip of a coin, in which case there is a positive probability that *X* would suffer a loss of U - V.) In this case we see also that deviating from a bid of *V* could never benefit *X*, but it can do harm.

We have demonstrated that submitting a bid equal to your reservation value is a dominant strategy for the Vickrey auction. The argument appeared to assume that individual *X* knew what the others would bid. To the contrary, we showed that even if *X* could read everyone else's mind, *X* could never profit by deviating from truthful revelation. And this holds true whether others bid wisely or not. (The proof didn't require us to make any assumption about the soundness of the other bidders' strategies.) Whatever the other bids are, and however they are arrived at, you can't do better than bidding your own reservation value in a Vickrey auction, whatever you know about the bids of others.

Because all individuals' bids equal their true reservation values, the asset will in fact be awarded to the individual with the highest value. Therefore, the Vickrey auction is asset efficient.

In terms of Example 2.1, *A* will bid 100, *B* will bid 70, *C* will bid 40, and *D* will bid 15. *A* will get the asset and pay 70 for it. But our argument was completely general. It applies to the auctioning of any object among any number of individuals. And once the object is allocated it is not possible for two individuals to engage in a mutually beneficial trade because the object goes to the person who values it most.

2.2 Social cost pricing

A mechanism uses *social cost pricing* if the individual taking an action incurs a cost equal to the cost that the action imposes on the rest of society. For the special case of the allocation of a single indivisible asset, if the asset is awarded to individual *A*, then the cost of this allocation to the rest of society is the highest payoff that would be realized if the asset were to go to someone else.

DEFINITION: Social cost

The cost to the rest of society of awarding the asset to one individual is the highest payoff that could be generated by giving the asset to someone else.

In determining the cost of giving the asset to individual *J* we calculate the payoff that would be realized by giving the asset to, say, individual *K* without deducting any payment that *K* might have to make. That is because the payment is a transfer from one person to another and thus is not a net loss to the group of individuals as a whole. However, if *K*'s reservation value is \$800 and *J*'s is \$500 then there is a net loss to the economy in giving the asset to *J*: The society as a whole loses \$300 of benefit. We say that the cost of giving the asset to *J* is \$800, so the net gain to society is +500 - 800 = -300.

We consider seven different mechanisms in which social cost pricing plays a central role, beginning with the Vickrey auction.

The Vickrey auction

The asset is awarded to the high bidder at a price equal to the second-highest bid. Because truthful revelation of the individual's reservation value is a dominant strategy, the second-highest bid will be the second-highest reservation value. Therefore, the price that the winner pays is equal to the second-highest reservation value, which is the cost to the rest of society of giving the asset to the winner of the Vickrey auction. In other words, the Vickrey auction uses social cost pricing.

Example 2.2: Four bidders again

As in Table 6.1, *A*'s reservation value is 100, *B*'s is 70, *C*'s is 40, and *D*'s is 30. If the asset were given to individual *A* then the cost to the rest of society is 70, because that is the highest payoff that could be generated by giving it to someone other than *A*. If the asset were given to *B* or *C* or *D* then the cost to the rest of society would be 100.

Before presenting the other six mechanisms we recall that social cost pricing in general involves charging an individual a fee equal to the cost that the individual's action has imposed on the rest of society.

Resource allocation

A general equilibrium is a configuration of prices at which every market simultaneously clears. A general *competitive* equilibrium is a general equilibrium in an economy in which each industry is competitive. Consider a private ownership market economy. At equilibrium, each consumer chooses a consumption plan at which the marginal rate of substitution between goods X and Y is equal to the ratio P_X/P_Y of the respective prices. This holds for any two goods X and Y that are consumed. The opportunity cost incurred by Jordan when he orders a unit of X is P_X/P_Y . It costs P_X dollars to buy a unit of X; each dollar will buy $1/P_Y$ units of Y, so P_X dollars spent on X could have been used to purchase $P_X \times (1/P_Y)$ units of commodity Y. Jordan takes the opportunity cost P_X/P_Y of X into consideration in determining his utility-maximizing consumption plan. Because the ratio P_X/P_Y also equals Leo's marginal rate of substitution (MRS_L), Jordan is being forced to take the preferences of Leo into consideration when Jordan formulates his consumption plan. Every unit of X consumed by Jordan is worth MRS_L to Leo, in the sense that $MRS_L = P_X/P_Y$ is the minimum amount of Y that would compensate Leo for the loss of a unit of X. We can say that MRS_L is the cost to society of Jordan taking a unit of good X for himself. In other words, P_X/P_Y is the cost that one imposes on society by consuming a unit of good X.

The ratio P_X/P_Y is also the amount of *Y* that could have been produced, given available technology, with the resources required to provide one more unit of *X* to consumers. This is another sense in which P_X/P_Y can be viewed as the cost individuals impose on society by ordering a unit of commodity *X* for their own use.

∂ Constrained optimization

Mathematical programming gives us another example of social cost pricing. Consider the problem

maximize f(x, y) subject to $g(x, y) \le a$ and $h(x, y) \le b$.

The function f represents the goal or objective, and we want to pick the values of x and y that maximize f. But there are constraints g and h, and they restrict the values of x and y that we can select.

The function f expresses the goals of society, but the society could be the set of shareholders of a particular firm, with f(x, y) denoting the profit from the production of x units of commodity X and y units of commodity Y. The constraints represent limitations such as warehouse and transportation capacity. The point is, that the example has a wide range of interpretations. If f is the value to society of the plan (x, y) then g and h reflect resource utilization by the plan of two inputs A and B—labor and capital, say—with a and b denoting the total amount available of A and B, respectively. The plan (x, y) uses g(x, y) units of labor, and that cannot exceed the total amount of labor, a, in the economy. Similarly, the plan (x, y) uses h(x, y) units of capital, and the economy has only b units of capital.

The solution of the constrained optimization program can be characterized by means of two *Lagrangian* (or *Kuhn-Tucker*) variables, α and β , associated

340 Auctions

with the respective constraints *g* and *h*. If x^0 and y^0 constitute a solution to the problem then there exist $\alpha \ge 0$ and $\beta \ge 0$ such that

$$\frac{\partial f(x^0, y^0)}{\partial x} - \alpha \frac{\partial g(x^0, y^0)}{\partial x} - \beta \frac{\partial h(x^0, y^0)}{\partial x} = 0,$$
[1]

$$\frac{\partial f(x^0, y^0)}{\partial y} - \alpha \frac{\partial g(x^0, y^0)}{\partial y} - \beta \frac{\partial h(x^0, y^0)}{\partial y} = 0.$$
 [2]

The variable α is a price in the sense that it is the value of the resource *A* underlying constraint *g*: If additional units of *A* can be obtained then α is the rate at which *f* will increase per unit of *A* added. And $\partial g(x^0y^0)/\partial x$ is the rate at which *A* is consumed at the margin. *B* and β are interpreted similarly.

Notice that we arrive at the same optimal plan (x^0 , y^0) if we maximize

$$f(x, y) - \alpha g(x, y) - \beta h(x, y)$$

treating α and β as given prices of *A* and *B* respectively. Therefore, α and β truly are social cost prices. (See Section 3 of Chapter 2 for an extensive treatment.)

A computer network

Suppose that the society that we are studying is actually a network of computers. Each computer is capable of carrying out a variety of tasks, but some agent must assign tasks to the individual computers. Computer scientist C. A. Waldspurger and colleagues at the Palo Alto Research Center (owned by Xerox) have programmed another computer to assign the tasks. One *could* program the central computer to gather data on the computational burden that each computer is currently carrying and then do the complex job of computing the optimal assignment of new jobs. Instead, the Xerox technicians have the central computer auction computer time. An individual computer can bid for time on other computers—each computer is given a "budget." Computational capacity is transferred from computers that "have time on their hands" to computers that currently do not have enough capacity to complete their assigned tasks. The price at which the transaction takes place is adjusted by the center in response to demand and supply.

Tort damages

A *tort* is an instance of unintentional harm to person *A* as a result of the action of person *B*. If the injury occurred because *B* did not exercise reasonable care

Millions of automobiles sold in the United States have been recalled as a result of safety defects that are then repaired at the manufacturer's expense. Two forces are at work: If one car maker does this the others have to follow suit to protect their reputations. But why would one manufacturer make the first move? To forestall civil suits by injured customers. That's the second force at work. then *B* can be held liable for the damages to *A* according to U.S. law and the law of many other countries. Frequently, the potential harm to *B* can be avoided by means of a contract between *A* and *B*. In such cases government intervention is not required, except to enforce the contract. For example, the contract signed by professional athletes and their employer can specify penalties in the event an athlete fails to show up for a game or even a practice. But in many cases, it would be too costly to arrange

2. The Vickrey Auction

all the contracts necessary for efficiency. I can't enter into a contract with every motorist who could possibly injure me as I walk down the sidewalk. By allowing me to collect for damages in civil court, tort liability implicitly imposes costs on anyone who unintentionally injures another. The closer the tort liability is to the amount of harm inflicted the greater the incentive an individual has to take decisions that incorporate the potential harm to others as a result of personal negligence.

The pivotal mechanism

The pivotal mechanism discussed in Section 2 of Chapter 8 induces truthful revelation of the benefit that an individual derives from a public project. It does so by imposing a tax surcharge on person A that is equal to the loss in utility suffered by everyone else as a result of A's participation. If A's participation has no effect on the outcome then there is no loss suffered by others and hence no surcharge paid by A. But if the outcome would have been F without A's participation and, as a result of A submitting A's benefit function, the outcome actually is G, then A's tax surcharge is the difference between the total utility that everyone but A would have derived from F and the total utility that everyone but A will derive from G. This makes the tax surcharge equal to the cost that A's action (participation) imposes on the rest of society.

Franchises

What payment schedule should the owner of a firm offer to the firm's manager to maximize the firm's contribution to the owner's wealth? The franchise solution comes closest to giving the manager maximum incentive. It does so *by giving all of the profit to the manager*—all of the profit over and above a fixed payment to the owner by the manager, that is. The manager then becomes the residual claimant: After the fixed payment (franchise fee) is made, every dollar of profit realized by the firm goes into the manager's pocket. This is an example of social cost pricing because the cost to the team—which you can think of as the manager-owner duo, or even society—of shirking by the manager is exactly equal to the cost borne by the manager. Even though the manager, not the firm's owner, is the residual claimant, the owner's return is maximized because the high degree of incentive under which the manager operates leads to high profits, and hence a high franchise fee can be set. If uncertainty introduces a random component to profit, then social cost pricing still maximizes the return to the owner of the firm as long as the manager is risk neutral.

2.3 Incentives, efficiency, and social cost pricing

We have shown that the Vickrey auction satisfies incentive compatibility and asset efficiency (defined in Sections 2.1 and 1.2, respectively). Now we show that it is the only auction mechanism satisfying those two properties plus the simple requirement that an individual who doesn't get the asset doesn't have to pay anything. This new criterion is called the participation constraint.

DEFINITION: Participation constraint

An auction mechanism satisfies this condition if an individual who is not awarded the asset doesn't make or receive a payment.

Consequently, participating in the auction cannot make you worse off.

We begin by confining attention to *direct* auction mechanisms, which simply ask all individuals to report their reservation values. Two simple rules identify a particular direct mechanism: Selection of the individual who receives the asset as a function of the reported reservation values and specification of how much that individual pays, as a function of the reported reservation values.

DEFINITION: Direct auction mechanism

All individuals are asked to report their reservation values, and the asset is awarded to one of these individuals, depending on the reported values R_1, R_2, \ldots, R_n of the *n* individuals. $P(R_1, R_2, \ldots, R_n)$ is the price paid by the person to whom the asset is awarded, as a function of the reported values.

The Vickrey auction satisfies the participation constraint and asset efficiency by definition. Section 2.1 demonstrated that it is incentive compatible. We now prove that it is the *only* direct mechanism that has all three properties.

Uniqueness of the Vickrey auction

The Vickrey auction is the only direct auction mechanism satisfying incentive compatibility, asset efficiency, and the participation constraint.

Here is the proof: Incentive compatibility means that each agent *i* reports his or her true reservation value. In symbols, we have $R_i = V_i$, for each individual *i*, where V_i denote's *i*'s true reservation value, known only to *i*, and R_i is *i*'s reported reservation value. Incentive compatibility and asset efficiency together imply that the asset is awarded to the individual with the highest R_i . Therefore, the only property of the auction mechanism to be determined is the payment schedule $P(R_1, R_2, ..., R_n)$. We show that our three criteria imply that it has to be the Vickrey payment schedule. That is, $P(R_1, R_2, ..., R_n)$ will be equal to the second-highest R_i .

Consider an individual acting alone, as opposed to someone representing a firm. That person's payoff is captured by the quasi-linear utility function U(x, y) = B(x) + y. An individual who is not awarded the asset pays nothing (because the participation constraint is satisfied): The individual's consumption of *X* is unchanged, and consumption of *Y* does not go down. Therefore,

2. The Vickrey Auction

$$R_2$$
 T_1 $P(R_1, R_2, \ldots, R_n)$

Figure 6.3

the change in utility of an individual who does not receive the asset cannot be negative. If, say, person 1 does gets the asset then her change in utility is

$$\Delta U_1 = B_1(1) + \Delta y_1 = B_1(1) - P(R_1, R_2, \dots, R_n)$$

= $V_1 - P(R_1, R_2, \dots, R_n),$

which is the benefit that she gets from the asset minus what she pays for it.

If the bidder is a firm, then its payoff is the effect of the auction on its profit, and that is equal to the reservation value minus the price paid, if the firm winds up with the asset. Therefore, whether agent 1 is a firm or an individual the change in its payoff is $V_1 - P(R_1, R_2, ..., R_n)$ if it wins the asset. For the rest of this section we refer to a bidder as an *agent*.

For convenience, we assume that the agents have been labeled so that $R_2 \ge R_i$ for all i > 2. In words, agent 2's bid is the highest, with the possible exception of agent 1. Suppose that $V_1 > R_2$, which means that agent 1's bid would be highest *if* she were to report truthfully. If she chooses some $R_1 < R_2$ she would not get the asset and her payoff would not fall. Therefore, incentive compatibility requires that agent 1's payoff is not negative when she bids $V_1 > R_2$ and is awarded the asset. A basic incentive compatibility condition, then, is

$$V_1 - P(V_1, R_2, \dots, R_n) \ge 0$$
 whenever $V_1 > R_2$ and $R_2 \ge R_i$ for all $i > 2$.

In words, the price paid by the winner can never exceed the reservation value reported by the winner. If we substitute the variable R_1 for V_1 this can be written as follows:

$$R_1 - P(R_1, R_2, ..., R_n) \ge 0$$
 whenever $R_1 > R_2$ and $R_2 \ge R_i$
for all $i > 2$. [3]

Suppose that $P(R_1, R_2, ..., R_n) > R_2$ for $R_1 > R_2$ and $R_2 \ge R_i$ for all i > 2, with $R_1 = V_1$. Then agent 1 will get the asset and pay $P(R_1, R_2, ..., R_n)$ for it. Intuitively, we see that it would be possible for 1 to lower her bid and still be the high bidder. She could get the asset, but at a lower price than when she reports truthfully, contradicting incentive compatibility. Therefore, incentive compatibility would seem to imply that $P(R_1, R_2, ..., R_n) \le R_2$ when $R_1 > R_2 \ge \cdots \ge R_n$.

To establish this rigorously we suppose to the contrary that $P(R_1, R_2, ..., R_n) > R_2$ and $R_1 > R_2 \ge R_i$ for all i > 2. Let T_1 be the average of $P(R_1, R_2, ..., R_n)$ and R_2 , as illustrated in Figure 6.3. That is, $T_1 = \frac{1}{2}P(R_1, R_2, ..., R_n) + \frac{1}{2}R_2$. This means that T_1 will be less than $P(R_1, R_2, ..., R_n)$ but more than R_2 . We have

$$P(R_1, R_2, \dots, R_n) > T_1 > R_2.$$
 [4]

$$P(R_1, R_2, \ldots, R_n)$$
 V_1 R_2 R_1

Figure 6.4

Therefore, [3] implies $P(R_1, R_2, ..., R_n) > T_1 \ge P(T_1, R_2, ..., R_n)$. Therefore

$$P(R_1, R_2, \ldots, R_n) > P(T_1, R_2, \ldots, R_n).$$

But we also have $T_1 > R_2$. Therefore, the strategy T_1 results in agent 1 getting the asset but at a lower price than when she bids R_1 . This results in a higher payoff for agent 1 than when she reports truthfully by bidding $R_1 = V_1$. Incentive compatibility therefore requires, when $R_2 \ge R_i$ for all i > 2,

$$P(R_1, R_2, ..., R_n) \le R_2$$
 whenever $R_1 > R_2$.

Suppose that we actually have $P(R_1, R_2, ..., R_n) < R_2$ and $R_1 > R_2 \ge R_i$ for all i > 2. Set $V_1 = \frac{1}{2}P(R_1, R_2, ..., R_n) + \frac{1}{2}R_2$. That is, suppose that agent 1's true reservation value is halfway between $P(R_1, R_2, ..., R_n)$ and R_2 , as in Figure 6.4. We have $P(R_1, R_2, ..., R_n) < V_1 < R_2$. When agent 1 (untruthfully) reports R_1 she gets the asset, and her payoff is $V_1 - P(R_1, R_2, ..., R_n) > 0$, which is greater than the payoff of zero that she gets by truthfully reporting V_1 : When $V_1 < R_2$ she does not get the asset if her bid is V_1 . Therefore, incentive compatibility rules out $P(R_1, R_2, ..., R_n) < R_2$ when $R_1 > R_2 \ge R_i$ for all i > 2. (We are allowed to "choose" person 1's reservation value because the mechanism is required to work for all possible combinations of individual reservation values. Hence, it has to satisfy the three criteria when V_1 is between $P(R_1, R_2, ..., R_n)$ and R_2 .)

There is only one possibility left: We have to have $P(R_1, R_2, ..., R_n) = R_2$ whenever $R_1 > R_2 \ge R_i$ for all i > 2, confirming our intuition. The mechanism must be the Vickrey auction. We started with an unknown mechanism. All we knew was that it had our three properties. We proved that these properties imply that it must actually be the Vickrey auction. We know that this scheme induces truthful revelation, so we must have $R_2 = V_2$ and $P(V_1, V_2, ..., V_n) = V_2$, which is the cost to society of giving the asset to agent 1. In general, if V_H is the highest reservation value and V_J is second highest, then we must have $P(V_1, V_2, ..., V_n) = V_J$ with the asset going to H.

With the Vickrey auction the agent who gets the asset must pay a price equal to the cost the agent imposes on the rest of society by making the asset unavailable for consumption by anyone else. Moreover, this social cost pricing scheme has been *derived from* considerations of efficiency and incentive compatibility.

We can extend our result to a much wider family of auction mechanisms. A *general auction mechanism* specifies for each agent *i* a set M_i of reports from which that agent is able to choose. The mechanism also specifies for each agent *i* a function σ_i that tells the agent what to report as a function of the agent's true reservation value. That is, if agent *i*'s true value is V_i then *i* is expected to report $\sigma_i(V_i)$, a member of M_i . For instance, if the mechanism is a direct one then $\sigma_i(V_i) = V_i$.

Example 2.3: Reporting a fraction of one's reservation value

There are *n* bidders, and all are asked to report the fraction (n-1)/n of their reservation values. In symbols, $\sigma_i(V_i) = [(n-1)/n] \times V_i$. The high bidder gets the asset at a price equal to the bid. Asset efficiency is satisfied by this mechanism. (Why?) Incentive compatibility is not, however. For instance, if n = 3, $V_1 = 300$, $V_2 = 150$, and $V_3 = 120$, then under truthful revelation agent 1 will bid 200, 2 will bid 100, and 3 will bid 80. However, 1's payoff would be higher with a bid of 101.

Example 2.3 may make you wonder if there is any point to considering more general auction mechanisms. By allowing a more detailed report by a bidder—say the reservation value plus additional information—the additional information may be used to arrive at an asset-efficient outcome in a way that satisfies some properties that the Vickrey mechanism lacks. Because the true payoff functions are still hidden information, the individual must have an *incentive* to behave according to σ_i . We say that truthful revelation is a dominant strategy if for each individual *i* and each V_i there is no message m_i in M_i such that *i*'s payoff is higher when *i* reports m_i than when *i* reports $\sigma_i(V_i)$.

Uniqueness of social cost pricing

If a general auction mechanism satisfies incentive compatibility, asset efficiency, and the participation constraint then the winner of the auction must be charged a price equal to the second-highest reservation value.

We prove this simply by constructing a direct auction mechanism from a given general mechanism satisfying asset efficiency and the participation constraint, and for which truthful revelation is a dominant strategy. Given the general mechanism *G*, construct a direct auction mechanism *D* by having each agent report his or her reservation value V_i , awarding the asset to the person with the highest V_i (as G must do, by asset efficiency), and then charging the winner the price $P(\sigma_1(V_1), \sigma_2(V_2), \ldots, \sigma_n(V_n))$, where *n* is the number of bidders and *P* is the pricing formula used by *G*. By the uniqueness theorem for *direct* mechanisms, $P(\sigma_1(V_1), \sigma_2(V_2, \ldots, \sigma_n(V_n))$ must equal the second-highest V_i . Therefore, at equilibrium, *G* must charge the winner a price equal to the second-highest reported reservation value.

We could modify the Vickrey auction's pricing rule so that individuals who don't receive the asset still have to pay a fee. But that would violate the *participation constraint*. We could have payments made *to* individuals who do not receive the asset. But who would make the payment? It can't be the person who is awarded the asset because that would increase the price that that person would have to pay. But any higher price than the second-highest bid would spoil the incentive to report truthfully, as we have seen. The payment can't come from one of the losers if the participation constraint is to be respected. Therefore, charging the losers precisely nothing and having the winner pay a price equal to the second-highest bid—and hence equal to the cost imposed on society by the winner's participation—is the only pricing scheme that satisfies asset efficiency, incentive compatibility, and the participation constraint.

But do we really have efficiency? Who *gets* the payment made by the winner? It can't be one of the bidders. Otherwise, one of them would have an incentive to submit a high bid, just under the winner's reservation value, to increase the fee paid by the winner and hence the amount of money going to those who don't get the asset. The problem with that is that individuals no longer have an incentive to submit bids equal to their respective reservation values. Therefore, the payment by the winner can't go to anyone. This represents waste and destroys the efficiency of the system. In this setting, efficiency is equivalent to the maximization of $\sum_{t=1}^{n} U_t$ subject to $x_t = 1$ for one and only one individual t, and $y_t \ge 0$ for all t, and $\sum_{t=1}^{n} y_t = \theta$, where θ is the total initial amount of Y available. However, if the one who gets the asset makes a payment that doesn't go to anyone else in the society, then we have $\sum_{t=1}^{n} y_t < \theta$ and hence an inefficient outcome.

Why don't we give the payment to the person who owned the asset initially? There are two objections to this. If we want to derive the efficient and incentivecompatible pricing schedule, private ownership should emerge as part of the solution; it shouldn't be assumed at the outset. Moreover, as soon as we put an original owner on stage and have the winner's payment go to the owner we again spoil the incentive for truthful revelation. Consider: Let agent 0 be the seller, whose reservation value is V_0 . Suppose that the seller's bid B_0 is used when determining the second-highest bid and hence the price to charge the winner. If the winner's payment goes to the seller then the seller has an incentive to overstate the reservation value to increase the payment that the seller will receive. However, suppose that B_0 is not taken into consideration when determining the price that the winner of the asset will pay. We just use B_0 to determine if the seller should keep the asset. Efficiency still demands that the asset go to the agent with the highest reservation value. If B_0 is higher than every other reservation value, efficiency requires that agent 0 keep the asset. If $B_t > B_0$ then the asset goes to whichever $t \neq 0$ has the highest B_t . But suppose that $B_1 > V_0 > B_2$. The asset will go to agent 1 at a price of B_2 . But the seller has to part with the asset and receives less than its worth to him. In this case the seller would have an incentive to misrepresent his reservation value and report $B_0 > B_1$.

If there is an initial owner of the asset we cannot "close the system" so that the winner's payment goes to the seller without destroying the incentive for truthful revelation. If, however, we have a large number of agents then there will be a very low probability that one and only one person has a reservation value above or close to that of a seller. In other words, the probability that $B_1 > V_0 > B_i$ for all i > 1 is very small if there is a large number of bidders. The probability that there is a significant efficiency loss will be very low with social cost pricing.
2. The Vickrey Auction

Sources

Vickrey (1961) pioneered the study of auctions in economic theory and his seminal article anticipated important discoveries in the theory of public goods in addition to the contemporary literature on auctions and bidding. In 1996 Vickrey was awarded the Nobel Prize in economics, along with James Mirlees, another seminal contributor to the theory of incentives. The "computer network" paragraph of Section 2.2 is based on Waldspurger et al. (1990).

Links

Milgrom (1987, 1989) provides introductions to the theory of auctions and bidding. Ashenfelter (1989) discusses the particular cases of wine (excuse the pun) and art. Makowski and Ostroy (1987) arrive at social cost pricing by a different route. (See also Roberts, 1979, and Makowski and Ostroy, 1991, 1993). Sternberg (1991) analyses the sale of the assets of failed banks under both the private values and the common values assumptions. Green and Laffont (1979) derive incentivecompatible mechanisms for allocating pure public goods. A more general result is presented in Walker (1978). Holmström (1979b) treats divisible private goods. There are artificial intelligence models that use market-like evaluation to direct the transition of a computer from one state to another. See, for example, Waldrup (1992, pp. 181–9). See Chapter 8 in Cooter and Ullen (1994) or Ullen (1994) for an extended discuusion of the economics of tort damage awards. Hurwicz and Walker (1990) prove that the inefficiency due to the inequality between the initial and final total *Y* consumption is almost inevitable. Their argument applies to a wide variety of models of resource allocation.

Problem set

- Suppose that when the Vickrey auction is used each bidder other than number 1 always (mistakenly) reports a reservation value equal to half his or her true reservation value. Suppose also that bidder 1 knows that. Is truthful revelation still a dominant strategy for bidder number 1? Explain.
- **2.** Ten different direct allocation mechanisms are described. Each participant i submits a bid S_i . Any money paid by the individual who gets the asset does not go to the other participants, unless there is an explicit statement to the contrary. In each case determine if the mechanism would satisfy (i) asset efficiency if the individuals reported truthfully, (ii) the participation constraint if the individuals reported truthfully, and (iii) incentive compatibility. If a criterion is not satisfied you have to give a numerical example to show that. If the criterion is satisfied then you have to prove that it is.
 - **A.** The asset goes to the individual i submitting the highest S_i at a price equal to that S_i . No one else pays anything or receives any money.
 - **B.** The asset goes to the individual submitting the highest S_i at a price equal to the second-highest S_i . The other individuals each receive \$5.
 - **C.** The Vickrey auction is used but there is an entry fee of \$100. This fee must be paid by each participant before the bidding starts.

- **D.** The asset goes to the individual submitting the *second-highest S_i* at a price equal to the *third-highest S_i*. No one else pays anything or receives any money.
- **E.** The asset is always given to individual 1—and free of charge. No one else pays anything or receives any money.
- **F.** The asset is always given to individual 1, who is then taxed \$100. No one else pays anything or receives any money.
- **G.** The asset goes to the individual submitting the highest *S_i* at a price equal to the *average* of the second-highest bid and the lowest bid. No one else pays anything or receives any money.
- **H.** The asset goes to the individual *i* submitting the highest bid at a price equal to the *average* of the second-highest bid and the high bid itself. No one else pays anything or receives any money.
- **I.** For this part only, assume that there are three individuals (n = 3). The asset goes to the individual submitting the highest bid at a price *P* equal to second-highest bid. The other two individuals each receive $\frac{1}{2}P$.
- **J.** For this part only, assume that there are two individuals (n = 2). A fair coin is tossed, and the asset goes to person 1 if it turns up heads and to person 2 if it turns up tails. Neither person pays any money or receives any money.
- **3.** A government agency is accepting tenders for the construction of a public building. There are *n* firms with an interest in undertaking the project. Each firm *i* has a minimum cost C_i that it would incur in construction. (C_i includes the opportunity cost of capital.) The contract will be awarded by having the firms submit sealed bids. Firm *i*'s bid B_i is the amount of money that it requires to undertake the project. The contract will be awarded to the firm submitting the lowest bid and that firm will be paid an amount of money equal to the second-lowest bid. Prove that a bid of C_i is a dominant strategy for arbitrary firm *i*.
- **4.** This question pertains to the Vickrey auction when the asset to be auctioned is owned by one of the participants, individual 0, whose true reservation value is V_0 . Answer the following two questions by means of specific numerical examples, one for *A* and one for *B*.
 - **A.** Show that if the owner's bid B_0 is used when determining the second-highest bid (and hence the price to charge the winner) then the incentive for truthful revelation is spoiled if the buyer's payment goes to the individual 0.
 - **B.** Now, suppose that B_0 is not taken into consideration when determining the price that the winner of the asset will pay. We just use B_0 to determine if agent 0 gets to keep the asset. Show that efficiency may be sacrificed.

3. Four Basic Auction Mechanisms

5. Prove the uniqueness of the Vickrey auction when we weaken the participation constraint to the following *normalization rule*: An individual whose reservation value is zero will not see his or her utility change as a result of participating in the auction. (Hint: All you have to do is show that asset efficiency, incentive compatibility, and the normalization rule imply the participation constraint.)

3 FOUR BASIC AUCTION MECHANISMS

We have already encountered the Vickrey auction. This section considers three other auction formulas. Each of them is frequently employed. We compare all four auction mechanisms and devote considerable time to working out the equilibria of two of them. (The other two have equilibria that are easy to identify.)

3.1 Vickrey, English, Dutch, and first-price auctions

The Vickrey auction was introduced in Section 2. It is a sealed-bid auction, as is the *first-price auction*, which awards the asset to the highest bidder but at a price equal to the winner's bid.

DEFINITION: First-price, sealed-bid auction

Each individual submits a bid, the high bidder receives the asset, and the high bidder pays a fee equal to that agent's own bid.

For both the Vickrey and first-price auctions there is only one round of bidding in which each agent submits his or her bid in a sealed envelope—that is, without disclosing the bid to anyone else—and when the deadline for submission is reached the envelopes are opened and the winner is announced. Don't jump to the conclusion that the winner pays less in a Vickrey auction than in a first-price auction. If Nan's reservation value is \$1000, Diane's is \$650, and everyone else's is below that, then in a Vickrey auction Nan will bid \$1000, Diane will bid \$650, and Nan will win the asset at a price of \$650. With a first-price auction Nan would not bid \$1000 because she would not gain anything by paying \$1000 for something worth a maximum of \$1000 to her. She would bid considerably less than \$1000 in a first-price auction. How much less? Sections 3.3 and 3.4 address that question.

The *English oral auction* is the one that we see in the movies. It has been used by the English auction house Sotheby's since 1744 and by Christie's since 1766. There are many quick rounds of bidding, and each round ends when someone shouts out a bid that is above the previous high. This continues until no one is willing to pay more for the asset than the previous high bid. It is then sold to the individual who made the last bid at a price equal to that bid. Of course, when this auction is used on the Internet—by eBay for instance—no one has to shout out the bid; it is submitted electronically.

DEFINITION: The English oral auction

The bidders interact directly with each other, in stages. Someone makes an initial bid, and anyone can raise it. This process continues until no one is willing to raise the bid. The asset goes to the last bidder at a price equal to his or her bid.

The *Dutch auction* has been used for centuries to allocate tulip bulbs in the Netherlands. It is the English auction turned upside down: The Dutch auction begins with the auctioneer announcing a ridiculously high price. No one will want the asset at that price, so it is lowered. And the price is lowered again and again, until someone shouts "I'll take it." The asset is then sold to that individual at that price.

DEFINITION: The Dutch auction

The auctioneer announces a very high price and then lowers it in small increments until one of the bidders declares that he or she will buy the asset at the current price. It is then sold to that agent at that price.

3.2 Outcome equivalence

Two auction mechanisms that look quite different, with different rules, can have the same outcome in the sense that the winner would pay the same price in either case. We say that the mechanisms are *outcome equivalent* if that would be true whatever the individual reservation values.

Example 3.1: The Vickrey and English auctions

Again we use the reservation values of Table 6.1 of Example 2.1: *A*'s reservation value is 100, *B*'s is 70, *C*'s is 40, and *D*'s is 30. If the Vickrey auction were used then *A* would win at a price of \$70. If the English auction were used, the bidding would not stop at a price below \$70 because either *A* or *B* would be willing to raise the bid. For either agent, there would be a new higher bid that is still below that agent's reservation value. If that new bid won, there would be a positive profit for the bidder and that would be preferred to the profit of zero that results when someone else gets the asset. Therefore, the bidding won't stop below \$70. If *A* raised the bid to \$70 then *B* would not be willing to bid more, because *B*'s reservation value is only \$70. Then *A* would get the asset for a price of \$70. The bidding would not stop below \$70, and it would not go above \$70. Therefore, the asset would go to *A* at a price of \$70. This is the same outcome as the Vickrey auction.

It is clear that the argument of Example 3.1 goes through with any number of bidders and any assignment of reservation values. However, it ignores one

3. Four Basic Auction Mechanisms

possibility. Suppose that *B* opens the bidding at \$50. Then *A* will raise the bid, but *A* won't know *B*'s reservation value. If *A* bids \$60 then *B* might respond with \$65. Then *A* will raise again, but *A*'s second bid might be \$72 or \$75. Strictly speaking, the best that we can do is claim that the winner of an English auction will pay something very close to the second-highest reservation value but not necessarily precisely that value. For practical purposes the outcomes of the Vickrey and English auctions are essentially the same, and from now one we speak as if they are always identical—to simplify the discussion. In fact, most Internet auction sites now use a technique that essentially turns their English auction into a Vickrey auction. To obviate the need for a bidder to sit at a computer terminal for hours, or even days, the software running the auction now allows a bidder to enter the maximum that the bidder is willing to pay. The algorithm then raises the bids submitted by others as long as the maximum has not been reached. This is called proxy bidding.

DEFINITION: Outcome equivalence

Two auction mechanisms are outcome equivalent if, however many bidders there are and whatever their reservation values, the same individual would be awarded the asset with either mechanism, and at the same price. Moreover, if the nonwinners have to make a payment it would be the same in the two auctions for a given specification of the individual reservation values.

The Vickrey and English auctions are outcome equivalent.

One advantage of the Vickrey auction over its English twin is the fact that the former does not require the bidders to assemble in the same place or even submit their bids at the same time. This is a consequence of the fact that truthful revelation is a dominant strategy for the Vickrey auction. Even if you knew what every other participant was going to bid, you could not do better than bidding your own reservation value. Consequently, information about the bidding of anyone else is of no value to a bidder in a Vickrey auction, and thus a bidder can submit a sealed bid at any time.

One defect of the Vickrey auction is that bidders may fear that the auctioneer will cheat and announce a second-highest bid that is substantially above the one that was actually submitted. This raises the selling price, of course, and thus the auctioneer's commission. This danger is even more acute if the auctioneer is also the seller. This sort of overstatement is not possible with the English auction because the bids come directly from the lips of the bidders. In addition, with a Vickrey auction the bidders may fear that a very high bid will tip the seller off to the asset's true value, resulting in the item being withdrawn. In the case of an English auction, neither the seller nor the auctioneer will find out how high the winner was prepared to go. However, because the two auctions are outcome equivalent, and the Vickrey auction is easier to analyze, we continue to give it serious consideration.

Surprisingly, the Dutch and the first-price auctions always lead to the same outcome.

Example 3.2: The Dutch and first-price auctions

A's reservation value is 100, B's is 70, C's is 40, and D's is 30. Suppose that the first-price, sealed-bid auction is used. We'll put ourselves in the shoes of agent A. He wants to outbid the other three, but at the same time wants to get the asset at the lowest possible price. He doesn't know the reservation values of the other three bidders, and even if he did he wouldn't know how much each would bid. Agents have to determine their bids as a function of their own reservation values and as a function the bids they expect the others to make, knowing that their bidding strategies will be based in part on what they think that others will bid. Suppose that A decides that a bid of \$75 maximizes his expected payoff when the first-price auction is used. It follows that if a Dutch auction is used instead, A would claim the asset when the price got down to \$75, provided that no one else claimed it at a higher price. Here's why: In a Dutch auction bidder A is in precisely the situation that he faces in deciding what to bid in a first-price auction. In either case he doesn't know what the others will bid, so he has to decide how much he will pay if no one else outbids him. Granted, in a Dutch auction the bidders get some information about what the others are prepared to bid. As the auctioneer brings the price down from \$200 to \$175 to \$150, and so on, they learn that the maximum anyone is prepared to pay is below \$150. But that is no longer useful information to anyone who has decided that he or she will not claim the asset at a price above \$75. It would be valuable information to someone who decided to claim the asset at a price of \$175. If that bidder knew in advance that no one else would pay more than \$150 then that bidder wouldn't have to pay \$175. But the only way to find that out in a Dutch auction is to let the price fall below \$175, and then the bidder might lose the asset to someone else although he or she would have been prepared to pay \$175. In short, the bidders have more information in a Dutch auction than in a first-price auction, but by the time they get that information it is no longer of value. With either auction, the bidder has to decide the price at which he or she will buy the asset, should that bidder be the high bidder, and he or she has to do it before the bidding starts.

Given the individual reservation values, the amount that each decides to bid in a first-price auction will be the same as in a Dutch auction. Therefore, the same individual will win in both cases, and the price will be the same.

The Dutch and first-price auctions are outcome equivalent.

A good way to show that the Dutch and first-price auctions are outcome equivalent is to turn one into the other. Imagine that *n* bidders have assembled to participate in a first-price auction. The auctioneer begins by saying, "I'm feeling too lazy to open a bunch of envelopes. I'll call out numbers, starting very

3. Four Basic Auction Mechanisms

high, and then lower them in small increments. Shout when I call the number that you have placed in your envelope. The first one to shout will be the high bidder, and hence the winner of the first-price auction. I'll check your envelope to make sure that the price at which you claimed the asset is in fact the bid that you inserted in your envelope."

Now, suppose that the auctioneer omits the last sentence. No one will check to see if the price at which you claim the asset is the same as the bid that you decided on when you thought it would be a conventional first-price auction. That means that you can claim the asset at any price you like, provided that no one else has claimed it first. Would you claim the asset at a price that is different from the bid that you wrote down before you knew about the rule change? In other words, is the information that you get when the price falls, and you discover that no one was willing to claim the asset at a higher price, of use to you in revising your bid? No. The same number of bidders remain—no one has claimed the asset—and you don't know what their bids are. As soon as someone does claim the asset you learn something, but it has no value; it's too late to be of use.

We've just shown that we can turn a first-price auction into a Dutch auction, and that the equilibrium bids will not change. Whatever bid is optimal for someone in the former will be optimal in the latter. Now, imagine that n bidders have assembled to participate in a Dutch auction. Before it gets under way the auctioneer circulates the following memo: "I have laryngitis. Instead of calling out prices, starting high and then slowly lower the price, I'm asking you to write down the price at which you've decided to claim the asset-assuming that no one has beaten you to it—and seal it in an envelope and hand it to me. I will then open the envelopes to see who would have won the Dutch auction if I had conducted it in the usual fashion." Would this change in procedure cause you to submit a price that is different from the one at which you had decided to claim the asset when you thought it would be a conventional Dutch auction? No, because you are in the same position in either case. Then we have shown that a Dutch auction can be turned into a first-price auction. The price at which an individual decides to claim the asset with the Dutch auction will be the bid that the individual submits in the first-price version. The two schemes are outcome equivalent.

We know that for both the Vickrey and English auctions the price paid by the winner will be equal to the second-highest reservation value. The seller won't know what that value is, so the seller won't know how much revenue to expect if either of those auctions is used. However, we do at least have a useful starting point. For the Dutch and first-price auctions we need to work out the price paid by the winner as a function of the individual reservation values.

3.3 Equilibrium bids in a first-price, sealed-bid auction

Suppose that you are one of the bidders in a first-price, sealed-bid auction of a single asset. You know that your reservation value is v_1 , but you don't know anyone else's. How should you bid? You don't want to bid v_1 because if you won

then you would be paying v_1 dollars for an asset that is worth no more that v_1 to you. Your payoff-maximizing strategy is to bid something less than v_1 . But how much less? To simplify our calculations, we'll assume that there is only one other bidder. We'll also assume that bidder 2's value v_2 is somewhere between 0 and 1, and that from your point of view any value in that interval is just as likely to be the actual v_2 as is any other value in the interval. (We are really supposing that both bidders agree that the asset has a maximum possible value of, say, \$10 million to anyone, and the value placed on the asset by bidder *i* is the fraction v_i of that number. Hence, if $v_2 = 0.72$ we're saying that bidder 2's reservation value is \$7.2 million.)

In assuming that bidder 2's value is a random draw from the interval from 0 to 1, with each value being as likely as any other, we are assuming the *uniform probability distribution* for v_2 . (See Section 6.5 of Chapter 2.) In short, this means that the probability that v_2 is less than a given number β is β itself. This holds for any value β in the interval. So, the probability that bidder 2's value is less than 0.8 is 0.8, the probability that bidder 2's value is less than 0.35 is 0.35, and so on. Now, the probability that your reservation value v_1 is higher than bidder 2's value is v_1 , because that's the probability that v_2 is less than v_1 . But you need to know the probability that b_2 is less than b_1 , where b_1 and b_2 are, respectively, the bids of individuals 1 and 2.

Suppose that the optimal strategy is to submit a bid equal to the fraction λ of one's reservation value. Then b_2 will equal λv_2 , but you still don't know the value of v_2 . But now you know that b_2 will never exceed λ , because v_2 cannot be larger than 1, so λv_2 cannot be larger than λ . That means that it is not payoff maximizing for you to submit a bid greater than λ . Of course a bid of $\beta > \lambda$ would win for sure, because $b_2 \leq \lambda$. But a bid halfway between λ and β would also win for sure, for the same reason. You'd still get the asset, but you'd pay less for if it than if you had bid β . In general, no bid greater than λ can be payoff maximizing for you. Therefore, you can restrict your attention to bids $b_1 \leq \lambda$. Because v_2 is uniformly distributed on the interval 0 to 1, we can think of $b_2 = \lambda v_2$ as being uniformly distributed on the interval 0 to λ .

What's the probability that a random draw from the uniform distribution on the interval 0 to λ is less than b_1 ? It is just the distance from 0 to b_1 as a fraction of the length of the interval 0 to λ itself.

Example 3.3: The probability that you have the higher bid

If $\lambda = \frac{3}{4}$ and $b_1 = \frac{3}{8}$ then λv_2 will be less than b_1 for half of the values of λv_2 in the interval from 0 to $\frac{3}{4}$. If $\lambda = \frac{3}{4}$ and $b_1 = \frac{1}{4}$ then $\lambda v_2 < b_1$ for one-third of the values of λv_2 in the interval from 0 to $\frac{3}{4}$. Suppose that $\lambda = \frac{1}{2}$. Then for $b_1 = \frac{3}{8}$ (respectively, $b_1 = \frac{1}{4}$) we have $\lambda v_2 < b_1$ for three-quarters (respectively, one-half) of the values of λv_2 in the interval from 0 to $\frac{1}{2}$.

In general, the probability that λv_2 is less than a given b_1 is b_1/λ . That's the probability that bidder 1's bid is higher than bidder 2's bid. Your payoff from a bid of b_1 is the probability of winning with b_1 multiplied by the profit you get

3. Four Basic Auction Mechanisms

when you do win. If you win, the asset is worth v_1 to you, but you paid b_1 for it, so your profit is $v_1 - b_1$. Therefore, your payoff from a bid of b_1 is

$$\frac{b_1}{\lambda} \times (v_1 - b_1)$$

because b_1/λ is the probability of winning with a bid of b_1 . Note that we are assuming that the individual is risk neutral. (See Section 6.2 of Chapter 2.)

To find your payoff-maximizing bid we merely have to determine the value of b_1 that maximizes $(b_1/\lambda) \times (v_1 - b_1) = (v_1/\lambda)b_1 - (1/\lambda)b_1^2$, a simple quadratic. Now, employ our formula for maximizing a quadratic. We get

$$b_1 = \frac{\nu_1/\lambda}{2/\lambda} = \frac{\nu_1}{2}.$$

Therefore, if you expect bidder 2 to submit a bid equal to some fraction of her reservation value, then you maximize your payoff by sending in a bid equal to *half* your reservation value. Of course, because your bid is a fraction of your reservation value, bidder 2 maximizes her payoff by setting her bid equal to half her reservation value. (We're assuming that bidder 2 is clever enough to deduce that you will set $b_1 = \frac{1}{2}v_1$.) We have a Nash equilibrium: Each person is playing a best response to the other's strategy.

Two bidders in a first-price auction

If the bidders are risk neutral and each models the other's reservation value as a random draw from the uniform probability distribution, then at a symmetric Nash equilibrium both will submit bids equal to half of their respective reservation values.

(We proved that for any λ , if bidder *j* sets $b_j = \lambda v_j$ then bidder *i*'s payoff will be maximized by setting $b_i = \frac{1}{2}v_i$. But it is possible that $\frac{1}{2}v_i > \lambda$, and we know that *that* does not maximize *i*'s payoff. A slightly smaller bid will guarantee that *i* wins, and the price paid will be slightly lower. Now, *i*'s payoff as a function of b_i is a hill-shaped quadratic, and thus if we maximize that payoff subject to $b_i \leq \lambda$ we get $b_i = \frac{1}{2}v_i$ if $\frac{1}{2}v_i \leq \lambda$, but if $\frac{1}{2}v_i > \lambda$ the solution must be $b_i = \lambda$. However, if $\lambda = \frac{1}{2}$ then we will certainly have $\frac{1}{2}v_i \leq \frac{1}{2}$ because $v_i \leq 1$. Therefore, we really do have a Nash equilibrium with two bidders when both submit bids equal to half their respective reservation values.)

We have discovered that if there are two bidders in a first-price or a Dutch auction then the seller's revenue will be exactly half of the larger of the two reservation values because that is the price paid by the winner.

Now, suppose that there are more than two bidders. The larger the number of bidders, the greater the probability that someone else has a high reservation value and hence is prepared to submit a high bid. Therefore, the more bidders there are, the greater the probability that the high bid among all the others is close to the maximum that you would be prepared to bid. That means that the greater the number of bidders, the higher you will have to bid to maximize

356 Auctions

your payoff. With n bidders we have an equilibrium in which each individual i sets

$$b_i = \frac{n-1}{n} \times v_i.$$

n bidders in a first-price auction

If the bidders are risk neutral and each models the others' reservation values as random draws from the uniform probability distribution, then at a symmetric Nash equilibrium all will submit bids equal to the fraction (n-1)/n of their respective reservation values.

If you know a little calculus you can prove this with ease, as we do in the next subsection.

It follows that if there are *n* bidders in a first-price or a Dutch auction then the seller's revenue will be the fraction (n-1)/n of the largest reservation value.

∂3.4 The case of *n* bidders

Suppose that you are in competition with n-1 other risk-neutral bidders in a first-price, sealed-bid auction. We continue to refer to you as bidder 1. As in Section 3.3, the probability that your bid is higher than individual *i*'s, when $b_i = \lambda v_i$, is b_1/λ . The probability that b_1 is higher than *everyone else*'s bid is the probability that b_1 is higher than b_2 , and b_1 is higher than b_3 , and b_1 is higher than b_4, \ldots and b_1 is higher than b_n . The probability that b_1 is higher than each other b_i is

$$\frac{b_1}{\lambda} \times \frac{b_1}{\lambda} \times \frac{b_1}{\lambda} \times \cdots \times \frac{b_1}{\lambda} = \frac{b_1^{n-1}}{\lambda^{n-1}}.$$

Therefore, your payoff from a bid of b_1 is

$$\frac{b_1^{n-1}}{\lambda^{n-1}} \times (v_1 - b_1) = \frac{v_1}{\lambda^{n-1}} b_1^{n-1} - \frac{b_1^n}{\lambda^{n-1}}.$$

We want to maximize this function. The first derivative (with respect to b_1) must be zero at the maximum, because $b_1 = 0$ can't be the solution. (With a bid of zero the probability if winning is zero, and hence the payoff is zero. But with $v_1 > 0$ and a bid of even $0.1v_1$ there is a positive, but very small, probability of winning and getting a positive profit of $0.9v_1$.) When we take the first derivative of bidder 1's payoff function and set it equal to zero we get

$$(n-1)\frac{\nu_1}{\lambda^{n-1}}b_1^{n-2} - n\frac{b_1^{n-1}}{\lambda^{n-1}} = 0.$$

Because b_1 is positive (and hence nonzero) we can divide both sides by b_1^{n-2} , yielding

$$(n-1)\frac{\nu_1}{\lambda^{n-1}} - n\frac{b_1}{\lambda^{n-1}} = 0,$$

3. Four Basic Auction Mechanisms

the solution of which is

$$b_1 = \frac{n-1}{n}v_1$$

With *n* bidders we have an equilibrium when all individuals submit a bid equal to the fraction (n-1)/n of their reservation values. (Note that if every i > 1 sets $b_i = (n-1)v_i/n$ then $(n-1)v_1/n$ does not exceed λ for $\lambda = (n-1)/n$. Therefore, for each bidder *j* setting $b_j = (n-1)v_j/n$ clearly is a best response by *j* to the strategy $b_i = (n-1)v_i/n$ for all $i \neq j$.)

Source

The paragraph on proxy bidding is based on Lucking-Reiley (2000).

Link

Krishna (2002) is a very technical, but insightful, presentation of auction theory.

Problem set

- **1.** Explain why the first-price, sealed-bid auction is *not* outcome equivalent to the Vickrey auction.
- **2.** Explain why the English auction is *not* outcome equivalent to the Dutch auction.
- **3.** There are two bidders in a first-price, sealed-bid auction. Bidder 1 has learned that bidder 2 plans to bid \$50. What is bidder 1's payoff-maximizing response as a function of his or her reservation value?
- **4.** There are two bidders in a first-price, sealed-bid auction. Bidder 1 knows that individual 2 will submit a bid of \$19 with probability $1/_2$ and \$49 with probability $1/_2$. Under each of the following four assumptions, calculate individual 1's payoff-maximizing bid, determine the probability of person 1 winning the asset, and calculate bidder 1's payoff.
 - **A.** Bidder 1's reservation value is \$100.
 - **B.** Bidder 1's reservation value is \$60.
 - **C.** Bidder 1's reservation value is \$30.
 - **D.** Bidder 1's reservation value is \$15.
- 5. There are two bidders in a first-price, sealed-bid auction. Bidder 1 knows that individual 2 will submit a bid of \$29 with probability ²/₃ and \$59 with probability ¹/₃. Under each of the following four assumptions, calculate individual 1's payoff-maximizing bid, determine the probability of person 1 winning the asset, and calculate bidder 1's payoff.
 - A. Bidder 1's reservation value is \$99.
 - **B.** Bidder 1's reservation value is \$60.
 - **C.** Bidder 1's reservation value is \$42.
 - **D.** Bidder 1's reservation value is \$15.

358 Auctions

- **6.** There are two bidders in an English auction. Bidder 1's reservation value is \$75. Determine bidder 1's payoff-maximizing bid, the winner of the asset, the price paid, and person 1's payoff, under each of the following four assumptions:
 - A. Bidder 2's reservation value is \$100.
 - **B.** Bidder 2's reservation value is \$60.
 - C. Bidder 2's reservation value is \$30.
 - **D.** Bidder 2's reservation value is \$15.
- **7.** Determine an individual's payoff-maximizing bidding strategy at equilibrium in a first-price, sealed-bid auction for the following four cases:
 - **A.** There are two bidders and each reservation value is drawn from the uniform probability distribution on the interval from 0 to 5.
 - **B.** There are two bidders and each reservation value is drawn from the uniform probability distribution on the interval from 2 to 5.
 - **∂C.** There are four bidders and each reservation value is drawn from the uniform probability distribution on the interval 0 to 1.
 - **∂D.** There are four bidders and each reservation value is drawn from the uniform probability distribution on the interval 1 to 11.
- **8.** There are two bidders, *A* and *B*. Each bidder's value is drawn from the uniform probability distribution, with values between zero and unity, inclusive. Will the first-price, sealed-bid auction and the Vickrey auction yield the same revenue when $V_A = \frac{3}{4}$ and $V_B = \frac{1}{4}$, where V_i is the value that *i* places on the asset?
- **9.** There are two bidders, *A* and *B*. Each bidder's value is drawn from the uniform probability distribution, with values between zero and unity, inclusive. Will the English auction and the first-price, sealed-bid auction yield the same revenue when $V_A = \frac{3}{4}$ and $V_B = \frac{1}{4}$, where V_i is the value that *i* places on the asset?

4 **REVENUE EQUIVALENCE**

The seller of an item at auction wants to make as much revenue as possible. Therefore, many different types of auctions have to be considered, to see which would be most profitable from the seller's point of view. This is problematic because auction *A* might be optimal for one range of buyer reservation values, whereas auction *B* is best for a different range of values. The buyers know their own reservation values, but these are unknown to the seller. From the seller's point of view, we can think of the buyer reservation values as random variables drawn from some probability distribution. The seller will want to employ the auction that maximizes the *seller's expected revenue*. Note that we assume in this section that buyers and seller are risk neutral.

The surprise is that there is a large family of auctions that generate the same expected revenue. Each has its own set of formulas to determine who wins and

how much each bidder pays but, astonishingly, the expected revenue is the same for each auction in the family, which we refer to as the set of standard auctions.

DEFINITION: Standard auction mechanism

If an agent with a reservation value of zero gets zero profit from participating in the auction, and the agent with the highest reservation value always gets the asset at equilibrium, then we say that the auction mechanism is a standard one.

A standard auction is not necessarily a direct mechanism. The first-price, sealed bid auction is obviously standard: No one who places a zero value on the asset will submit a positive bid, and the higher the reservation value the higher is the individual's optimal bid at equilibrium. Therefore, the high-value agent will win the asset at an equilibrium of a first-price, sealed-bid auction. But it is not a direct mechanism because individuals are not asked to report their reservation values. At equilibrium, all individuals bid amounts equal to a fraction of their respective reservation values. Fortunately, in proving the revenue equivalence theorem, we do not have to go into detail as far as the bidding is concerned. We map individuals' reservation values into their payoffs at equilibrium, embedding all the details in this mapping.

As is the case with the first-price auction, the agent with the highest reservation value may not submit a bid equal to his or her reservation value. But as long as the *equilibrium* strategies result in the asset going to the agent with the highest reservation value, the second defining condition of a standard auction will be satisfied.

The revenue equivalence theorem

If each of the *n* agents is risk neutral and each has a privately known value independently drawn from a common probability distribution, then all standard auctions have the bidders making the same expected payments at equilibrium, given their respective values, and thus the seller's expected revenue is the same for all standard auctions.

To see what's behind the revenue equivalence theorem, compare the firstprice, sealed-bid auction with the all-pay auction. The all-pay auction requires each participant to submit a sealed bid, and the high bidder gets the asset at a price equal to his or her bid. However, *all* participants have to pay the seller the amount of their bids. The fact that you pay whether you win or not depresses your bid—for two reasons. First, you know that you will have to pay even if you lose, so every dollar you bid has a higher expected cost than it would in a firstprice auction. Second, you know that others are in the same situation and hence

Table 6.2

	Seller's revenue				
Auction	Case A	Case B	Average		
Vickrey First-price	200 120	70 150	135 135		

will be submitting low bids, so the benefit of adding a dollar to your bid is also lower—it's not as likely to be key to winning. So, everyone will be paying the seller a small amount of money in an all-pay auction, and the seller's expected revenue turns out to be the same as in a first-price auction. Before proving the general theorem we'll illustrate what revenue equivalence *is* with an elementary situation.

Example 4.1: Two bidders and two pairs of reservation values

There are two bidders and only two possible scenarios: Case *A*, in which v_1 , agent 1's reservation value, is 240 and $v_2 = 200$. For Case *B*, $v_1 = 70$ and $v_2 = 300$. (See Figure 6.5.) With the Vickrey auction all individuals' bids are equal to their reservation values. Hence, in Case *A* if the Vickrey auction were employed the asset would go to agent 1 at a price of 200. However, if the first-price, sealed-bid auction were used, agent 1 would bid 120 and agent 2 would bid 100. (All individuals will submit bids equal to half their reservation values.) Therefore, agent 1 would get the asset for 120. If the Vickrey auction were employed in Case *B*, the asset would go to individual 2 at a price of 70, but if the first-price, sealed-bid auction were used instead, agent 2 would get the asset for 150 because agent 1 would bid 35 and agent 2 would bid 150.

Now, suppose that Case *A* occurs with probability $\frac{1}{2}$, and so does Case *B*. Then the expected revenue from the Vickrey auction is $\frac{1}{2} \times 200 + \frac{1}{2} \times 70 = 135$, and expected revenue from the first-price auction is $\frac{1}{2} \times 120 + \frac{1}{2} \times 150 = 135$ also, as shown in Table 6.2.

The two auctions provide the same expected revenue in Example 4.1. This is not true in general when there are only two possible scenarios. The purpose





Figure 6.5



Figure 6.6

of the example is to show what revenue equivalence means: It's weaker than outcome equivalence because we're only claiming that revenue will be the same *on average* for any two standard auctions. To prove this we need to assume that the possible reservation values stretch over a wide range.

4.1 Revenue equivalence for the four basic auctions

This subsection gives an intuitive explanation of revenue equivalence for a narrow but important family of cases. (The formal proof is in Subsections 4.5 and 4.6. The latter is shorter, but it employs integral calculus.)

Assume that all reservation values are drawn from the uniform distribution. We show that Vickrey and first-price auctions are revenue equivalent. We begin with the case of two bidders. Because the values are uniformly distributed in the interval 0 to 1, the average high bid v_H and the average low bid v_L divide the interval into three segments of equal length (Figure 6.6). The average second price is 1/3, and hence the expected revenue from the Vickrey auction is 1/3.

Because bidders in a first-price auction submit bids equal to half their reservation values, the average reservation value of the winner is $\frac{2}{3}$ with a bid of half that, or $\frac{1}{3}$. Therefore, the expected revenue from the first-price auction is $\frac{1}{3}$, the same as for the Vickrey auction.

Now, let's do the general case, with *n* bidders. Again, we assume that the reservation values are uniformly distributed in the interval 0 to 1, but there are *n* of them this time. They will divide the interval into n + 1 segments of equal length, as shown in Figure 6.7. The average second high bid is (n-1)/(n+1), and hence the expected revenue from the Vickrey auction is (n-1)/(n+1).

In a first-price auction with *n* bidders, payoff maximization requires the individuals to submit bids equal to the fraction (n-1)/n of their reservation values. The average high value is n/(n+1), and thus the average price paid by the winner is $[(n-1)/n] \times [n/(n+1)] = (n-1)/(n+1)$, which is then the seller's expected revenue from the first-price auction. We see that the expected revenue from the first-price auction is the same as it is for the Vickrey auction.

Finally, because the first-price auction is outcome equivalent to the Dutch auction, and the Vickrey auction is outcome equivalent to the English auction, we have established the revenue equivalence of all four auctions when the reservation values are drawn from the uniform distribution. (If two auction mechanisms are outcome equivalent, then for *any* specification of the

Figure 6.7

362 Auctions

individual reservation values, the price paid by the each agent will be the same for either auction, and thus the seller's actual revenue will be the same.) The next subsection uses integral calculus to prove the revenue equivalence of the four basic auctions when the reservation values have the uniform probability distribution.

$\partial 4.2$ Expected revenue is equal for the Vickrey and first-price auctions

We again assume that there are two bidders, and that each treats the other's reservation value as a random draw from the interval 0 to 1. We begin by calculating expected revenue for the Vickrey auction.

We know that all individuals will submit bids equal to their reservation values. Let *r* denote the value of one of the bidders and let *s* denote the value of the other. Then one person will bid *r*, and the other will bid *s*. Consider a particular value of *r*. When *s* is less than *r*, the bidder who submitted *r* will win the asset and will pay *s*, the second-highest bid. When *s* is more than *r* the second-highest bid will be *r*, and that will be the price paid for that range of values of *s*. Therefore, given *r*, the seller's expected revenue from the Vickrey auction is

$$\int_0^r s \, ds \, + \int_r^1 r \, ds = 0.5r^2 + r(1-r) = r - 0.5r^2.$$

This is obviously a function of r, which is not fixed—it's a random variable. Therefore, the seller's expected revenue (ER) is

$$ER = \int_0^1 (r - 0.5r^2) dr.$$

Because $\int r dr - \int \frac{1}{2} r^2 dr = r^2/2 - r^3/6$ we have $ER = \frac{1^2}{2} - \frac{1^3}{6} - \frac{0^2}{2} - \frac{0^3}{6} = \frac{1}{2} - \frac{1}{6} = \frac{1}{3}$, as we claimed in the previous subsection.

Now let's calculate ER for the first-price auction for which all individuals will submit bids equal to half their reservation values. When *r* and *s* are the values then the bids will be $\frac{1}{2}r$ and $\frac{1}{2}s$, respectively. For a particular value of *r*, when *s* is less than *r* the winner (the one bidding *r*) will pay $\frac{1}{2}r$. When *s* is more than *r* the winner (the one bidding *s*) will pay $\frac{1}{2}s$. Therefore, given *r*, the seller's expected revenue from the first-price auction is

$$\int_0^r 0.5r \, ds \, + \int_r^1 0.5s \, ds.$$

Now, $\int \frac{1}{2}r \, ds = \frac{1}{2}rs$ and $\int \frac{1}{2}s \, ds = \frac{1}{4}s^2$. Therefore,

$$\int_0^r 0.5r \, ds + \int_r^1 0.5s \, ds = (0.5r \times r - 0.5r \times 0) + (0.25 \times 1^2 - 0.25 \times r^2)$$
$$= 0.25r^2 + 0.25.$$

This is obviously a function of the random variable r. Therefore, the seller's expected revenue is

$$ER = \int_0^1 \left(0.25r^2 + 0.25 \right) dr.$$

Because $\int \frac{1}{4}r^2 dr = r^3/12$ and $\int \frac{1}{4}dr = \frac{1}{4}r$, we have

$$ER = \frac{1^3}{12} + \frac{1}{4} \times 1 - \left(\frac{0^3}{12} - \frac{1}{4} \times 0\right) = \frac{1}{12} + \frac{1}{4} = \frac{1}{3}$$

the same as the expected revenue for the Vickrey auction.

4.3 Other probability distributions

Let's assume that the reservation values are drawn from a distribution that is not uniform. After all, we would expect relatively small probabilities for values that are extremely high or extremely low. We won't actually specify the distribution in this section, but we do assume that it is known by all *n* individuals. Therefore, we can't calculate the equilibrium configuration of strategies for the first-price auction. We merely let $\sigma(v_i)$ denote the optimal bid for an individual with reservation value v_i . Of course, $\sigma(v_i)$ will be higher as v_i is higher. Therefore, the asset will be won by the individual with the highest reservation value for a price of $\sigma(v_H)$, where *H* denotes the individual with the highest v_i .

Now, consider a sealed-bid auction in which the asset goes to the high bidder for a price that is *four times* that bid. It is not hard to see that the optimal strategy for someone participating in this auction is to bid $\frac{1}{4}\sigma(v_i)$. Therefore, the winner will pay $4 \times \frac{1}{4} \times \sigma(v_H)$, which is the same as the price paid with the first-price auction. Therefore, the two auctions are outcome equivalent, and hence they generate the same revenue. We soon see that even standard auctions that are not outcome equivalent are revenue equivalent. *On average*, the expected payments by a given bidder will be the same in the two auctions. To prove this we have to track the payments made by a bidder in equilibrium.

4.4 Equilibrium payoffs

We begin by reducing an auction to its bare essentials. To see what the seller has to gain, we have to spend some time figuring out what the buyers will do. We are going to highlight the strategy and the profit of a generic agent whose value for the asset to be auctioned is represented by v. From the point of view of the seller, v is a random variable drawn from a particular probability distribution: The agent's reservation value v is known precisely to the agent, but because v is unknown to the seller, the seller will calculate his expected revenue as though the individual reservation values were random variables.

Assume a particular auction mechanism, and let $\mu(v)$ denote the expected payoff of our bidder when his or her reservation value is v. If the agent is a business then $\mu(v)$ is expected profit in the usual sense. If the agent is a household, bidding on a painting for the home, say, then $\mu(v)$ will denote expected utility net of the purchase price. We assume that all bidders are risk neutral, which simply means that they seek to maximize $\mu(v)$.

Let p(v) be the probability that an agent with reservation value v gets the asset. Given that the agent knows his or her own v, and that the agent knows that the reservation values of the other agents are drawn from a probability distribution, the agent can calculate the probability p(v) of getting the asset after submitting the bid that is optimal at equilibrium, given what the agent

knows. Therefore, the expected value of the asset is $v \times p(v)$. In words, it is the value the agent places on the asset multiplied by the probability of getting it. But if the agent wins the asset then he or she will have to make a payment to the seller, and we let e(v) denote the expected value of the payment at equilibrium. Finally, we have

$$\mu(v) = vp(v) - e(v).$$

This is the basic identity on which everything else hinges. It merely says that an agent's expected profit is equal to expected revenue minus expected cost.

Example 4.2: $\mu(v)$ and e(v) for a first-price auction with two bidders

Suppose that there are two bidders, person 1's reservation value is \$120, and the probability of winning is $\frac{1}{3}$ with the bid that is payoff maximizing at equilibrium for someone with v = 120. At equilibrium, a bid of $\frac{1}{2} \times 120 = 60$ is payoff maximizing for this individual. When this individual wins, the profit is 120 - 60 = 60, and that happens $\frac{1}{3}$ of the time. Hence $\mu(120) = \frac{1}{3} \times 60 = 20$. Now let's calculate $\mu(120)$ by using $\mu(v) = vp(v) - e(v)$: We have $p(120) = \frac{1}{3}$ by assumption. Person 1's payment when he or she wins is 60, so $e(120) = \frac{1}{3} \times 60 = 20$. Then $\mu(120) = 120 \times \frac{1}{3} - 20 = 20$.

There are auctions in which even the losers have to pay. An all-pay auction requires each participant to submit a bid. The winner is the high bidder, and the price is the amount that the winner bid. But all losers also pay the amounts that they bid. In that case, all bids will be depressed relative to a first-price auction. The fact that my bid is lower in an all-pay auction than it would have been in a first-price auction is due to the fact that I have to pay my bid even if I lose, *and* I know that the other bidders will be in the same position and hence will discount their bids. According to the revenue equivalence theorem, the total amount taken in by the seller on average will be the same in an all-pay auction as in a first-price auction.

Example 4.3: The winner has to pay four times the bid

Suppose that there are *n* bidders in an auction that awards the asset to the high bidder at a price equal to *four times* the winner's bid. No one else pays anything. It is easy to see why this auction will yield the same revenue as the first-price, sealed-bid auction. (We do not necessarily assume the uniform probability distribution.) Let $b_1, b_2, b_3, \ldots, b_n$ denote the equilibrium bids in a conventional first-price auction. Then $\frac{1}{4} b_1, \frac{1}{4} b_2, \frac{1}{4} b_3, \ldots, \frac{1}{4} b_n$ will be the equilibrium bids in the new auction. Here's why: An individual's probability of winning will be the same with both auctions: The probability that $b_i > b_j$ is the same as the probability that $\frac{1}{4} b_i > \frac{1}{4} b_j$. Moreover, the profit *if* you win is the same, because $v_i - b_i = v_i - 4 \times \frac{1}{4} b_i$. Therefore, $\frac{1}{4} b_j$ maximizes *i*'s expected payoff in the new auction if b_i maximizes *i*'s payoff in the first-price auction.

4.5 Proof of the revenue equivalence theorem

Assume a particular auction mechanism. (I'll let you know when we use the assumption that it is a standard one.) We put the spotlight on a particular bidder, and begin by proving that for a given reservation value the bidder's expected payoff is the same in any two auctions. We do this by showing that expected payoff depends only on the individual's reservation value and on the details of the probability distribution from which the reservation values are drawn, not on any details of the auction mechanism itself.

We can solve the identity $\mu(v) = vp(v) - e(v)$ for e(v). We obtain

$$e(v) = vp(v) - \mu(v).$$

In words, the difference between the expected value of the asset and the expected payoff from owning the asset must be the expected payment that one must make to have a chance of acquiring the asset.

So far we haven't said much, but we begin to make progress by considering the possibility that the agents can misrepresent their reservation values to increase their expected payoff. But if $\mu(v)$ is an agent's expected payoff at equilibrium, it must be the highest payoff that the agent can get, given what this agent knows about others. This agent may be misrepresenting his or her reservation value, but at equilibrium the agent does so in a way that maximizes the return. Consider a different strategy *s*, by which we mean adopting the strategy that would be optimal for someone with reservation value *s*. Let $\mu(v|s)$ denote the expected payoff to an agent with reservation value *v* given that this agent masquerades as someone with reservation value *s*. We have

$$\mu(v|s) = v \times p(s) - e(s).$$

Let's explain this formula. The agent is behaving as an *s* type, so he or she has to pay the amount e(s) that an *s* type would be required to pay. And the agent will win the asset with probability p(s). However, the agent's true reservation value is v, so her expected revenue is $v \times p(s)$. Consequently, $\mu(v|s) = v \times p(s) - e(s)$.

Example 4.4: Misrepresentation with a first-price auction with two bidders

Suppose that person 1's reservation value is \$120, in which case a bid of $\frac{1}{2} \times 120 = 60$ is payoff maximizing for this individual at equilibrium. If person 1 were to bid 50 he or she would be masquerading as an individual whose reservation value is s = 100. Therefore, $\mu(120|100) = 120 \times p(100) - e(100)$.

We know that $e(s) = s \times p(s) - \mu(s)$. Therefore

$$\mu(v|s) = v \times p(s) - s \times p(s) + \mu(s)$$

or

$$\mu(v|s) = \mu(s) + (v - s) \times p(s)$$

In words, the expected payoff to a v type from employing the strategy that would be payoff maximizing for an s type is the payoff $\mu(s)$ that an s type would get, plus the difference in the actual value of the asset to a v type (v - s) weighted by the probability that the individual would get the asset by employing the strategy of an s type.

Because $\mu(v)$ is the best that a type-v agent can do, we must have $\mu(v) \ge \mu(v|s)$. If we had $\mu(v|s) > \mu(v)$ then the agent would do better masquerading as a type-s agent than the agent does at equilibrium, contradicting the fact that $\mu(v)$ is her expected return at equilibrium, where each agent maximizes her expected payoff. (If the agent can do better, we can't be at equilibrium.) Therefore, $\mu(v) \ge \mu(v|s)$. Because $\mu(v|s) = \mu(s) + (v - s) \times p(s)$, we have

$$\mu(v) \ge \mu(s) + (v - s) \times p(s).$$
^[5]

Given *v*, this is true for all *s*.

Before returning to the formal argument, we pause to highlight the intuition behind our theorem: Suppose that we could establish $\mu(s + 1) = \mu(s) + p(s)$ for all *s*. Then for *s* = 0 we have

$$\mu(1) = \mu(0) + p(0)$$

and because $\mu(2) = \mu(1) + p(1)$ we can state that

$$\mu(2) = \mu(0) + p(0) + p(1)$$

after replacing $\mu(1)$ with $\mu(0) + p(0)$. And because $\mu(3) = \mu(2) + p(2)$ we have

$$\mu(3) = \mu(0) + p(0) + p(1) + p(2).$$

Continuing in this manner, we find that for any reservation value v we have

$$\mu(v) = \mu(0) + p(0) + p(1) + p(2) + p(v-2) + p(v-1).$$

This then gives us revenue equivalence because $\mu(0)$ is zero for all standard auctions, and for any *s* the probability p(s) is the same for all standard auctions: The probability of winning if your reservation value is *s* is just the probability that *s* is higher than any other bidder's reservation value.

But how do we prove that $\mu(s + 1) = \mu(s) + p(s)$ holds? If $\mu(s + 1) > \mu(s) + p(s)$ then $\mu(s)$ is strictly less than $\mu(s + 1) - p(s)$, which is "almost" $\mu(s|s + 1)$, the payoff to someone whose reservation value is *s* from adopting the equilibrium behavior of someone whose value is s + 1. Note that $\mu(s|s + 1) = \mu(s + 1) - p(s + 1)$. If we had p(s) = p(s + 1) then we would have $\mu(s|s + 1) = \mu(s + 1) - p(s)$, in which case $\mu(s + 1) > \mu(s) + p(s)$ is inconsistent with the fact that $\mu(s)$ reflects optimizing behavior of someone with reservation value *s*. If this person is optimizing, he or she can't do better by pursuing another strategy. Similarly, $\mu(s + 1) < \mu(s) + p(s)$ leads to a contradiction. But this argument depends on the equality of p(s) and p(s + 1). As it is, the two probabilities will be almost identical. The difference will be tiny compared to $\mu(s + 1)$ and hence

will not be large enough to invalidate the theorem. Now, return to the formal argument.

Suppose s = v - 1. In other words, suppose that agent *v* masquerades as an agent whose reservation value is v - 1. Then [5] becomes

$$\mu(s+1) \ge \mu(s) + p(s).$$
 [6]

When s = 0 we have

$$\mu(1) \ge \mu(0) + p(0).$$
[7]

When s = 1 statement [6] yields

$$\mu(2) \ge \mu(1) + p(1)$$
[8]

and because we already have inequality [7], we can replace $\mu(1)$ in [8] with $\mu(0) + p(0)$, which will be equal to $\mu(1)$, or smaller, and thus we have

$$\mu(2) \ge \mu(0) + p(0) + p(1).$$
[9]

(If $\mu(2)$ is at least as large as $\mu(1) + p(1)$, and $\mu(1)$ is at least as large as $\mu(0) + p(0)$, then $\mu(2)$ is at least as large as $\mu(0) + p(0) + p(1)$.)

Now suppose that s = 2. In that case inequality [6] reduces to $\mu(3) \ge \mu(2) + p(2)$ and we can replace $\mu(2)$ by the right-hand side of [9] without invalidating the inequality. Therefore

$$\mu(3) \ge \mu(0) + p(0) + p(1) + p(2).$$

In general, for any reservation value v we will have

$$\mu(v) \ge \mu(0) + p(0) + p(1) + p(2) + p(3) + \dots + p(v-2) + p(v-1).$$
[10]

Proof of statement [10]

We know that [10] is true when v = 1 because we established this as [7]. Suppose that [10] is true for all reservation values up to and including *t*. We want to show that it is also true for t + 1. From [6] we have

$$\mu(t+1) \ge \mu(t) + p(t).$$
[11]

By hypothesis, [10] is true for *t* so we also have

$$\mu(t) \ge \mu(0) + p(0) + p(1) + \dots + p(t-2) + p(t-1).$$
[12]

Now, substitute the right-hand side of [12] for $\mu(t)$ in statement [11]. We then get

$$\mu(t+1) \ge \mu(0) + p(0) + p(1) + \dots + p(t-2) + p(t-1) + p(t),$$

which is statement [10] when v = t + 1. We have established that [10] is true for t = 1, and that if [10] is true for arbitrary reservation value v then it is true for v + 1. This tells us that [10] is true for all values of v. We are trying to show that $\mu(v)$ depends only on $\mu(0)$ and on the probabilities $p(0), p(1), \ldots, p(v-2), p(v-1), p(v)$. We are halfway there by virtue of [10]. What we need is a statement that puts an *upper* limit on the magnitude of $\mu(v)$. Go back to [5] and suppose this time that s = v + 1. From [5] we get

$$\mu(s) \le \mu(s-1) + p(s).$$
[13]

If s = 1 for example, [13] tells us that

$$\mu(1) \le \mu(0) + p(1) \tag{14}$$

and for s = 2 we get

$$\mu(2) \le \mu(1) + p(2).$$
[15]

Now substitute the right-hand side of [14] for $\mu(1)$ in [15] to get

$$\mu(2) \le \mu(0) + p(1) + p(2).$$
[16]

For s = 3 we get $\mu(3) \le \mu(2) + p(2)$, and when we substitute the right-hand side of [16] for $\mu(2)$ we get

$$\mu(3) \le \mu(0) + p(1) + p(2) + p(3).$$

In general, we have

$$\mu(v) \le \mu(0) + p(1) + p(2) + p(3) + \dots + p(v-1) + p(v).$$
[17]

It is left to you to prove [17] in the same way that we established [10]: We know that [17] is true when v = 1, and thus proving that [17] is true for v + 1 if it is true for v gives us the general result.

Statement [10] gives us a lower bound on $\mu(v)$ and [17] gives an upper bound. Combining the two yields

$$\mu(0) + p(0) + p(1) + p(2) + \dots + p(v-2) + p(v-1) \le \mu(v)$$

$$\le \mu(0) + p(1) + p(2) + \dots + p(v-1) + p(v).$$

Therefore

$$\mu(0) + p(0) + p(1) + \dots + p(v-1) \le \mu(0) + p(1) + \dots + p(v-1) + p(v).$$
[18]

The left-hand side of [18] is almost identical to the right-hand side. To get the latter from the former we add p(v) and subtract p(0). Now p(0) is the probability of winning when your reservation value is zero. That probability will be zero, so the difference between the lower bound and the upper bound is the presence of p(v) in the latter. But that is not a big number at all—not relative to the sum of the other probabilities. Suppose that your reservation value is exactly \$1 million. Then p(1,000,000) is the probability of your winning when your reservation

value is exactly 1 million. Even if that number is not close to zero, each of the terms in the sum $p(900,000) + p(900,001) + p(900,002) + \cdots + p(999,998) + p(999,999)$ will be fairly close to p(1,000,000). The sum of 100,000 positive numbers that are significantly greater than zero will be rather large. In other words, p(v) will be tiny compared to the sum of the numbers that precede it. Therefore, the difference between the right-hand and the left-hand sides of [18] is very tiny.

Example 4.5: A simple illustration

Suppose that $p(1,000,000) = \frac{1}{2}$ but $p(900,000) = \frac{1}{4}$. Then the sum of the probabilities for reservation values between 900,000 and 999,999 inclusive cannot be smaller than $\frac{1}{4} \times 100,000 = 25,000$, and this is extremely large compared to p(1,000,000).

So we ignore the difference between the left-hand and the right-hand sides of [18], and say that $\mu(v)$ is "equal" to

$$\mu(0) + p(0) + p(1) + p(2) + p(3) + \dots + p(v-1) + p(v)$$

Formally,

$$\mu(v) = \mu(0) + p(0) + p(1) + p(2) + p(3) + \dots + p(v-1) + p(v).$$
[19]

We have proved that [19] is true—well, approximately true—for every value of v. Therefore, it is true for every bidder, whatever his or her reservation value v.

Equation [19] takes us to the threshold of the revenue equivalence theorem. Compare two auctions *A* and *B* that each award the asset to the buyer with the highest reservation value at equilibrium. It follows that for any reservation value *v*, the probability p(v) of winning is the same for the two auctions: If there are *n* bidders, p(v) is just the probability that *v* is higher than the other n - 1randomly drawn reservation values. That means that every term in the righthand side of [19] will be the same for the two auctions, except perhaps for $\mu(0)$. But if $\mu(0) = 0$ in both auctions, then that term will also be the same for *A* and *B*. (*Note that we have now assumed that A and B are standard auctions*.) Therefore, for any reservation value *v* the payoff $\mu(v)$ will be the same for each auction. This tells us that a buyer's expected payoff is the same in the two auctions, given his or her reservation value.

Recall our starting point: The identity $\mu(v) = v \times p(v) - e(v)$. We have proved that $\mu(v)$ is identical for the two auctions, given v. Because, by assumption, p(v) is the same for the two auctions, it follows that e(v) must be the same for the two auctions. (For any six numbers a, b, c, x, y, z, if a = b - c, x = y - z, a = x, and b = y, we must have c = z.) For a given bidder, for each possible value v of that bidder the expected payment e(v) is the same for any two standard auctions. Therefore, that bidder's expected payment averaged over all possible

reservation values must be the same for any two standard auctions. If every bidder's expected payment is the same for any two standard auctions, then the total of those expected payments over all bidders must be the same for the two auctions. But the total of the bidders' payments is the seller's revenue. We have proved that standard auctions *A* and *B* yield the same expected revenue.

Example 4.6: How good is the approximation?

There are two bidders, and the reservation value of each is no lower than α and no higher than ω . Let *m* be a large integer. Divide the set of numbers between α and ω (inclusive) into m - 1 subintervals of equal length. Set $\lambda = (\omega - \alpha)/(m - 1)$, the length of each subinterval. Including the left and right endpoints, the numbers α , $\alpha + \lambda$, $\alpha + 2\lambda$, $\alpha + 3\lambda$, ..., $\alpha + (m - 2)\lambda$, ω are the endpoints of the subintervals. We treat these *m* numbers as the possible reservation values. Assume that the probability that any one of these is the agent's reservation value is 1/m. The probability that a particular agent with reservation value *r* will win is just the probability that the other agent's reservation value is not higher. (To simplify the calculations we assume that a tie goes to the other agent.) If *t* denotes the value of the other agent, then the probability that our agent wins the asset when his or her reservation value is *r* is

$$prob(t = 0) + prob(t = 1) + prob(t = 2) + \dots + prob(t = r - 2)$$

+ $prob(t = r - 1)$.

Each of these probabilities is 1/m, and there are *r* of these terms, so

$$\operatorname{prob}(t < r) = r \times \frac{1}{m} = \frac{r}{m}.$$

This is the probability that our agent with reservation value r will win. Now, suppose that our agent's reservation value is actually $v = \alpha + k\lambda$ for some value of k. Note for a given value if v, the larger is m, the smaller is λ , and hence the larger is k. There are k reservation values below v, and thus the *sum* of the probabilities of winning from r = 0 through r = v - 1 is

$$\frac{1}{m} + \frac{2}{m} + \frac{3}{m} + \dots + \frac{k-1}{m} + \frac{k}{m} = \frac{k}{m} + \frac{k(k-1)}{2m} = \frac{k^2 + k}{2m}.$$

(See the box following the example.) This approximates the sum of the probabilities on the right-hand side of [18]. The difference is that we have added p(0) and subtracted p(v). That is, we have added 1/m and subtracted k/m. How much difference does the net subtraction of (k-1)/m make? The ratio of (k-1)/m to the sum $(k^2 + k)/2m$ is less than 2/k. For a given value of v, the integer k increases with m. Therefore, if m is sufficiently large the difference between the left-hand and the right-hand sides of [18] will be arbitrarily small.

Sum of an arithmetic progression

Let *S* denote the *n*-term sum $a + (a + d) + (a + 2d) + \dots + [a + (n - 2)d] + [a + (n - 1)d$. The first number is *a*, and every subsequent term is higher than its predecessor by the amount *d*. The first term plus the last term equals a + [a + (n - 1)d] = 2a + (n - 1)d. The second term plus the second-last term equals (a + d) + [a + (n - 2)d] = 2a + (n - 1)d. The third term plus the third-last term equals (a + 2d) + [a + (n - 3)d] = 2a + (n - 1)d. Continuing is this fashion, we establish that

$$2S = n \times [2a + (n-1)d]$$

and thus S = an + [n(n-1)d]/2.

$\partial 4.6$ Integral calculus proof of the revenue equivalence theorem

We begin with inequality [5] from the previous section:

 $\mu(v) \ge \mu(s) + (v - s) \times p(s)$. This holds for all v and s. Let v = s + ds. Then v - s = ds, and we have $\mu(s + ds) \ge \mu(s) + ds \times p(s)$; hence

$$\mu(s+ds)-\mu(s)\geq ds\times p(s).$$

If ds > 0 we can divide both sides of this inequality by ds without changing the direction of the inequality: We get

$$\frac{\mu(s+ds)-\mu(s)}{ds} \ge p(s)$$

As ds > 0 approaches zero, the left hand side of this inequality approaches the derivative $\mu'(s)$. This establishes that $\mu'(s) \ge p(s)$ holds for all *s*.

However, if ds < 0 then we do change the direction of the inequality when we divide ds into both sides of $\mu(s + ds) - \mu(s) \ge ds \times p(s)$. This yields

$$\frac{\mu(s+ds)-\mu(s)}{ds}\leq p(s).$$

As ds < 0 approaches zero through negative values, the left-hand side of this last inequality also approaches the derivative $\mu'(s)$, and thus $\mu'(s) \le p(s)$ for all s.

We have shown that $p(s) \le \mu'(s) \le p(s)$ for all *s*, and thus we must have $\mu'(s) = p(s)$ for all *s*. By the fundamental theorem of calculus, $\int \mu'(s) ds = \mu(s)$. Therefore, $\mu(s) = \int \mu'(s) ds = \int p(s) ds$. It follows that

$$\mu(v) = \int_0^v p(s) \, ds + C.$$
 [20]

C is a constant, which must be equal to $\mu(0)$ because, by [20], $\mu(0) = \int_0^0 p(s) ds + C = 0 + C$. Therefore,

$$\mu(v) = \int_0^v p(s) \, ds + \mu(0).$$
[21]

By assumption $\mu(0)$ is zero for any standard auction. And because any two standard auctions award the asset to the buyer with the highest reservation value at equilibrium, the probability p(s) of winning is the same for the two auctions,

for any value of *s*. It's simply the probability that *s* is higher than the reservation value of any other bidder. Therefore, $\mu(v)$ is the same for all standard auctions.

By definition, $\mu(v) = v \times p(v) - e(v)$, and we now know that $\mu(v)$ is identical for any two standard auctions. And because p(v) is the same for any two such auctions, it follows that e(v) is the same. Finally, if for each reservation value v, a bidder's expected payment e(v) is the same for the two auctions, then the total of those payments over all bidders must be the same for the two auctions. But the total of the bidders' payments is the seller's revenue. We have proved that any two standard auctions will yield the same expected revenue.

Sources

Section 4.6 is based on Klemperer (1999). The revenue equivalence theorem was discovered (as a special case) by Vickrey (1961), where a proof was also given. The general version first appeared simultaneously in Myerson (1981) and Riley and Samuelson (1981).

Links

Klemperer (2004) and Illing and Klüh (2003) take you deeper into auction theory and practice. The first book is a general treatment, and the second one is specifically devoted to the recent auctions of the radio spectrum in Europe.

Problem set

- **1.** Determine the seller's expected revenue in a first-price, sealed-bid auction for the following four cases:
 - **A.** There are two bidders and each reservation value is drawn from the uniform probability distribution on the interval from 0 to 5.
 - **B.** There are two bidders and each reservation value is drawn from the uniform probability distribution on the interval from 2 to 5.
 - **C.** There are four bidders and each reservation value is drawn from the uniform probability distribution on the interval 0 to 1.
 - **D.** There are four bidders and each reservation value is drawn from the uniform probability distribution on the interval 1 to 11.
- **2.** There are *n* bidders in an auction that awards the asset to the high bidder at a price equal to 20% of the winner's bid. No one else pays anything. Without using mathematics, and without appealing to the revenue equivalence theorem, explain why this auction yields the same revenue as the first-price, sealed-bid auction. (You should be able to do this without assuming the uniform probability distribution.)
- **3.** The following questions pertain to a first-price, sealed-bid auction with exactly two bidders. Each individual's reservation value is drawn from the uniform probability distribution on the interval 0 to 1. Calculate the three quantities e(v), $v \times p(v)$, and $\mu(v)$ for each of the five values of v listed. Calculate $\mu(v)$ in two different ways:

$$\mu(v) = v \times p(v) - e(v)$$

and $\mu(v) = (v \text{ minus this individual's bid}) \times p(v)$.

- **A.** v = 1.
- **B.** v = 0.
- **C.** $v = \frac{2}{3}$.
- **D.** $v = \frac{1}{4}$.
- **E.** Generic *v*. That is, leave the reservation value as *v*, so your answers will be functions of *v*.
- **4.** Why does the expected (i.e., average) revenue of the first-price, sealed-bid auction increase when the number of bidders increases?
- **5.** What is the expected (i.e., average) revenue from the first-price, sealed-bid auction when there are four bidders and each reservation value is drawn from the uniform probability distribution on the interval 0 to 1?
- **6.** What is the expected (i.e., average) revenue from the Vickrey auction when there are four bidders and each reservation value is drawn from the uniform probability distribution on the interval 0 to 5?
- **7.** What is the expected (i.e., average) revenue from the first-price, sealed-bid auction when there are four bidders and each reservation value is drawn from the uniform probability distribution on the interval from 1 to 11?
- **8.** Why does the expected (i.e., average) revenue of the second-price (Vickrey) auction increase when the number of bidders increases?
- **9.** What is the expected (i.e., average) revenue from the Dutch auction when there are two bidders and each reservation value is drawn from the uniform probability distribution on the interval 0 to 1?
- **10.** There are *n* bidders. What is the expected (i.e., average) revenue of the first-price, sealed-bid auction, assuming that the reservation values are drawn from the uniform probability distribution (with values between zero and unity, inclusive)? Explain briefly.
- **11.** Prove statement [11] in Section 4.5.
- **12.** Assuming exactly two bidders, construct a simple example of revenue equivalence between the first-price, sealed-bid auction and English auction when there are two possible pairs of reservation values—case 1 and case 2—and they are equally likely.
- **13.** Assuming exactly two bidders, construct a simple example of a *failure* of revenue equivalence between the first-price, sealed-bid and English auctions when there are only two possible pairs of reservation values—case 1 and case 2—and they are equally likely.
- **14.** Consider the following new auction mechanism: After the bidders have gathered, the auctioneer flips a coin. If it turns up heads then the (English) ascending auction is used, but if tails turns up then the (Dutch) descending auction is used. Will this new auction be revenue equivalent to the first-price, sealed-bid auction? Explain.

5

APPLICATIONS OF THE REVENUE EQUIVALENCE THEOREM

If any two standard auctions yield the same expected revenue, should the seller devote any effort to choosing or designing an auction mechanism? Yes. Many

The government (through the Resolution Trust Corporation) sold the assets of almost 1000 failed banks and savings and loan institutions in the 1980s and early 1990s. It solicited sealed bids but conducted the auctions prematurely, hence there were few bidders, and the auctions yielded far less than their potential (Sternberg, 1991).

A 1999 spectrum auction in Germany used the ascending bid format, but required a new bid to be at least 10% more than its predecessor. Mannesman opened by bidding 18.18 million deutschmarks on licenses 1 through 5, and 20 million deutschmarks on licenses 6 through 10. The only other credible bidder was T-Mobil, and its opening bids were lower. One of the T-Mobil managers reported that there was no explicit agreement with Mannesman, but the T-Mobil team understood that if it did not raise the bid on lots 6-10 then T-Mobil could have lots 1-5 for 20 million, which is slightly more that 10% greater than 18.18 million. That is in fact what happened: The auction ended after two rounds of bidding (Klemperer, 2004, pp. 104-5). A 1997 spectrum auction in the United States was expected to raise \$1.8 billion but realized only \$14 million. Bidders used the final three digits of their multimillion-dollar bids to signal the market code of the area that they intended to go after (Cramton and Schwartz, 2000).

practical issues do not arise within the abstract framework employed in the previous section. In particular, the revenue equivalence theorem takes the number of bidders and the absence of collusion for granted. In any particular sale, both issues should receive careful consideration. We deal with them in turn.

We expect the seller's revenue to increase with the number of bidders. All other things being equal, the seller should employ the auction mechanism that attracts the most bidders. An English auction can discourage entry, particularly if it is known that there are one or two bidders with very high reservation values. The weaker bidders know that they will be outbid and thus will not even compete. This leaves only two or three bidders, who will then be tempted to collude. However, a first-price, sealed-bid auction gives weak bidders at least a chance of winning, because everyone knows that every firm will submit a bid below its reservation value. A sealed-bid auction might even attract firms who have no intention of using the asset but simply hope to sell the asset for a profit after the auction. (It's hard to profit from resale in the case of an English auction. The winner will be the high-value agent, and so no one else would be willing to pay more than the asset is worth to the winner of the auction.)

The very advantage of sealed-bid auctions from the standpoint of encouraging entry low-value agents have a chance of winning—is a disadvantage from an efficiency standpoint. Paul Klemperer has proposed a middle ground, the *Anglo-Dutch auction*. The first stage is an English auction, which is allowed to run its course until only two bidders remain. These two then enter sealed bids, which must not be

less than the last bid from the English stage. The asset is sold to the high bidder in stage two for a price equal to that bid. (The sealed-bid stage is "Dutch" because the Dutch auction is outcome equivalent to the first-price, sealed-bid auction.)

5. Applications of the Revenue Equivalence Theorem

Collusion by the bidders can significantly depress the seller's revenue. The English auction is vulnerable to bidder collusion for two reasons: First, there are several rounds of bidding, so members of a cartel have a chance to punish a member who deviates from the cartel strategy. Second, the bids are not sealed and thus can be used to signal information to other bidders.

The analysis of Section 4 applies even to mechanisms that are not auctions in the conventional sense of the word. The revenue equivalence theorem is valid for any two mechanisms that use a formula for allocating a single asset—or something of value—provided that an agent with a reservation value of zero gets zero profit on average, and the agent with the highest reservation value always gets the asset at equilibrium. The mechanism can even allocate the asset randomly, as a function of the bids or messages submitted by the agents. Moreover, if you go back and check the proof, you'll see that we can weaken the assumption that an agent with a reservation value of zero gets zero profit on average. As long as the expected profit of an agent with the lowest possible reservation value is the same across two auction mechanisms, then they will generate the same expected revenue (provided that the agent with the highest reservation value always wins). Here are five significant applications of the theorem.

5.1 Multistage auctions

The U.S. and British airwaves auctions were designed by academic economists to allocate radio frequencies to companies selling personal communication devices and broadcast licenses. Bidding takes place in several rounds, and bidders can revise their bids after observing what happened on the previous round. The revenue equivalence theorem doesn't say anything, one way or another, about how many stages an auction can take, so it applies to multistage auctions. All that we need to know about an auction is the probability of winning, as a function of the reservation value, and the expected payoff of an agent whose reservation value is zero.

5.2 Adoption of a standard

There are auctions in which the losers have to make a payment, in addition to the winner. This is true of a "war of attrition," a term that covers a family of allocation problems not normally thought of as auctions. Suppose that *n* firms are lobbying the government, each to have its own technology adopted as the standard for the industry. For example, telecommunications firms compete for the prize of having their own technology for third-generation mobile phones adopted as the industry standard. The amount spent on lobbying is the bidder's payment. If we assume that the standard that's adopted is the one employed by the company that spends the most on lobbying, then we have an auction—call it auction *A*. The asset that the winner gets is the mandating of its technology for all firms in the industry. Let *B* refer to the second-price (or Vickrey) auction. We already know that *B* is standard. In the case of *A*, if you spend nothing then your technology will not be chosen as the standard. A firm's reservation value is the profit it expects to make from the adoption of its own technology as the standard. The higher the reservation value, the more a firm is willing to spend lobbying

376 Auctions

the government (or industry) committee that will make the decision. Therefore, in equilibrium the agent with the highest reservation value will get the prize, and hence *A* is also a standard auction. Therefore, the expected payment with auction *B* must also be the expected total payment (over all firms) for auction *B*. In other words, the total amount of money expected to be dissipated in the war of attrition is equal to the expected payment by the winner in a Vickrey auction. The latter will be much easier to compute.

5.3 Civil litigation

Currently, the laws governing civil suits in the United States require all contestants to pay their own expenses. If the law were changed so that the loser were also required to pay the winner an additional amount equal to the loser's expenses, would expenditures on lawsuits be reduced? A party would have to pay more if it lost, but every additional dollar paid by the loser is an additional dollar gained by the winner, so the expected value of a lawsuit might not change. In fact, even under the new rule, if a party spent nothing it would not win and thus would gain nothing. So the first part of the definition of a standard auction is confirmed. If we assume that the party that spends the most wins the suit, then we also have the second part. Therefore, the two systems result in the same total expenditure on civil suits. Because the expected profit from a lawsuit is the same for the two systems, the incentive to bring an action is the same. Hence the same number of lawsuits are contested in the two systems, so total expenditure really would be the same.

5.4 Procurement

When a government or a firm puts a contract up for bids, the winning bidder will have to deliver an asset (i.e., it will have to construct a hospital, road, or office building, etc.), and in return the winner is *paid* an amount of money. This is an auction in reverse. The bidder's reservation value v is a cost (the cost of construction) and thus it is a negative number. The winner's payment is negative (it is a receipt) so e(v) is negative. If we let c denote cost and r(c) denote the *bidder's* expected revenue as a function of the bidder's cost c we have

$$\mu(c) = r(c) - c \times p(c),$$

which says that expected payoff is equal to the bidder's expected revenue minus expected cost.

Now, let c = -v and r(v) = -e(c). Then $\mu(c) = r(c) - c \times p(c)$ becomes

$$\mu(v) = v \times p(v) - e(v),$$

which was the starting point for the revenue equivalence theorem, which now tells us that the winner's expected profit will be the same and the government's expected expenditure will be the same in any two procurement auctions in which the low-cost supplier always wins the contract at equilibrium, and the highest cost supplier will get zero profit at equilibrium.

6. Interdependent Values

5.5 Car sales

There is great enthusiasm in Europe for Internet sales of automobiles as a substitute for dealer sales. They are gaining popularity in the United States. Prices are more transparent on the Internet, and the assumption is that consumers benefit from this because of the reduction in search costs. The assumption may be wrong. Certainly there are many more sellers competing for a given customer's favor in an Internet sale. This means that Internet sales approximate the standard English, oral ascending auction, whereas purchase at a dealership is similar to a first-price, sealed-bid auction. The traditional dealership sale is a sealed-bid auction because the buyer has no way of credibly reporting one dealer's offer to another, particularly when dealers so rarely put an offer in writing. The offers are, in effect, sealed.

An internet sale to one customer can be treated as a separate auction. The bidders are sellers, not buyers, and the bids are lowered until only one seller remains—the car is then sold at that survivor's bid. So it is a procurement auction and, after inserting minus signs, equivalent to an ascending auction. The revenue equivalence theorem tells us that the expected outcome is the same in the two situations. However, that theorem assumes away *collusion* on the part of the bidders. Collusion among automobile sellers is much easier to orchestrate in the case of Internet sales. Early rounds can be used to signal information from one seller to another. Finally, sealed-bid procurement auctions typically generate lower prices for the buyer.

Source

These examples are drawn from Klemperer (2003). Klemperer (1998) proposed the *Anglo-Dutch auction*.

Links

Chapters 3 and 4 of Klemperer (2004) provide insight into the practical side of designing an auction. The former is also available as Klemperer (2002a). Paul Klemperer played a central role in the design of the British spectrum auctions. John McMillan and Paul Milgrom played a key role in designing the U.S. spectrum auction. Milgrom (2004, Chapter 1) discusses the practical side of auction design, and McMillan (2002, Chapter 7) is a superb account of modern auctions.

6 INTERDEPENDENT VALUES

Up to this point we have assumed that the auctions in question apply to private values cases, by which we mean each agent has a reservation value that is statistically independent of the reservation value of any other agent. The pure common values case is the opposite: The asset has a market value that is the same for all bidders, but no bidder knows the true value. Each bidder has some information about the true value, but the information is different for different agents. The intermediate case is encountered most often: The asset is worth more to some firms than to others—perhaps because the former have other resources that

378 Auctions

combine well with the asset that is up for sale—but the values of the different bidders are closely related, perhaps because the asset would be used to serve a

The value of a TV license is higher for bidder *A* than bidder *B* because *A* owns a baseball franchise and can use the TV station to hide some of the profit from baseball operations, say by buying broadcast rights from the baseball team at below market value. This would allow the owners of the baseball team to tell the players that they can't have a salary increase because there is no profit (Chapter 4 of Zimbalist, 2004). particular consumer group, such as cell phone users.

The standard example of the pure common value situation is the competition between oildrilling firms for the right to extract crude oil from a presently undeveloped tract of land. Each firm will employ experts to estimate the size of the underground reserve and the cost of extracting it. The estimates won't agree, and the data will not be made public until the bidding is over—if then. Each firm keeps its estimate to itself. There will be *some* publicly available information—perhaps the amount of oil

pumped from nearby land. Hence, each firm has a *signal* (i.e., estimate) of the value of the tract, based on public information and its private information, and the signals are different for different firms.

DEFINITION: Bidder signal

Firm *i*'s signal σ_i is its estimate of the asset's worth to *i* itself.

In a private values auction a firm's signal is just its own reservation value. Even if the firm knew the reservation values of the other firms, its own estimate

The techniques for estimating the oil reserves trapped in a geological formation are significantly more reliable than a hundred years ago, but geologists can still disagree about a particular tract. Thousands of licenses for drilling oil off the coasts of Louisiana and Texas have been auctioned by the U.S. Department of the Interior, but many of the sales attracted only one bid (McAfee, 2002, pp. 307–8). of the asset's value would be unaffected. (But the information would be useful in guessing how much the other agents would bid.) In a pure common value auction the firm's signal is its own particular estimate of the value of the asset. The asset is worth the same amount to each bidder—the common value—but that number is not precisely known by any bidder. Each bidder knows that the average of all the bidders' estimates would be a much more reliable indicator of the value of the asset, but no firm will know any other firm's estimate before bidding begins.

In this last section we abandon the private values assumption and reconsider auctions when one agent's signal embodies information that is relevant to the value of the asset to other agents. Happily, much of what we learned in the private values case can be adapted to the general framework.

For the pure common value case, in which the asset has precisely the same (unknown) value to all the bidders, asset efficiency is satisfied by *any* assignment of the asset! This suggests that an auction is unnecessary; just allocate the asset

6. Interdependent Values

by lottery. On the contrary, there are two reasons why an auction should still be used. The first is a fairness argument. If the asset (e.g., oil reserves) belongs to all citizens, it shouldn't be given free to any of them. A competitive auction would eliminate much (sometimes all) of the excess profit from asset ownership. Second, even if the asset may have the same value to all bidders, this won't be the case for all firms in the economy. Firms that have no expertise in exploiting the asset would not be able to realize the "common" value. An auction attracts only firms that can realize the asset's potential. But if the value is the same for all *bidders*, then asset efficiency is satisfied, regardless of the outcome.

For situations that are intermediate between the pure private values and the pure common value case, asset efficiency can be problematic, as the next example demonstrates.

Example 6.1: Interdependent values

There are three bidders, called 1, 2, and 3. Their respective signals are σ_1 , σ_2 , and σ_3 . The values v_1 and v_2 of the first two bidders each depend on all three signals, but the value v_3 of the asset to agent 3 is a function of 3's signal only. Specifically:

$$v_1 = \sigma_1 + \frac{2}{3}\sigma_2 + \frac{1}{3}\sigma_3$$
, $v_2 = \sigma_2 + \frac{1}{3}\sigma_1 + \frac{2}{3}\sigma_3$, and $v_3 = \sigma_3$.

The asset is to be auctioned, and each firm has to submit a bid before knowing the signal received by the other two. Suppose that $\sigma_1 = \alpha = \sigma_2$ and $\sigma_3 = \alpha + \epsilon$, where α is positive, although the random variable ϵ can be positive or negative. We have

$$v_1 = 2\alpha + \frac{1}{3}\epsilon$$
, $v_2 = 2\alpha + \frac{2}{3}\epsilon$, and $v_3 = \alpha + \epsilon$.

If $\epsilon < 0$ then $v_1 > v_2 > v_3$, in which case asset efficiency requires that agent 1 is the winning bidder. If $0 < \epsilon < 1.5\alpha$ then $v_2 > v_1 > v_3$, and asset efficiency is satisfied only if agent 2 is the winning bidder. Suppose that an agent's bid can depend only on his or her own signal. (That would be the case in any sealed-bid auction.) Then when $\sigma_1 = \alpha = \sigma_2$ neither the bid of 1 or 2 will be influenced by the sign of ϵ . If agent 1 is the winning bidder then asset efficiency is violated when $1.5\alpha > \epsilon > 0$, and if agent 2 is the winning bidder then $\epsilon < 0$ is inconsistent with asset efficiency, whatever sealed-bid auction is used.

From now on we will confine our attention to the common value case. The asset has the same value to each bidder, but the common value is unknown to each.

6.1 **Revenue equivalence**

For the family of sealed-bid auctions, the revenue equivalence theorem is valid for the common value case. All we have to do to prove this is replace the agent's reservation value v in the proof of Section 4.5 (or 4.6) with the agent's signal σ . (In the case of an open auction, a bidder make inferences about the signals received

380 Auctions

by others when they hear their bids. Thus, replacing the reservation value by the signal is invalid.) The bidder's behavior is now based on σ instead of v, but the mathematics will not have to be changed—after replacing v with σ (or even better, interpreting v as the agent's signal). It needs to be emphasized, though, that the assumption that each bidder's signal is statistically independent of any other agent's signal is crucial. This rules out intermediate interdependent values cases, such as Example 6.1.

The common value version of the revenue equivalence theorem

If each of the *n* agents is risk neutral, and each has a privately known signal independently drawn from a common probability distribution, then all standard sealed-bid auctions have each bidder making the same expected payment at equilibrium, given his or her value, and thus the seller's expected revenue is the same for all standard auctions.

6.2 The winner's curse

Before bidding, each potential buyer hires a team of experts. In the case of oil drilling, the bidder will employ a team of geologists to determine how much oil is under the tract of land up for auction and how difficult it will be to extract the oil. Economists will also be called on to determine the future market value of oil. Each bidder hires a different team of experts, and the estimates of the asset's market value will disagree. There are three reasons why the potential buyers will *not* exchange their information before the auction is run. First, if they did then each buyer would have little incentive to fund research because it would get the results of others' estimates for free. Second, bidder *A*'s estimate would help bidder *B*, but *A*'s goal is to profit at the expense of *B*. Third, each bidder would have an incentive to mislead the others.

Each bidder has an estimate of the common value—the value of the asset and no bidder will know the estimates obtained by the others. Some of the estimates will be on the high side and some will be on the low side. The average of all the estimates will be a good approximation to the common value, but the highest estimate will not. But an agent's bid will be proportional to the estimate that that agent has obtained. It follows that the high bidder will be the one with the highest estimate. Therefore, as soon as an agent learns that he has won, he knows that he has paid too much, because he had the extreme estimate (on the high side) of the value of the asset.

Example 6.2: Three hats

In this instance the asset is a hat containing five pieces of paper. Each of the five slips has a number on it, and if you obtain the asset by outbidding your rival you will be paid an amount of money equal to 100 times the average of the five numbers. You are allowed to sample before deciding how much to bid.

6. Interdependent Values

Specifically, you can draw one piece of paper at random from the hat and look at it. You have to replace it before your rival, who cannot see what number you drew, takes a sample. And you won't know what number your rival drew before entering your bid. Moreover, neither of you know which of three hats is the one you are bidding for: Hat A contains the numbers 1, 2, 3, 4, and 5, so that asset is worth \$300. Hat B contains 2, 3, 4, 5, and 6, so it is worth \$400. Hat C is worth \$500 because it contains the numbers 3, 4, 5, 6, and 7. You draw a 4. How much should you bid? Four hundred dollars is the value of asset *B*, so it might be well to suppose that you are bidding for hat B. Much more useful information would be obtained by averaging your sample with the other person's. (If the average of the two draws is 6.5 then the asset is certainly hat C. If the average is 5.5 then it is certainly not hat A.) But comparing sample values is against the rules. Now, you submit your sealed bid, based solely on the information you possess, and you are told that you won the asset because you are the high bidder. That means that your rival drew a lower number than you and submitted a lower bid. Drawing a number smaller than 4 is much more likely to happen when sampling from A than from C. It is 50% more likely with A than with B. Conditional on winning the first-price, sealed-bid auction a draw of 4 should lead to a bid below \$300 because the chances are good that you are competing for asset A. If you draw the number 4 and bid, say, \$325 on the supposition that the asset is more likely to be *B* than *A* or *C* you have a good chance of experiencing the winner's curse. (When two samples are taken from hat A the average low draw is 2.28 and the average high is 3.8. When two samples are taken from B, the average high is 4.8, although the mean of the numbers in hat *B* is 4.)

Any mechanism in which firms or individuals compete with each other for a single asset, or a handful of assets, can be viewed as an auction. Accordingly,

Oil companies appear to have fallen victim to the winner's curse during the auctions for offshore oil-drilling rights. Book publishers often feel that by outbidding rivals for the right to publish a book they have paid more than they will ever recoup in profits from book sales. Baseball teams have often outbid other teams for a free agent only to find that they have paid too much for the player's services. The phenomenon also occurs in corporate takeover battles (Thaler, 1992, pp. 57–8 and Dyer and Kagel, 2002, p. 349). the winner's curse can emerge in a wide variety of market contexts.

If the bidders *did* exchange their information before the auction they could produce a common estimate of the asset's market value. The average of all the estimates can be expected to be as close as anyone could forecast to the actual market value of the asset. The buyer whose team of experts produced the lowest estimate of the asset's value would know that its team underestimated. More significantly, the buyer whose team produced the highest estimate of the asset's value would know that its team overestimated. But that buyer would be the one with the highest reservation value going

into the auction and hence would be the winning bidder. Without an exchange of information, the appropriate strategy is to adjust one's estimate, and hence one's bid, to avoid falling victim to the winner's curse.

Example 6.3: Correcting for the overestimate

There are two bidders (to keep the calculations simple), and each draws a single sample from a probability distribution. Each knows everything about the distribution except the mean μ , which is also the common value of the asset. An individual will draw μ + 0 with probability 0.4, μ + 100 with probability 0.2, μ - 100 with probability 0.2, μ + 200 with probability 0.1, and μ - 200 with probability 0.1 as summarized by Table 6.3. The average draw is μ , but typically one of the bidders will have an above-average draw and the other will have a below-average estimate. Because neither knows the value of μ , or the other's estimate, neither will know if his or her own estimate is too high or too low. Let's calculate the average high estimate. Table 6.4 displays the probability 0.2 and *B* will draw μ + 200 with probability 0.2. and *B* will draw μ + 200 with probability 0.5. A model of the bidders will have μ + 0.0 with probability 0.2 and *B* will draw μ + 200 with probability 0.2. and *B* will draw μ + 200 with probability 0.5. The probability of both happening is $0.2 \times 0.1 = 0.02$, and when it does the high estimate is μ + 200. You can use Table 6.4 to calculate the average high estimate, which is μ + 60. Therefore, bidders should neutralize the winner's curse by subtracting \$60 from their estimates.

Note that the winner's curse becomes more severe as the number of bidders increases. That's because as the number of bidders increases, so does the

Table 6.3

$egin{array}{cccc} \mu-200 & 0.1 \ \mu-100 & 0.2 \ \mu & 0.4 \ \mu+100 & 0.2 \end{array}$	Draw	Probablity
$\mu + 200 \qquad 0.1$	$\mu - 200 \ \mu - 100 \ \mu \ \mu + 100 \ \mu + 200$	0.1 0.2 0.4 0.2 0.1

probability of someone drawing an extremely high sample estimate. In the case of Example 6.3 the probability of someone drawing either $\mu + 100$ or $\mu + 200$ when there are only two agents is 1 minus the probability that both individuals draw one of the other three numbers. The probability of a single individual drawing one of the three lowest numbers is 0.1 + 0.2 + 0.4 = 0.7. Therefore the probability of one of the two bidders drawing an estimate greater than μ is $1 - 0.7 \times 0.7 = 0.51$. (Alternatively, eliminate the last two columns and the last two rows of Table 6.4, add the remaining numbers, and

then subtract the result from 1.) If there are three agents, the probability of someone drawing $\mu + 100$ or $\mu + 200$ is $1 - 0.7 \times 0.7 \times 0.7 = 0.657$. With four bidders the probability increases to 0.7599.

Table	6.4
-------	-----

	B's estimate				
A's estimate	$\mu - 200$	$\mu - 100$	μ	$\mu + 100$	$\mu + 200$
$\mu - 200$	0.01	0.02	0.04	0.02	0.01
$\mu-100$	0.02	0.04	0.08	0.04	0.02
μ	0.04	0.08	0.16	0.08	0.04
$\mu + 100$	0.02	0.04	0.08	0.04	0.02
μ + 200	0.01	0.02	0.04	0.02	0.01
6. Interdependent Values

Is the winner's curse more dangerous with the first-price, sealed-bid auction or with the English auction? The former. If you seriously overestimate the asset's value in an English auction, you won't pay too much if you are the only one to overestimate, because you won't have to pay more than the second-highest estimate of its value. Moreover, as the auction proceeds you acquire information about some of the other estimates. As bidders drop out, you get an upper bound on their estimates of the asset's value.

Sources

Example 6.1 is from Maskin (2003).

Links

McAfee (2002, pp. 307–1) has some valuable observations on the winner's curse. The winner's curse emerges in laboratory experiments, but with experience the subjects learn to mitigate its effects, so that the winner realizes some profit while still bidding too high. Experimental subjects submit bids that are closer to the Nash equilibrium levels in English auctions than in first-price, sealed-bid auctions. See Kagel and Levin (2002b).

Voting and Preference Revelation

1.	Voti	ng Schemes	. 385
	1.1	Majority rule	386
	1.2	Other voting schemes	391
	1.3	Value-restricted preferences	396
		Problem set	400
2.	Pref	erence Revelation in General	. 402
	2.1	The Gibbard-Satterthwaite Theorem	404
	2.2	Proof for two individuals and three alternatives	405
		Problem set	410
3.	Gen	eral Proof of the Gibbard-Satterthwaite	
	The	orem	411
	3.1	Proof for two individuals and three or more	
		alternatives	412
	3.2	Proof for two or more individuals and three or more	
		alternatives	414
		Problem set	417
4.	The	Revelation Principle	418

This chapter examines decision making by a *community* (or any group) in a simple model: The community must choose from a finite set of mutually exclusive alternatives. (The next chapter endows the model with much more structure by specifying individual utility functions and a production function—and resource constraints. The utility functions will have classical economic properties.)

We look at situations in which a group must make a decision that will be binding on all of its members. For example, a class has to determine a time for a review session, a town has to decide whether to build a new school, a nation has to elect a legislature. The resulting choice will have no other implications for personal consumption—in this chapter. Think of the alternatives *X*, *Y*, *Z*, and so forth from which a choice is to be made as alternative ways of spending a fixed amount of government revenue, with the same individual tax burdens in each case. In this setting we can't rule out any ranking of the available options as a possible preference scheme for a member of the group. This makes it very difficult to induce truthful revelation of the hidden characteristic, which in this chapter and the next is the individual's true preference scheme. We want the individuals to reveal enough information about their preferences to enable the system to select the outcome that best reflects those individual preferences.

Clearly, one could write an entire book on the criteria for determining the alternative that "best" reflects individual preferences. Hundreds have been written on that theme. In this chapter we set that issue aside and simply determine which selection rules elicit truthful information about individual preferences. A selection rule is essentially a mapping from individual preferences into a social choice, and it must be defined for each possible specification of individual preferences.

1 VOTING SCHEMES

Although this section examines voting procedures, you are encouraged to think of the candidates standing for election not as individuals seeking careers in government but as alternative packages of public projects. Candidate (or alternative) *X* may, for example, be a proposal to reduce expenditure on the space shuttle program by a specific amount and to use the proceeds to fund research on the production of energy by nuclear fusion. Alternative *Y* may be a proposal to maintain the level of expenditure on the space shuttle while increasing federal expenditure on health care at the expense of grants to university professors. Other candidates or alternatives may present mixtures of *X* and *Y*, and each would be identified by its own label, *W*, *Z*, and so forth. The voters may be the members of the legislature or even the citizens themselves who are asked to vote directly for public projects.

The fact that citizens do not presently vote directly on these matters should not stand in our way. The observation that a particular voting scheme is not widely used is not a persuasive argument against its introduction if it would serve society better than the methods presently in use. When examining alternative

voting systems our main concern is the extent to which a voting procedure embodies incentives that induce individuals to mark their ballots to reflect their preferences in the way prescribed by the systems rules. (Other criteria are important, of course, but this book examines incentives.)

1.1 Majority rule

Our aim is to employ a voting scheme that induces individuals to reveal their preferences truthfully. Majority rule meets the case *when there are only two alternatives on the agenda*: Assume that there are two alternatives, *X* and *Y*, and the one that is preferred by a majority will be selected. (Some tie-breaking rule is employed when necessary.)

Majority rule theorem for two alternatives

Majority rule induces truthful preference revelation—even by coalitions—if there are only *two* alternatives.

Proof

If *X* wins but Vince voted for *Y* because he prefers *Y* to *X* then Vince cannot precipitate an outcome that he prefers to *X* by changing his vote. With only two alternatives Vince can only change his vote to *X*, but *X* already wins even when he votes for *Y*. Note that we can apply this argument to a coalition of individuals: If everyone in group *G* prefers *Y* to *X* but *X* wins the election when the members of *G* vote for *Y* over *X*, then there is nothing they can do individually or as a group to secure the election of *Y*. If there are only two alternatives, then majority rule always renders a decision that is invulnerable to manipulation by misrepresentation of preference.

When the feasible set has only two alternatives, *X* and *Y*, the majority winner is efficient: If alternative *X* is not efficient then no one strictly prefers *X* to *Y* and at least one person strictly prefers *Y* to *X*. But then *Y* would defeat *X* in the election.

Even if there is a large number of alternatives we can induce truthful revelation by selecting two alternatives, X^* and Y^* , and have the rule select the majority winner between X^* and Y^* . According to this decision scheme, the other alternatives (*Z*, *A*, *B*, etc.) don't have a chance, even if everyone ranks *Z* at the top and X^* and Y^* either last or second last. Disqualifying all but two of the alternatives from ever receiving consideration, whatever the individual preferences, is an extremely poor way to solve the hidden information problem.

The majority rule theorem is of no comfort in an economic context because there are always more than two alternatives. For one thing, a given public project can be financed in an astronomically large number of ways, by adjusting individual tax burdens. To see how this affects our evaluation of majority rule we revisit the case of two alternatives but use utility functions to express individual preferences.

Example 1.1: Two projects F and G and three individuals

Table 7.1 represents individual preferences by means of the utility levels U_i (net of any taxes paid). If the two projects come up for election then *G* will win because Rosie and Edie both prefer *G* to *F*. (If you like, you can suppose there are 3 million individuals, with each voter of the table belonging to a group of 1 million individuals with identical preferences.) Outcome *G* is inefficient, however, if we expand the feasible set by allowing side payments. If Soren were to compensate

Table 7.1

Table	7.	2
-------	----	---

	Soren	Rosie	Edie		Soren	Rosie	Edio
$U_i(F)$ $U_i(G)$	5000 3000	100 105	150 160	$U_i(F)$ $U_i(G)$ $U_i(H)$	5000 3000 4800	100 105 200	150 160 250

the other two individuals by paying them each \$100 for throwing their support to *F* we have a new outcome *H*, which everyone prefers to *G*, as illustrated by Table 7.2. That is, we create *H* from *F* by reducing Soren's consumption of the private good by 200 and increasing the private good consumption of Rosie and Edie by 100 each. (There is an implicit assumption here of quasi-linear preferences that justifies the claim that \$100 would be sufficient compensation.)

We have uncovered a general principle: If we are using majority rule in an economic context, and we restrict the agenda to two alternatives to ensure truthful revelation of individual preference, then we could wind up with an inefficient outcome. Note that we could modify Example 1.1 so that the difference between total utility from F and the total utility from G is as large as we like.

Majority voting with a two-alternative agenda can select an inefficient outcome because an individual who cares very little about the choice between two projects is given the same number of votes as an individual who has a great deal at stake. The intensity of an individual's preference is not recorded, and therefore there is no way to ensure that the majority's gain outweighs the minority's loss. We attempt to correct this defect by considering a voting scheme that allows individual to cast a variable number of votes.

Example 1.2: A variable number of votes

Each individual *i* is required to report $U_i(F)$ and $U_i(G)$ when *F* and *G* are the two projects on the ballot. The net benefit figure reported by *i* can be interpreted as the number of votes cast by *i* for the project in question, and all individuals are required to cast a number of votes equal to their net utility. If the project that receives the largest total vote is selected *and* individuals report truthfully then

an inefficient outcome will not survive this new voting process: If the private goods consumption at F can be rearranged so that all individuals have more utility than they have at outcome G then total utility at F exceeds total utility at G.

Of course individuals have a strong incentive *not* to report truthfully. Suppose the first table of Example 1.1 gives the true utility levels for the two alternatives F and G on the agenda. If Rosie declares that her net utility from G is 5105 (i.e., casts 5105 votes for G) and in all other cases each individual reports truthfully, the result is displayed in Table 7.3. Inefficient outcome G will receive the most

Та	bl	е	7	.3

	Soren	Rosie	Edie
$U_i(F)$	5000	100	150
$U_i(G)$	3000	5105	160

votes. (Remember, *G* is inefficient with respect to the true preferences if side payments are allowed and preferences are quasilinear.) Of course this voting pattern does not constitute an equilibrium. All individuals have an incentive to overstate their preference for the project they prefer, however slight their intensity of preference for one option over the other, and there is no limit to the number of votes individuals would be willing to cast for their preferred outcome in the case of this voting mechanism.

Suppose that we modify Example 1.2 by imposing an upper limit. Let's consider the implications of a cap of 10,000 on the number of votes that one can cast for an outcome. An individual is asked to cast a number of votes for an alternative in accord with the net utility that the individual would receive from that alternative, but in no case can he or she exceed 10,000 votes. Then all individuals will cast exactly 10,000 votes for the option that they prefer and 0 votes for the other project, so the mechanism imitates majority rule. (The number of votes cast has no effect on the tax formula used to finance the winning project—the financing formula is part of the definition of a project.) Each alternative will now receive 10,000 times as many votes as under majority rule, so the two mechanisms yield the same decision. The problem with this new scheme is that there is still insufficient restraint on individuals' desire to overstate the benefit derived from their preferred alternative, even if it gives them only slightly more utility than the other. Therefore, we want to consider a modification of the variable number of votes model in which there is a built-in incentive not to exaggerate. We don't investigate this until Chapter 8.

To forestall the potential inefficiency of majority rule let's have all of the feasible alternatives on the ballot. From now on we assume that there are three or more feasible alternatives. It is certain that a majority winner will be efficient if

every feasible alternative is on the ballot: If X defeats every other alternative by a majority then there is no other feasible alternative Y that everyone prefers to X. If there were such an alternative then it would have defeated X by a majority—an overwhelming majority. For the same reason, there can be no other feasible alternative Y such that some people prefer it to X and the rest are indifferent.

Efficiency of majority rule

If every feasible alternative is on the ballot and *if* there is an alternative that defeats every other by a majority then that alternative is efficient.

The winner from Example 1.1 was not efficient because there were feasible alternatives not on the ballot, which offered only two outcomes, F and G. However, even with only three alternatives, majority rule can fail to deliver a clear winner, as we are about to see. For the rest of this chapter, we represent individual preference as a ranking of the feasible alternatives.

DEFINITION: Individual preference ordering

We represent an individual's preference scheme as a ranking of the alternatives in order of preference: most preferred, second most preferred, and so on. We refer to it as an *ordering* (or ranking) and display it as a column, with the most preferred alternative at the top and the other alternatives (or outcomes) arranged below in order of preference.

Example 1.3: A majority rule cycle with three feasible alternatives

Table 7.4 illustrates how we use an ordering to represent individual preferences. There are three voters. Person 1 prefers *X* to *Y* and *Y* to *Z* (and hence *X* to *Z*).

	Та	ble	7.4	
--	----	-----	-----	--

Person 1	Person 2	Person 3
X	Y	Z
Y	Z	X
Z	X	Y

Person 2 prefers Y to the other two policy options and so on. There is such diversity of individual preference here that it is far from obvious how one of these

In 1956 the U.S. House of Representatives voted on a program of grants for school construction—option G. An amendment was introduced to deny aid to any state with segregated public schools—option A. The third alternative was the status quo, S. Southern Democrats preferred G to the other two but also preferred *S* to *A*. A second group, mostly northern Democrats, preferred A to G and G to S. The remaining representatives, mostly Republicans, preferred S to either of the other alternatives but preferred A to G. The three groups were roughly equal in size; hence a voting cycle. The House voting rules precipitated *S* as the outcome in this case (Brams, 1976, p. 44).

policies can be selected as the best for this three-person group. Each person has a different most-preferred alternative, and any individual's first choice is someone's last choice. The alternative that defeats the other two by a clear majority is the one that will be selected. Consider first the contest between Y and Z. Persons 1 and 2 will vote for *Y* over *Z*. (You can imagine three groups of roughly equal size if you think the case of three individuals is too unrealistic.) But Y will not defeat both of the other two alternatives by a majority. Persons 1 and 3 will vote for X over Y. This seems to leave us with X as the group choice. But alternative Z defeats X by a clear majority, with persons 2 and 3 both voting for Z in preference to X. We say that there is a majority rule cycle: X beats Y, Y beats Z, and Z beats X.

We can fix the indecisiveness problem by augmenting majority voting to incorporate a supplementary rule for selecting one of the alternatives in case of a cycle. However, one of the implications of the theorem that we prove in Section 3 is that there is no way of doing this without creating opportunities for someone to profit by misrepresenting his or her individual preference, as illustrated by the next example.

Example 1.4: The status quo receives special treatment

If there is an overall majority winner, then that will be the outcome. If there is a cycle and hence no overall winner, as with Example 1.3, then the outcome will be *X*, which we take to be the status quo. On one hand, suppose Table 7.5 gives

|--|

Person 1	Person 2	Person 3
X	Y	Z
Z	Z	X
Y	X	Y

the true individual preferences. Then the outcome is *Z* because *Z* defeats *X* by two votes to one (persons 2 and 3 prefer *Z* to *X*) and *Z* defeats *Y* by two votes to one (persons 1 and 3 prefer *Z* to *Y*). But if individual 1 does not report truthfully and declares a preference for *X* over *Y* and *Y* over *Z* then the reported preference pattern is the one of Table 7.4: There will be no overall majority

winner. Alternative *Y* defeats *Z* by two votes to one, *X* defeats *Y* by two votes to one, and *Z* defeats *X* by two votes to one. According to the definition of our new decision rule the outcome will be *X*. Because person 1 actually prefers *X* to *Z* person 1 has profited by misrepresenting his or her preference scheme. (Note that the preferences of persons 2 and 3 are identical in the two situations.) On the other hand, suppose that Table 7.4 gives the true individual preferences. Then the outcome is *X* if everyone reports truthfully. However, if person 2 reports an ordering with *Z* on top then *Z* will be the majority winner because *Z* will be the top alternative for 2 and 3. Then *Z* will be selected, and person 2 prefers *Z* to *X* according to his or her true preferences. In that case person 2 has benefitted by misrepresentation.

When the preferences are as in Table 7.4, whatever the outcome, it will be someone's last choice. That's the source of the incentive to misrepresent. (For instance, if person 3 were the chair and decided the outcome in case of a cycle, then Z would be chosen and that is person 1's lowest-ranked alternative.)

When an individual precipitates the election of a different alternative by deviating from truthful revelation we say that the person has *manipulated* if he or she prefers that alternative to the one that would have been selected had he or she reported truthfully.

DEFINITION: Manipulation

We say that individual *i* can *manipulate* the voting rule if there is some assignment p of individual preferences at which individual *i* can profit by reporting a preference scheme that is not *i*'s true preference. That is, individual *i* can manipulate at p if, given *i*'s preferences at p, *i* prefers the outcome that emerges when *i* reports a different preference to the one that is selected when *i* truthfully reports his or her actual preference at p and everyone else continues to report their preferences at p.

How do we know that the individual preference pattern of Table 7.4 will arise? Hidden information requires us to define a rule that can handle any preference input because we don't know what people's preferences might be. Note that even if we are committed to using majority rule, every specification of the rule for selecting an outcome when there is a voting cycle defines a different voting scheme. Moreover, we do *not* assume that public decisions have to be made via a process that is based on majority rule. We *do* assume that the process is fundamentally democratic, but we're not sure yet exactly what that means.

1.2 Other voting schemes

The Gibbard-Satterthwaite Theorem of Section 2 establishes that if there are three or more alternatives, then every nondictatorial voting scheme can be manipulated. (A dictatorial scheme selects an individual in advance, say

person *J*, and always selects the alternative that is at the top of *J*'s reported preference scheme.) In spite of the theorem, there is much that we can learn about preference revelation by attempting to find a scheme that is invulnerable to manipulation. First, even if invulnerability to manipulation implies that one individual has considerable power, why can't that power ever be mitigated? The next example shows what can go wrong.

Example 1.5: Sharing the power

There are two voters. Suppose that we let the mechanism select person 1's top alternative, unless it is ranked at the bottom by person 2, in which case person 2's top alternative is selected. Consider the situation of Table 7.6. The "True"

Та	bl	e	7.	.6

Ti	ue	Reported		
Person 1	Person 2	Person 1	Person 2	
X	Y	X	Y	
Y	X	Y	Ζ	
Ζ	Ζ	Ζ	X	

cell gives the respective true preferences of the individuals. Under truthful revelation, alternative *X* is top ranked by person 1 but is not bottom ranked by person 2. Thus the rule requires that *X* be selected. But if individual 2 changes her reported preference ordering to the one represented as the second column in the "Reported" cell, then *X* will not be selected because it is bottom ranked by person 2, and according to the rule, alternative *Y* will be selected. Note that person 2 does not require 1's cooperation. Person 1's reported ordering is the same in the two situations. Person 2 can do better than reporting her true preference by reporting a different ordering. When person 2 does this she precipitates the selection of an alternative that she prefers, according to her true preference scheme, to the alternative that is selected when she reports truthfully.

We run into the same difficulty if we seek a compromise.

Example 1.6: Compromise

There are two individuals and three alternatives. If the individuals have the same top-ranked alternative then that is selected. Suppose that they have different top alternatives. If each person's top is the other's second-ranked alternative, or each person's top is the other's third-ranked alternative, then select whichever of the tops precedes the other in the alphabet. Otherwise, one person's top will be

the other's second-ranked alternative and one person's top is the other's thirdranked alternative, in which case select the alternative that ranks first for one person and second for the other. Consider the situation of Table 7.6. Alternative *X* is selected under truthful revelation, but if person 2 reports the ranking in column 2 of the "Reported" cell instead then *Y* will be selected. Again, person 2 can misrepresent in a way that precipitates the selection of an alternative that she prefers, according to her true preference scheme, to the alternative that is selected when she reports truthfully.

The voting rule of Example 1.5 treats alternatives symmetrically but treats individuals asymmetrically. Example 1.6 treats individuals symmetrically but alternatives asymmetrically, as does the next example.

Example 1.7: Veto power

There are three feasible alternatives, and all voters are asked to name the one they most prefer. If at least one person names X then X is adopted; otherwise Y is selected, unless everyone nominates Z, in which case Z is adopted. In other words, any individual can veto Z by naming either X or Y, and any individual can veto both Y and Z by naming X. Suppose that there are three voters 1, 2, and 3 with the true preference orderings given in Table 7.7. If each reports

Та	bl	е	7	7

Person 1	Person 2	Person 3
Z	Y	Y
X	Z	Z
Y	X	X

truthfully then the outcome is *Y*. However, person 1 can profit from misrepresentation. If person 1 nominates *X* then *X* will be selected, and person 1 prefers that alternative to *Y*. When person 1 profits from misrepresenting his preferences the resulting outcome *X* is less desirable than *Z* in terms of *everyone's* true preferences.

What about plurality rule? With this procedure the voters simply declare their most-preferred alternative, and the one receiving the most votes is the outcome. It is well known that this rule is not guaranteed to induce truthful revelation. Before looking at an example, we specify a tie-breaking rule, because we want the decision scheme to always select one and only one alternative. Let's agree that in case of a tie the alternative coming first in the alphabet is selected from the tied alternatives receiving the most votes.

Example 1.8: Plurality rule

There are three individuals and three alternatives. Again, think of the left-hand cell of Table 7.8 as the true preferences of the respective individuals. Then the

Table 7	.8
---------	----

True			Reported		
Person 1	Person 2	Person 3	Person 1	Person 2	Person 3
X	Y	Ζ	X	Ζ	Ζ
Y	Ζ	X	Y	Y	X
Ζ	X	Y	Ζ	X	Y

three alternatives are tied, with one vote each, so X is selected. But if person 2 reports the preference scheme of the middle column of the "Reported" cell the winner will be Z with two votes, assuming that the other two report truthfully. By misrepresenting her preference and claiming that she prefers Z to the other alternatives, person 2 precipitates the selection of an outcome that she prefers according to her true preference—the middle column of the "true" cell—which ranks Z above X. This is a case of voting for your second choice instead of your first choice, because your favorite candidate doesn't have much chance of victory.

Suppose that we use plurality rule, but if there is a tie then we have a second election involving only the first-round winners. The next example shows how easily this scheme can be manipulated.

Example 1.9: Plurality rule with runoff

There are five individuals and four alternatives. The true preferences of each person are given in the top half of Table 7.9. If everyone reports their mostpreferred alternative truthfully then X will receive two votes and Y, Z, and W will receive one vote each. Then the outcome will be X, which is the alternative preferred least by person 5. If 5 deviates from the truth and claims Y as his mostpreferred alternative—see the last column in the bottom half—then there will be a tie between X and Y. Alternative Y will win the runoff, receiving votes from 3, 4, and 5, each of whom prefers Y to X. Individual 5 obviously prefers Y to X according to his true preference ordering.

Consider now conventional rank order voting—also called the *Borda rule* after the eighteenth-century French mathematician J.-C. Borda. If there are four alternatives, all individuals cast four votes for the alternative that ranks at the top of their preference ordering, three votes for the alternative that ranks second, two votes for the alternative that ranks third, and one vote for the last-placed alternative.

Table	7.9)
-------	-----	---

		True		
Person 1	Person 2	Person 3	Person 4	Person 5
X	X	Y	Ζ	W
Y	Y	Z	Y	Ζ
Ζ	Z	X	X	Y
W	W	W	W	X
Reported				
Person 1	Person 2	Person 3	Person 4	Person 5
X	X	Y	Ζ	Y
Y	Y	Ζ	Y	W
Z	Ζ	X	X	Ζ
W	W	W	W	X

DEFINITION: The Borda (or rank order) rule

If there are *m* alternatives, all individuals cast *m* votes for the alternative that ranks at the top of their preference ordering, m - 1 votes for the alternative that ranks second, and so on. Then the alternative selected is the one with the most total votes.

Example 1.10: Borda or rank order voting with four alternatives

There are three individuals and four alternatives (Table 7.10). Again, the lefthand cell displays the true preferences of each person. Then for the true preference pattern alternative X receives nine votes, Y receives eight, and Z and W get seven and six votes respectively. Therefore, X would be selected by the Borda rule

Table	7.10
-------	------

	True			Reported	
Person 1	Person 2	Person 3	Person 1	Person 2	Person 3
X	Ζ	W	X	Ζ	Y
Y	X	Y	Y	X	W
Ζ	Y	X	Ζ	Y	Z
W	W	Ζ	W	W	X

in this situation. The right-hand cell displays the reported preferences resulting from a change in person 3's reported preference ordering: This time *Y* receives nine votes, *X* and *Z* each receive eight, and *W* gets five votes. Then *Y* would be selected. Person 1's ranking is the same in the two tables, and so is 2's. If, instead of reporting truthfully (the third column of the "True" cell), person 3 reports the ranking in the third column of the "Reported" cell then the outcome will be *Y*, which individual 3 prefers to *X* according to his true preference ordering.

We could investigate many more voting rules, but our search is destined to fail. The Gibbard-Satterthwaite Theorem of Section 2 implies that any voting rule can be manipulated unless it is sensitive to the preferences of only one individual or only two of the feasible alternatives are ever allowed to be selected. Suppose, however, that one of the three orderings of Table 7.4 could not possibly represent an individual's preferences for the chosen application. Perhaps the remaining preference orderings could never give rise to a voting cycle. The next section considers the possibility that majority rule is invulnerable to manipulation when some restriction is placed on individual preference.

1.3 Value-restricted preferences

Suppose that the feasible alternatives can be arranged in some intrinsic order that has significance for individual preferences. For instance, the three alternatives are L (left), M (middle), and R (right), as illustrated in Figure 7.1. They



may represent three different levels of expenditure on a public fireworks display, with *L* requiring the smallest amount of money and *R* the largest. Similarly, the alternatives could be

three different regulatory policies for a particular industry, with L requiring the greatest amount of government regulation and R the least. In some contexts we know in advance that anyone who prefers L to the other two alternatives would never prefer R to M. Similarly, anyone who prefers R to the other two could not conceivably have a preference for L over M. That means that we can rule out any preference scheme that has M as the least-preferred alternative. (If neither L or R is the most-preferred alternative then M must be preferred to the other two.) If we restrict the true *and* the reported preferences to the set of all logically possible orderings except the ones that have M ranked last then we can be sure of two things:

- There will be an alternative that would not be defeated by any other alternative under majority rule.
- If a majority winner were selected as the outcome then no individual can manipulate.

When the alternatives can be naturally ordered on a line, and an alternative is never ranked below one to its right *and* one to its left we say that individual preferences are *single peaked*. However, we can prove statements 1 and 2 under a much weaker restriction on individual preference. A set of preference orderings

is *value restricted* if we *cannot* find three preference orderings in that set and three alternatives that generate the preference pattern of Table 7.4.

DEFINITION: Value-restricted preferences

Given three alternatives *x*, *y*, and *z*, a set of preference orderings is *x* restricted if alternative *x* does not rank above *both y* and *z* for any of the orderings in the set, *or x* does not rank below both *y* and *z*, for any of the orderings in the set, *or x* does not rank in between *y* and *z* any of the orderings in the set. We say that the set of preference orderings is *value restricted* if for any three alternatives, *x*, *y*, and *z*, the set is either *x* restricted, *y* restricted, or *z* restricted.

The three preferences of Table 7.4 are not value restricted because each of the three alternatives is first for one individual preference, last for one individual, and ranked in between the other two for the remaining individual.

For the rest of this section we assume an odd number of voters. The results go through with an even number of voters and some rule to select a winner in case of a tie, but by restricting attention to an odd number of individuals we are able to greatly simplify the argument. Up to this point we have implicitly assumed away ties at an individual level. That is, an individual is never indifferent between distinct alternatives. We maintain this assumption—for convenience. With an odd number of individuals and the absence of individual indifference, then for any two alternatives x and y, either a majority of individuals prefers x to y or a majority prefers y to x.

Sen's majority rule theorem

If the number of individuals is odd, the number of alternatives is finite, individual indifference is ruled out, and individual preferences are value restricted, then there is one and only one alternative that defeats every other alternative by a clear majority. We call this alternative the *unique majority winner*.

Proof

We begin by showing that there cannot be a majority cycle. Suppose to the contrary that x defeats y by a majority, y defeats z by a majority, and z defeats x by a majority. Let J be the set of individuals who prefer x to y, and let K be the set of individuals who prefer y to z. Then J contains more than 50% of the individuals, and so does K. Therefore, J and K must have at least one individual in common. (If J is a majority, then the individuals *not* in J constitute a minority.) Therefore, there is at least one individual who prefer x to y and y to z. Now, let H be the set of individuals who prefer z to x. We know that H is a majority. The

individuals in *K* prefer *y* to *z*, and the individuals in *H* prefer *z* to *x*. Because *K* and *H* each contain a majority of individuals the two sets must have at least one individual in common. Therefore, there is someone who prefers *y* to *z* and *z* to *x*. Similarly, because *H* and *J* are each majorities there is at least one individual who prefers *z* to *x* and *x* to *y*. But now we have the three preference orderings of Table 7.4, contradicting value restriction.

The previous paragraph rules out majority cycles. Now we show that this implies that there is a unique majority winner. That is certainly the case if there are only two feasible alternatives. Suppose that there is a unique majority winner whenever there are m or fewer feasible alternatives, but we are investigating a set A containing exactly m + 1 feasible alternatives. Choose any member x of A. Let B denote the set of all alternatives in A apart from x. Then B has m members, so it has a unique majority winner y. Either y defeats x by a majority or x defeats y by a majority. If y defeats x by a majority, but x is not the unique majority winner in A. If x defeats y by a majority, but x is not the unique majority winner in A then there is an alternative z in B that defeats x by a majority. But y defeats z by a majority because y is the unique majority winner from B. Now we have a cycle: x defeats y, y defeats z, and z defeats x. But there are no cycles with value-restricted preferences, so we are forced to drop the assumption that x is not the unique majority winner in A when x defeats y by a majority. We are finished our proof of Sen's theorem.

Now that we know that there is always a unique majority winner whenever the individual preferences belong to the value-restricted family, we are ready to show that no individual or group can profit by deviating from truthful revelation, provided that the reported preferences must also belong to this family. First we extend the definition of manipulability to groups of individuals. They can manipulate if they can jointly deviate from truthful revelation in a way that leaves all members of the group better off according to their true preferences.

DEFINITION: Manipulation by a group

We say that the individuals in group C can *manipulate* the voting rule if there are two different assignments p and q of individual preferences such that, for any individual i not belonging to C, p and q assign the same preference ordering to i and, according to the individual preferences at p, everyone in C prefers the outcome that emerges when they report the preference assigned to them at q to the alternative that is selected when all members of C truthfully report their actual preferences at p.

Group *C* could consist of a single individual, so the definition covers manipulation by an individual. We think of the preferences assigned by *p* as the true preferences of the respective individuals. The members of *C* can manipulate if they can coordinate their strategies by reporting the preferences assigned to them at some *q*, *and* each member of *C* prefers the resulting outcome to the one

that emerges when each reports truthfully, *and* each individual not in *C* has the same preference at *q* that the individual has at *p*. The individuals not in coalition *C* cannot be expected to assist the coalition, hence the assumption that no one outside of *C* reports a different preference at *q* than the one reported at *p*.

Suppose that some family F of value-restricted preferences is known to contain the true preference ordering of each individual in the community, although the actual preference of an individual is still hidden information. If the reported preference of each individual must belong to F then majority rule cannot be manipulated by any individual or group, as we now prove.

Nonmanipulability of majority rule for value-restricted preferences

Suppose that the number of individuals is odd, the number of alternatives is finite, individual indifference is ruled out, and both the true and reported individual preferences belong to some set F that is value restricted. Then majority rule cannot be manipulated by any individual or coalition.

In other words, truthful revelation is a dominant strategy for groups as well as individuals.

Proof

Suppose that under truthful revelation alternative x is the unique majority winner but everyone in group C prefers y to x. If no one outside of C changes his or her reported preference ordering then there is nothing the members of C can do to precipitate the selection of y, because x defeats y by a clear majority even when everyone in C truthfully declares a preference of y over x. Whatever preferences are reported by the members of C, they must belong to F, so there will be a unique majority winner by Sen's theorem. But that winner cannot be y as we have just demonstrated. We have finished the proof.

The assumption that the reported preferences must also belong to the family *F* is essential. Otherwise a voting cycle could be precipitated, in which case the definition of majority rule would have to be augmented to specify the alternative selected when there is a cycle, and that opens the door to manipulation by an individual. That fact is a consequence of the striking theorem introduced and proved in the next section.

Source

The possibility of a majority voting cycle appears to have been noticed first in Condorcet (1785). The rank order method of voting was advocated in Borda (1781). Sen's theorem is from Sen (1966).

Links

Neufeld, Hausman, and Rapoport (1994) present an example of a voting cycle from the U.S. Senate. Brams (1976, pp. 43–7) discusses the strategic dimensions

of the U.S. House voting cycle presented in this section. Campbell and Kelly (2000, 2003) explore the normative and strategic properties of majority rule in some depth. Arrow (1951a, 1963a) began the formal study of the criteria for determining how an alternative should be selected as a function of individual preferences, and his impossibility theorem is one of the most significant discoveries in the history of ideas. Campbell and Kelly (2002) prove that theorem and its close relatives and review the literature since Arrow.

Problem set

- 1. Prove that in the case of two alternatives, if majority rule is used to select one of the alternatives, supplemented by some tie-breaking rule, when necessary, then whatever tie-breaking rule is employed, no individual (or coalition) can every profit by misrepresenting preference.
- **2.** Example 1.3 presented a majority rule cycle with three alternatives and three individuals. Explain how one can construct a cycle with three alternatives and any odd number of individuals (except one).
- **3.** List all the logically possible linear orders of the three alternatives *x*, *y*, *z*. (A linear order is one that can be represented as a column, with higher alternatives being preferred to lower ones; a linear order rules out the possibility that two distinct alternatives are indifferent to each other.) Now, identify all the families of single-peaked subsets of preferences.
- **4.** Does the following family of preferences over the three alternatives *x*, *y*, *z* have the single-peak property? Explain.

x	\mathcal{Z}	Z	у
Z	x	у	Z
У	у	x	x

(There are four preference orderings, each represented by a column, and a higher alternative is preferred to a lower one.)

- **5.** Suppose that an odd number of individuals have single-peaked preferences. Explain why the most-preferred alternative of the median voter—the one whose peak lies between the peaks of the other two voters—will defeat every other proposal by a majority of votes.
- **6.** Show that there are many families of preferences that are value restricted. In particular, for the case of three feasible alternatives, identify all the families of value-restricted preferences that are maximal in the sense that one can't introduce another preference ordering without precipitating a violation of value restriction.
- **7.** Consider the following voting scheme for selecting an outcome from a set of three available alternatives, *x*, *y*, and *z*: All individuals report their preference ordering to a referee. If according to the reported preferences one alternative

defeats each of the two others by a majority then that alternative is the chosen outcome. If there is no such majority winner then the outcome is either *x* or *y*, whichever of those two alternatives defeats the other by a majority. Is reporting one's true preference ordering always a dominant strategy in the case of this particular mechanism? If it is, explain briefly why it is. If it is not, prove that truthful revelation is not a dominant strategy by means of an example.

8. Consider majority rule with six people. Because there is an even number of individuals there is a possibility of ties. We will say that α is a unique majority winner if there is no feasible alternative that defeats α by a majority, and for every other feasible alternative β there is at least one other feasible alternative that defeats β by some majority. Use Tables 7.11 and 7.12 to show that the rule that selects the unique majority winner can be manipulated by a single individual.

Table 7.11	
------------	--

Person 1	Person 2	Person 3	Person 4	Person 5	Person 6
x	x	w	у	Z	z
у	w	У	Z	x	w
Z	У	x	w	w	У
w	Z	Z	x	У	x

Table 7.12

Person 1	Person 2	Person 3	Person 4	Person 5	Person 6
x	x	у	у	Z	z
у	w	Z	Z	x	w
Z	У	w	w	w	У
w	Z	x	x	У	x

- **9.** For the voting rule of Example 1.7, for each configuration p of individual preference orderings, determine the outcome that emerges at a Nash equilibrium when each individual reports a preference ordering that is a best response—in terms of the individual's preference ordering at p—to the preferences reported by the others.
- **10.** Assume that the number of individuals is odd, that there are exactly three alternatives, and that individuals are never indifferent between distinct alternatives. Prove that if the subset of preference orderings *F* has the property that for every assignment of preferences to individuals there is a unique majority winner then *F* is value restricted.

2 PREFERENCE REVELATION IN GENERAL

The previous section featured a number of voting rules. With unrestricted individual preferences, almost every rule gave individuals an incentive to profit from misrepresenting their preferences in some situations. Only two things worked: restricting the number of alternatives on the ballot and restricting the set of admissible individual preferences. If there are only two alternatives on the ballot then conventional majority rule induces truthful revelation of individual preference. But that approach will—except by accident—run afoul of efficiency when there are actually three or more feasible alternatives.

What about arbitrarily restricting individual preferences? If we have a priori information establishing that each person's preference ranking belongs to a particular family of single-peaked or value-restricted preferences, then we can apply majority rule and know that there will be an alternative that will survive an election when paired with any other alternative and that truthful revelation is a dominant strategy for each person. Specifically, truthful revelation is a dominant strategy if no one is allowed to report a preference scheme that is outside of the restricted family.

This section investigates the possibilities for truthful revelation when there are three or more alternatives and we are unable to rule out any ranking of the alternatives as a possible preference scheme for any individual. We prove that if each alternative is selected in at least one situation—perhaps when everyone ranks that alternative at the top—then the only rules that induce truthful revelation are dictatorial. Before stating this formally, we specify our social choice framework.

At the very least, democracy means that decisions that have widespread social consequences should be the result of a process in which individual assessments of the alternatives have a bearing on the outcome-in a positive way. The society that we wish to study can be anything from a small committee to an entire nation. Once we have identified the society, the next step is to identify the policy options at issue. These alternatives have a different character in different applications. If a college class is selecting a president then the alternatives are the names of the eligible candidates. If we are interested in a nation selecting a health care policy, then the options are the various proposals under consideration. In most situations, one of the options will be "no policy"-that is, the status quo. We can even model the market system in this abstract way, with each alternative being a particular configuration of production and consumption activities. The framework will be flexible enough to embrace a panel of physicians deciding which patient is to receive a kidney transplant when the next organ becomes available, or NASA determining which experiments will be performed by a space probe of Saturn, or the board of directors of a modern corporation deliberating on a proposed merger. The analysis pertains to any group decision in which the individuals care about the alternatives but where there is rarely unanimous agreement about what should be done.

The group's decision should be a function of the individual preferences of the group members. Therefore, we need to model the individual assessments

2. Preference Revelation in General

of the policy options, and we do this by representing the *order* in which each individual ranks the alternatives, as we did in Section 1. In this chapter, the social choice process does not use any information about the *intensity* of an individual's preference for one alternative over another. Consequently we only require information about the order of individual preference—how one alternative ranks relative to another. Does *x* rank above or below *y* in individual *i*'s preference scheme? In the next chapter we see that allowing additional preference input—specifically, information about intensity of preference—opens the door to a social choice procedure that induces truthful revelation even though it is sensitive to everyone's preferences. This procedure has significant drawbacks, however.

We say nothing about *why* the individuals have the preferences that they have. One person will be very selfish and rank the policy options according to their impact on the basket of goods and services that person will consume. Another may be altruistic, taking into account the impact on those less fortunate. Some individuals will not rank alternatives according to what they want but according to what they think society should have; in some people this normative judgment is selfishly based, and in others it is not. Whatever the basis of the individual's preference ordering of the alternatives, this is what motivates the individual's participation in the decision process, and these orderings are the basic input in our model. The output is the decision rendered; a selection of one of the options as the decision taken by society. In other words, the *social choice rule* selects a feasible alternative as a function of individual preferences.

DEFINITION: The social choice framework

There are *n* individuals and a given set *X* of feasible alternatives. A *pro-file p* assigns an ordering p(i) of the members of *X* to each individual *i*. *A social choice rule g* selects a member g(p) of *X* for each profile *p*. We require g(p) to be defined for every logically possible profile *p* for which no individual is ever indifferent between two distinct alternatives and for which every individual's preference is transitive. This is the *unrestricted preferences* assumption. (Transitivity means that if *x* is preferred to *y* and *y* is preferred to *z* then it must be the case that *x* is preferred to *z*.) The rule *g* is *nonimposed* if for each member *x* of *X* there is some profile *p* such that g(p) = x.

A profile is interpreted as a record of the individual preferences in a particular situation. Sometimes the profile will represent the true preferences of the respective individuals, and sometimes it will represent the preferences *reported* by the individuals. No restriction is placed on the either the true or the reported preferences. In other words, the social choice rule must render a decision for each logically possible assignment of preference orderings to individuals. We would expect a social choice rule to select *x* at any profile for which everyone

ranked *x* at the top. Nonimposition merely requires that there be *some* profile at which *x* is selected.

The social choice rule g is not necessarily a voting system—and we are certainly not restricting attention to rules that have actually been tried. In principle g could represent the market system, with g(p) being the market equilibrium for the configuration of preferences represented by p. However, the theorem that we prove in this section applies only when the unrestricted preferences assumption is appropriate-that is, when every logically possible ordering of the feasible set is a conceivable individual ordering. This is certainly not the case for the allocation of private goods. Let x denote the feasible allocation that delivers every unit of every good to individual 1. Then the social choice rule does not have to be defined for a profile that has x on the top of person 2's preference ordering. However, our proof in this chapter makes use of this ordering. Therefore, the theorem is relevant only for situations in which the unrestricted preferences assumption is appropriate. For instance, if the members of X represent different mixes of public projects that are financed with a given amount of tax revenue, then we cannot a priori rule out any ordering of X as a possible preference that an individual might have.

2.1 The Gibbard-Satterthwaite Theorem

We seek a nonimposed social choice rule that induces individuals to report their preferences truthfully. In other words, no individual can manipulate. A social choice rule for which truthful preference revelation is a dominant strategy is said to be *strategy proof*.

DEFINITION: Strategy proofness

Individual *i* can *manipulate* the social choice rule *g* if there are two profiles p and q such that q(j) = p(j) for all $j \neq i$ and g(q) ranks higher in p(i) than g(p) does. The social choice rule *g* is *strategy proof* if no individual can manipulate.

In interpreting the definition of manipulation we think of p(i) as the true preference of individual *i*. By reporting q(i) instead of p(i) individual *i* precipitates the selection of an alternative g(q) that *i* prefers to the alternative g(p) that is selected when *i* reports truthfully. We require q(j) = p(j) for all $j \neq i$ because we cannot assume that others will change their reported preferences to suit individual *i*.

Here are two strategy-proof social choice rules: (i) Select two specific alternatives and label them x^* and y^* . The outcome is x^* unless the reported individual preference orderings have y^* ranking above x^* in a majority of cases, in which case y^* is the outcome. There may be many available alternatives other than these two, but only x^* and y^* can ever be selected. Why is this scheme strategy proof? Suppose that x^* wins under truthful revelation, but individual j prefers

2. Preference Revelation in General

 y^* to x^* . Then *j* has voted for y^* but x^* won nonetheless. Misrepresentation by *j* would result in another vote being cast for x^* , in which case x^* will still win. Similarly, if y^* wins under truthful revelation, but *j* prefers x^* there is nothing that *j* can do to prevent the election of y^* . In most situations there are more than two alternatives to consider but with this voting scheme the outcome must be either x^* or y^* in every case. Suppose there is a third alternative *z*, and everyone ranks *z* as most preferred. The outcome will not be efficient; everyone prefers *z* to the winner, which must be either x^* or y^* . This procedure is strategy proof but very unsatisfactory on efficiency grounds.

Our second procedure *is* efficient: (ii) There are any number of alternatives and the outcome is always the most-preferred alternative according to the reported preference scheme of person 1. No one other than person 1 can affect the outcome so no one other than person 1 has an incentive to manipulate. And it is in person 1's self-interest to ensure the selection of 1's most-preferred alternative so person 1 will always report truthfully. The procedure is strategy proof. It is obviously dictatorial and therefore completely unsatisfactory.

DEFINITION: Dictatorship

The social choice rule g is *dictatorial* if there is some individual i such that, for each profile p, g(p) is the top-ranked member of p(i). We say that individual i is a *dictator* in that case.

In plain words, an individual is a dictator if the rule *always* selects the alternative that the individual most prefers. If there are at least three feasible alternatives, a nonimposed social choice rule cannot be strategy proof if it is not dictatorial.

The Gibbard-Satterthwaite Theorem

Suppose that the feasible set has three or more members, preferences are unrestricted, and the social choice rule *g* is nonimposed. Then *g* is dictatorial if it is strategy proof.

Proof of this theorem for the case of two individuals and three alternatives is quite easy, and we take care of that in the next section. You may wish to stop there, in which case you will have a good grasp of the import of the theorem and a good idea of how the general proof works.

2.2 Proof for two individuals and three alternatives

We assume in this section that the feasible alternatives are *x*, *y*, and *z*, and no others, and that persons 1 and 2 are the only individuals whose preferences are to be considered. We show that either person 1 or person 2 must be a dictator.

That means that a strategy-proof rule must select an individual in advance, say individual 1, and always select 1's top-ranked alternative at every profile. The crux of the proof consists in presenting a profile at which either person 1 or person 2's top alternative must be selected—because the third alternative is bottom ranked by both. (It is easy to show that an alternative can't be selected by a strategy proof rule if it is bottom ranked by everyone.) Then we show that the individual whose top-ranked alternative is selected must inevitably have absolute power. But even if strategy proofness implies that one individual has considerable power, why can that power ever be mitigated? Example 1.5 in the previous section shows what can go wrong.

The proof begins with the unanimity lemma: If both persons 1 and 2 have a common top-ranked alternative then that alternative must be selected if *g* is strategy proof. Then we use that lemma to show that strategy proofness implies that *z* will not be selected at a profile at which both persons rank *z* at the bottom. We use the symbol \Box at the end of an argument to show that the proof is complete.

Step 1: The unanimity lemma

If both 1 and 2 have *x* ranked at the top at *p* then g(p) = x if *g* is strategy proof and *x* is selected at *some* profile.

Proof

Let *q* be a profile at which *x* is selected, and let *p* be a profile at which both persons 1 and 2 have *x* ranked at the top. Strategy proofness implies that *g* must select *x* when person 1's true preference ordering is *p*(1) and person 2 reports the ordering *q*(2). That is, g(p(1), q(2)) = x. If we did not have g(p(1), q(2)) = x then person 1 would have an opportunity to manipulate by reporting *q*(1) because we know that g(q(1), q(2)) = x and *x* is ranked at the top of *p*(1). Therefore, strategy proofness implies g(p(1), q(2)) = x. That fact itself implies that strategy proofness requires that *g* select *x* when person 2's true preference ordering is *p*(2) and person 1 reports *p*(1). That is, g(p(1), p(2)) = x. If we did not have g(p(1), p(2)) = x then person 2 would have an opportunity to manipulate by reporting *q*(2) because we know that g(p(1), q(2)) = x. Therefore, strategy proofness implies g(p(1), q(2)) = x. Therefore, strategy proofness g(p) = g(p(1), p(2)) = x.

Now we show that if both persons 1 and 2 have a common *bottom*-ranked alternative then that alternative must not be selected if *g* is strategy proof.

Step 2

If both persons 1 and 2 have *z* ranked at the bottom then $g(p) \neq z$ if *g* is strategy proof and nonimposed.

Proof

If the two individuals have the same top-ranked alternative, then that alternative will be selected (by the unanimity lemma), and hence the common bottom-ranked alternative will not be selected.

2. Preference Revelation in General

 p(1)
 p(2)

 x
 y

 y
 x

 z
 z

Table 7.13: Ifg(p) = z then person 2 would have an opportunity to manipulate by reporting an ordering with x at the top: The unanimity lemma (step 1) implies that x would then be selected, and person 2 prefers x to z according to 2's true preference ordering p(2). Therefore, strategy proofness implies that $g(p) \neq z$.

What if the top-ranked alternatives are different? Consider profile p of

We have just discovered that g(p) = x or y at the profile p of Table 7.13. We show that if g(p) = x then person 1 is a dictator for g, and if g(p) = y then person 2 is a dictator for g. We suppose until further notice that g(p) = x.

Step 3

If g(p) = x for the profile p of Table 7.13 then g(q) = x for the profile q of Table 7.14.

Proof

If g(q) = z then person 1 would have an opportunity to manipulate when the profile of true preferences is q, because, by the unanimity lemma, person 1

Table 7.14. Profile q

q (1)	q (2)
x	у
y z	$z \\ x$

could then precipitate the selection of *y* by reporting an ordering with *y* at the top. Therefore, strategy proofness implies that $g(q) \neq z$.

If g(q) = y then person 2 would have an opportunity to manipulate when the profile of true preferences is *p* of Table 7.13 because g(p) = x by assumption, at *q* person 1's ordering is the same as it is at *p*, and person 2 prefers *y* to *x* at *p* according to 2's true preference p(2). Therefore, strategy proofness implies that $g(q) \neq y$.

We have ruled out g(q) = z and also g(q) = y. Therefore, we must have g(q) = x.

Step 4

If g(p) = x for the profile *p* of Table 7.13, then g(r) = x for *any* profile *r* that has *x* ranked at the top for person 1.

Proof

Note that *x* is ranked at the bottom by person 2 at the profile *q* of Table 7.14. Therefore, strategy proofness implies that *x* will be selected at any profile for which 1's reported ordering is q(1)—that is, whenever person 1 has *x* at the top and *z* at the bottom. If in that situation there is an ordering that person 2 could report to secure the selection of an alternative other than *x* then 2 could manipulate at *q*, where person 2 prefers anything to *x*. Therefore, g(s) = x for any profile *s* that has s(1) = q(1).

Now, if $g(r) \neq x$ and x is ranked at the top for person 1 then person 1 can manipulate at r by the argument of the previous paragraph, because person 1

```
Table 7.15. Profile t
```

<i>t</i> (1)	<i>t</i> (2)			
у	z			
x	у			
z	x			

can precipitate the selection of 1's most-preferred alternative by reporting the ordering q(1). (To make this claim we suppose that r(1) is the true preference of person 1.) Therefore, we must have g(r) = x, at any profile r for which x is top-ranked by person 1.

We haven't established that person 1 is the dictator. So far, we only know that alternative x will be selected at any profile at which x is top ranked by person 1. Might we have x selected in some situation in which it is bottom ranked for person 1?

That would disqualify person 1 as a dictator. We now proceed to show that this can't happen.

Step 5

If g(p) = x for the profile *p* of Table 7.13, then g(t) = y for profile *t* of Table 7.15.

Proof

If g(t) = x then person 2 would have an opportunity to manipulate when the profile of true preferences is *t*, because, by the unanimity lemma (step 1), person 2 can precipitate the selection of *y* by reporting an ordering with *y* at the top. Therefore, strategy proofness implies that $g(t) \neq x$.

If g(t) = z then person 1 would have an opportunity to manipulate when the profile of true preferences is *t* by reporting an ordering with *x* on top. That would precipitate the selection of *x* (by step 4) and at *t* person 1 prefers *x* to *z*. Therefore, strategy proofness implies that $g(t) \neq z$.

We have ruled out g(t) = x and also g(t) = z. Therefore, we must have g(t) = y.

Step 6

If g(p) = x for the profile *p* of Table 7.13, then g(u) = y for profile *u* of Table 7.16.

Proof

If $g(u) \neq y$ then person 1 would manipulate when the profile of true preferences is *u* by reporting *t*(1) of step 5. (Note that *u*(2) and *t*(2) are identical.) Therefore, strategy proofness implies that g(u) = y.

Now we complete the proof of the Gibbard-Satterthwaite Theorem by applying steps 2, 3, and 4 to the profile *u* (of step 6) to show that *y* is selected at *any*

Table	7.16.	Profile	и
-------	-------	---------	---

u (1)	u (2)				
y	z				
z	y				
x	x				

profile for which person 1 has *y* ranked at the top. (This time, *y* plays the role of *x*, and *x* plays the role of *z*. Note that *x* is bottom ranked for both persons at *u*.) Having established that, we use the arguments of steps 5 and 6 to show that *z* is selected at the profile with *y* at the bottom for both persons, *z* at the top for person 1, and *x* at the top for person 2. Having established that, we can apply steps 2–4 again to show that *z* is selected at any profile at which person 1 has *z* ranked at the top. Because

2. Preference Revelation in General

x, *y* , and *z* are the only three alternatives, we then have proved that individual 1 is a dictator.

All of this has been predicated on the assumption that *x* is selected at profile *p* of Table 7.13. We know that it has to be either *x* or *y*. What if it's *y*? Then we apply the same arguments to person 2, to show first that *y* will be selected at any profile for which it is top ranked by *person 2*, then we go on to show that *x* is selected at any profile for which it is top ranked by person 2, and then finally we show that *z* is selected at any profile for which it is top ranked by person 1 is a dictator or person 2 is a dictator. That is, every strategy-proof and nonimposed rule is dictatorial if there are three alternatives and two individuals.

Remark 1. Nonimposition played a role in the proof right at the beginning when we assumed that *x* is selected at some profile. Similarly, the proof depends on the assumption that *y* is selected at some profile, and so is *z*. (The rule that selects *x* at *every* profile is obviously strategy proof, and so is the rule that *never* selects *z* but selects *x* when both individuals prefer *x* to *y* and otherwise selects *y*.)

Remark 2. With only two individuals, there will be profiles at which x is first and y is second for one person and the other has y first and x second, as in the case of profile p of Table 7.13. One of the two individuals must be favored in that instance, and once that happens strategy proofness forces the rule to give that individual absolute power. However, the Gibbard-Satterthwaite Theorem is not special to the case n = 2. If n > 2 there will be favoritism of some individual at some profile, but it will be more subtle. Nevertheless, strategy proofness will still force the rule to give the favored individual absolute power. For instance, begin with profile α that has x at the top for everyone and y second. Then x will be selected. Now, move y to the top and x down to second place for one individual at a time. When we are finished we will be at profile β that has y first and x second for everyone, and hence y will be selected at β . We started with x being selected and finished with y. At some intermediate stage we must have switched the ordering of x and y for one individual and caused the selection to change from x to y. That individual is given special treatment at that point, and strategy proofness will imply that that person is a dictator.

Sources

This remarkable theorem was discovered independently in the 1970s by the philosopher Alan Gibbard and the economist Mark Satterthwaite (Gibbard, 1973; Satterthwaite, 1975). This section presented a different proof from either of theirs. (Quite a number of different proofs of the Gibbard-Satterthwaite Theorem are now available in the literature.)

Links

There is a vast literature on the preference revelation problem. Significant surveys include Chapter 10 of Kelly (1988), Chapter 1 of Saari (1994), Barberà (2001, 2004), Jackson (2001), and Maskin and Sjöström (2002).

Problem set

- 1. Consider the following social choice rule defined for two individuals, 1 and 2, and three alternatives *x*, *y*, *z*: The individuals report a preference ordering of the three alternatives, and if one alternative is ranked at the top of both reported preference orderings then that alternative is the selected outcome; otherwise person 1's top-ranked alternative is the outcome. (Assume that two distinct alternatives are never indifferent in anyone's preference ordering.) Prove that truthful revelation is a dominant strategy for this social choice rule. Does this contradict the Gibbard-Satterthwaite theorem? Explain.
- 2. For the purposes of this question we expand the set of possible individual preferences to include cases where an individual is indifferent between two or more alternatives. Consider the following social choice rule defined for two individuals, 1 and 2, and three alternatives x, y, z: The individuals report a preference ordering of the three alternatives, and if there is a unique alternative that is ranked at the top of both reported preference orderings then that alternative is selected, and otherwise person 1's top-ranked alternative is the outcome if there is a single alternative that is top ranked by person 1. If there are two or more alternatives that are tied for top rank in person 1's reported preference scheme then the outcome is person 2's most preferred alternative (according to 2's reported preference scheme) in the set of topranked alternatives in the preference ordering reported by person 1. If there are two or more alternatives that are top ranked in person 1's reported preference scheme and within that set there is not a unique outcome that is highest ranked in person 2's reported scheme then the outcome is x. Is truthful revelation a dominant strategy for this social choice rule? Is person 1 a dictator in this case? Explain.
- **3.** There are three individuals and three alternatives, *x*, *y*, and *z*. Social choice rule g_1 selects the majority winner unless there is a voting cycle, in which case the top-ranked alternative of person 1 is selected. Social choice rule g_2 selects the majority winner unless there is a voting cycle, in which case *x* is selected if a majority prefers *x* to *y* and *y* is selected if a majority prefers *y* to *x*. Social choice rule g_3 selects the majority winner unless there is a voting cycle, in which case *x* is social choice rule g_3 selects the majority winner unless there is a voting cycle, in which case *x* is selected. For each of these three social choice rules, work out the Nash equilibrium that results when there is a cycle and each person reports a preference ordering strategically. At equilibrium it will be the case that, *given the preferences reported by others*, no individual is able to modify his or her reported preference scheme in a way that elicits an outcome that person likes better according to his or her true preference scheme.
- **4.** There are three feasible alternatives and three individuals. The rule is based on the total number of votes received by each alternative when the Borda rule (Example 1.10) is used to rank the alternatives. The alternative with the largest total score is selected. If the three alternatives have the same score then *x* is selected. If there are exactly two alternatives with the highest

3. General Proof of the Gibbard-Satterthwaite Theorem

score, then whichever of those two alternatives comes before the other in the alphabet is selected. Show that this rule can be manipulated.

5. There are three feasible alternatives, *x*, *y*, and *z*, and three individuals. The rule is based on the total number of votes received by each alternative when the Borda rule (Example 1.10) is used to rank the alternatives. The alternative with the largest total score is selected. If there are two or more alternatives with the highest score then whichever of those alternatives ranks highest in the reported preference ordering of person 1 is selected. Show that this rule can be manipulated.

Each of the next four questions pertains to a different social choice rule *g* with unknown properties *except* that truthful revelation is a dominant strategy. You can't assume that the rule is dictatorial, or is majority rule, or any other rule. You only know that the rule can't be manipulated and whatever additional properties I give you in stating the question. The profiles in question are displayed in Table 7.17.

Table 7.17

P (1)	P (2)	Q (1)	Q (2)	R (1)	R (2)	S (1)	S (2)	<i>T</i> (1)	<i>T</i> (2)
x	y r	x	у 7	y r	X 7	x	x 7	y r	z
y z	x Z	y z	$\frac{z}{x}$	x Z	y y	y z	y y	x Z	y x

- **6.** Alternative *x* is selected at profile *P*, and whenever there is a common top alternative that common top is selected. Determine the alternative that is selected at profile *Q* and explain why it must be selected.
- **7.** Alternative *x* is selected at profile *R*. Determine the alternative that is selected at profile *S* and explain why it must be selected.
- **8.** Alternative *x* is selected at any profile at which person 1 has *x* at the top, and whenever there is a common top alternative that common top is selected. Determine the alternative that is selected at profile *T* and explain why it must be selected.
- **9.** Alternative *y* is selected at profile *R*. Determine the alternative that is selected at profile *T* and explain why it must be selected.

3 GENERAL PROOF OF THE GIBBARD-SATTERTHWAITE THEOREM

The most accessible evidence for the fact that people *do* attempt to manipulate is the observation that in a three-candidate election we often choose not to vote for our most-preferred candidate when that candidate has little chance of being elected. In that situation we often vote for our second choice, in an attempt to prevent the candidate who is our last choice from being elected. But what's

wrong with *that*? It probably makes the election more responsive to individual preferences. OK. Then let's take that kind of behavior as given and build it into the definition of the rule g. In other words, let g(p) be the outcome as a result of an equilibrium configuration of strategies by the n individuals. The Gibbard-Satterthwaite Theorem tells us that even *that* rule is manipulable.

Section 2 proved the Gibbard-Satterthwaite Theorem for the case of two individuals and three alternatives. In this section we complete the proof by using that fact to show that the result holds for any number of individuals and any number of alternatives greater than two. First we hold the number of individuals at two and allow the number of alternatives to increase.

3.1 Proof for two individuals and three or more alternatives

First, note that the proof of the unanimity lemma in the previous section is valid for any number of alternatives. Now, assume that there are more than three feasible alternatives (but only two individuals).

Step 1

Choose any three of the alternatives, and let *Y* denote the set containing just those three. Consider the family of profiles obtained by fixing the individual ordering of the remaining alternatives with the members of *Y* ranked above them all. (It doesn't matter how we order the remaining alternatives, as long as we stick to it and as long as they are all below *Y*.) As we change the individual orderings of *Y*, with the others ranking below in a fixed way, we trace out the set of all profiles on *Y*. In that way we can generate a social choice rule g_Y based on the given strategy-proof rule *g*: At any profile π on *Y* we define g_Y by having g_Y select the alternative that would be selected by *g* at the profile where the individual orderings of *Y* are as specified by π and the other alternatives are ordered below the members of *Y* in the fixed, predetermined way.

It is obvious that g_Y is strategy proof because g is. We can use the unanimity lemma to prove that g_Y always selects a member of Y. Suppose that it did not, and it selected some alternative b not belonging to Y. Then person 1 could manipulate g at the profile in question by reporting an ordering with the same top alternative as person 2. That alternative would be selected by g by the unanimity lemma, and it would be preferred by person 1 to b because b ranks below every member of Y(in the orderings of both persons 1 and 2). Therefore, strategy proofness implies that g_Y always selects a member of Y. And every member of Y is selected by g_Y at *some* profile because g will always select an alternative that is top ranked by both individuals (by the unanimity lemma). Therefore, g_Y is a dictatorial rule, by the previous section.

Let i(Y) be the name of the dictator for g_Y . That is, set i(Y) = 1 if person 1 is the dictator for g_Y and set i(Y) = 2 if person 2 is the dictator for g_Y .

Step 2

If *x* belongs to *Y*, then at any profile for which individual i(Y) has *x* on top, the original rule *g* will select *x* at that profile.

3. General Proof of the Gibbard-Satterthwaite Theorem

Proof

We know that g_Y will select x if individual i(Y) has x above the other two members of Y and the other person has x ranking below the other two members of Y. Therefore, g itself will select x if i(Y) has x ranking above the other two members of Y and the other person has x ranking below the other two members of Y, and the remaining alternatives are ordered in the agreed-upon way below Y. Let pdenote such a profile. Now, create a new profile from p simply by moving x down to the bottom of the ordering of individual j, the one who is *not* the dictator for g_Y . Call this new profile q.

Let *Y* consist of the three alternatives *x*, *y*, and *z*. Note that *q* will have either *y* or *z* at the top of *j*'s ordering because, before we moved *x* to the bottom, we had each member of *Y* ranking above every other alternative. It follows that g(q) must be a member of *Y*. Suppose to the contrary that g(q) = b and *b* does not belong to *Y*. Then individual i(Y) prefers both *y* and *z* to *b*. By the unanimity lemma, individual i(Y) can precipitate the selection of *y* or *z*, whichever is at the top of *j*'s ordering, simply by reporting an ordering with the same top alternative as *j*. This contradicts strategy proofness, so g(q) must belong to *Y*. But individual *j* prefers both *y* and *z* to *x* at *p*. Therefore, if g(q) is not equal to *x* then individual *j* will have an opportunity to manipulate at *p* because q(i(Y)) = p(i(Y)). It follows that g(q) = x.

Now, let *r* be *any* profile at which individual i(Y)'s ordering is the same as it is at *q* (and *p*). If g(r) is not equal to *x* then individual *j* will prefer g(r) to g(q) = x at profile *q*, contrary to the strategy proofness of *g*. (Alternative *x* is at the bottom of *j*'s ordering at *q*, and person *j* is the only one whose preferences change in going from *q* to *r*.) Therefore, g(r) = x. In words, the rule *g* selects *x* at any profile at which person i(Y)'s ordering is the same as it is at *p* (and *q*). Because *x* is at the top of i(Y)'s ordering at *p*, strategy proofness implies that *g* selects *x* at *any* profile at which i(Y) has *x* at the top. (If, say, profile *s* has *x* at the top of i(Y)'s ordering the ordering but g(s) is not *x*, then person i(Y) could manipulate at *s* by reporting the ordering person i(Y) has at profile *p*, because that would precipitate the selection of *x*.)

Let's summarize what we have learned so far. For every three-element subset Y of feasible alternatives there is an individual i(Y) such that whenever that person has a member of Y at the top then that top-ranked alternative will be selected. It is obvious that if i(Y) is the same individual for every three-element subset Y then that individual is a dictator for g itself. We conclude this subsection by proving that i(Y) is the same individual for every three-element subset Y.

Step 3

The rule g is dictatorial.

Proof

Let Y_1 , Y_2 , ..., Y_ℓ be any list of three-element subsets of the basic set of feasible alternatives such that every adjacent pair has two alternatives in common (for instance $Y_1 = \{a, b, c\}$ and $Y_2 = \{b, c, d\}$) and every feasible alternative belongs

to at least one of the sets Y_k . Suppose that x and y belong to both Y_1 and Y_2 . If individual $i(Y_1)$ is different from $i(Y_2)$, let p be any profile for which individual $i(Y_1)$ has x at the top and $i(Y_2)$ has y at the top. By step 2 applied to $i(Y_1)$ we have g(p) = x, but by step 2 applied to $i(Y_2)$ we get g(p) = y. The contradiction forces us to conclude that $i(Y_1) = i(Y_2)$. For the same reason, $i(Y_2) = i(Y_3)$, and so on. We have $i(Y_1) = i(Y_2) = \cdots = i(Y_\ell)$. Let h denote this individual. Let q be any profile, and let a be the alternative at the top of q(h), the ordering of individual h at q. Alternative a belongs to some set Y_k , and $i(Y_k) = h$, so by step 2 g(q) = a, the alternative at the top of h's ordering at q.

We now have a proof of the Gibbard-Satterthwaite Theorem for two individuals and any number of alternatives greater than two.

3.2 Proof for two or more individuals and three or more alternatives

This section extends the proof to the three-person case by showing how to add one more person. We do this by showing that if the theorem is true for n people then it must be true for n+1 individuals. This means that having gone from two to three, we can then go from three to four, and hence from four to five, and so on. By showing how to add one more individual we will have proved the Gibbard-Satterthwaite Theorem for any number of individuals because we already have a proof for the two-person case.

We begin by assuming that we can prove the theorem for arbitrary number $n \ge 2$ of persons. Then we investigate a strategy-proof social choice rule g for n + 1 persons. If every member of X is selected by g at some profile, we wind up demonstrating that g is dictatorial. The key is to use g to define a social choice rule f for n persons. We can then be assured that f is dictatorial. Finally, we show that the dictator for f must be a dictator for g. In that way we will have demonstrated that any strategy-proof social choice rule for a society of n + 1 persons is dictatorial if it is not imposed. (Of course, we assume that X has three or more alternatives.)

Step 1

Given the strategy-proof rule g, which is defined for a group of n + 1 individuals, we define the rule f for an n-person society: If $p = p(1), p(2), \ldots, p(n)$ is an arbitrary profile for n persons we define f by setting $f(p) = g(p(1), p(2), \ldots, p(n), p(n))$. In words, we simply copy person n's preference ordering and let person n + 1 have that ordering also. Then we let the alternative selected by f at p be whatever alternative is selected by g when person n + 1 reports the same ordering as n.

Step 2

The rule *f* is nonimposed.

Proof

For any *x* in *X*, if all n + 1 persons have the same ordering with *x* on top then *g* selects *x*. That is a consequence of the unanimity lemma and the strategy proofness of *g*. Therefore, there is an *n*-person profile at which *f* selects *x*. Hence, *f* is nonimposed.

3. General Proof of the Gibbard-Satterthwaite Theorem

Step 3

The rule f is strategy proof.

Proof

If individual i < n can manipulate f at some profile p(1), p(2), ..., p(n) then that individual can obviously manipulate g at profile p(1), p(2), ..., p(n), p(n) because

$$f(p(1), p(2), \dots, p(n)) = g(p(1), p(2), \dots, p(n), p(n)).$$

Suppose for instance that i = 1 and person 1's true preference ordering is p(1), but when person 1 announces the ordering q(1) the rule f selects alternative z, which ranks higher in p(1) than the alternative y that is selected by f when person 1 reports the ordering p(1). By definition of f, we have

$$z = f(q(1), p(2), \dots, p(n)) = g(q(1), p(2), \dots, p(n), p(n))$$

and

$$y = f(p(1), p(2), \dots, p(n)) = g(p(1), p(2), \dots, p(n), p(n))$$

But this means that the rule g selects z when person 1 announces q(1) and g selects y when person 1 announces p(1), although z ranks higher in p(1) than y. Consequently, person 1 can manipulate g, contradicting the fact that g is strategy proof. We have to drop the supposition that individual 1 can manipulate f. Similarly, we can show that no individual i < n can manipulate f.

Now, consider person *n*. We want to show that person *n* cannot manipulate *f*. This means that we have to show that when persons *n* and n+1 both announce the same ordering q(n) the alternative selected by *g* does not rank higher in p(n) than the alternative selected by *g* when *n* and n+1 both submit the ordering p(n). In symbols, we have to show that for any *n*-person profile $p(1), p(2), \ldots, p(n-1), p(n)$ and any ordering q(n) we cannot have the alternative $f(p(1), p(2), \ldots, p(n-1), q(n))$ strictly preferred to $f(p(1), p(2), \ldots, p(n-1), p(n)$ according to p(n). This simply means that the outcome that is selected by *f* when *n* reports q(n) does not rank higher in her true preference p(n) than the outcome that is selected by *f* when she truthfully reports p(n).

Let *x* denote the outcome g(p(1), p(2), ..., p(n-1), p(n), p(n)), with y = g(p(1), p(2), ..., p(n-1), q(n), p(n)), and z = g(p(1), p(2), ..., p(n-1), q(n), q(n)). Because *g* is strategy proof, if $x \neq y$ then *x* is preferred to *y* according to p(n). Similarly, if $y \neq z$ then *y* is preferred to *z* according to p(n+1), which is identical to p(n). It follows (by transitivity of p(n)) that either x = z or *x* is preferred to *z*. Because

$$x = g(p(1), p(2), \dots, p(n-1), p(n), p(n)) = f(p(1), p(2), \dots, p(n-1), p(n))$$

and

$$z = g(p(1), p(2), \dots, p(n-1), q(n), q(n)) = f(p(1), p(2), \dots, p(n-1), q(n))$$

we have shown that person *n* cannot manipulate *f* when p(1), p(2), ..., p(n-1), p(n) is the profile of true preferences. Because *p* was chosen arbitrarily, we have proved that individual *n* cannot manipulate *f*. We have already demonstrated that no *i* < *n* can manipulate *f*. Therefore, *f* is strategy proof. \Box

Step 4

The rule g is dictatorial.

Proof

It follows from steps 2 and 3 that f is dictatorial because we are assuming that the Gibbard-Satterthwaite Theorem has been proved for *n* persons. (For instance, we could have n = 2.) We prove that the dictator for *f* is a dictator for *g*. Suppose that individual i < n is the dictator for f. Let x be an arbitrary member of X. Let p be a profile for which person j ranks x at the top and everyone else ranks x at the bottom. Then f(p) = x because person j is a dictator for f. But f(p) = x $g(p(1), p(2), \ldots, p(n), p(n))$. Choose any individual $i \le n+1$ different from j. Let profile *q* be the same as p(1), p(2), ..., p(n), p(n) except that we replace p(i)with any other ordering q(i) on X. No matter what that ordering is, the rule g must select x at the new profile because x is ranked at the bottom of p(i) and any change in the outcome when person i changes i's reported ordering would give person i a chance to manipulate at the original profile p. Now, beginning with this new profile we can replace any other individual's ordering, other than that of the dictator *i*, with any other ordering on X. The rule g must still select x, otherwise the individual whose preference has been changed could manipulate g. Proceeding one individual at a time, we can replace the ordering of any individual (other than *j*) with any other ordering. Alternative x will still be selected. This proves that x will be selected at *any* profile for which individual *i* has x ranked at the top. Because x was selected arbitrarily, we have proved that person j is a dictator for g.

Finally, suppose that individual *n* is the dictator for *f*. We show that either person *n* or person n + 1 is a dictator for *g*. Let $\pi(1), \pi(2), \ldots, \pi(n-1)$ be some specific assignment (that we hold fixed until further notice) of preferences to individuals 1 through n - 1. We show that either *g* selects the top-ranked alternative of person *n* for any specification of p(n) and p(n + 1), or else *g* selects the top-ranked alternative of person n + 1 for any specification of p(n) and p(n + 1). We do this by defining a new rule *h* for a two-person society:

$$h(\alpha, \beta) = g(\pi(1), \pi(2), \dots, \pi(n-1), \alpha, \beta)$$

for any specification α and β of the preferences of individuals n and n+1. Because person n is a dictator for f, for arbitrary alternative x in X we must have $h(\alpha, \beta) = x$ when $\beta = \alpha$ and x is ranked at the top of α . Therefore, the rule h is not imposed. Obviously, h is strategy proof because g is. Therefore, his dictatorial. Suppose that $h(\alpha, \beta)$ is the alternative ranked at the top of α for every choice of α and β . Then g selects the alternative ranked at the top of the ordering reported by person n whenever $p(i) = \pi(i)$ for all i < n. Let y denote the alternative ranked at the top of $\pi(1)$ and let z be any other alternative. Choose

3. General Proof of the Gibbard-Satterthwaite Theorem

p(n) and p(n+1) so that *z* is at the top of p(n) and *y* is at the top of p(n+1). Then $z = h(p(n), p(n+1)) = g(\pi(1), \pi(2), \dots, \pi(n-1), p(n), p(n+1))$.

Let q(1) be any ordering of *X*. Define the two-person rule *k* by letting $k(\alpha, \beta)$ be the alternative selected by *g* at profile $q(1), \pi(2), \ldots, \pi(n-1), \alpha, \beta$. We cannot have $g(q(1), \pi(2), \ldots, \pi(n-1), p(n), p(n+1)) = y$ or else person 1 could manipulate *g* at profile $(\pi(1), \pi(2), \ldots, \pi(n-1), p(n), p(n+1))$. Then the argument of the previous paragraph can be used to show that $k(\alpha, \beta)$ is the alternative ranked at the top of α for any choice of α and β . The rule *k* must be dictatorial, and the dictator cannot be the individual whose preference is listed second because k(p(n), p(n+1)) is not *y*.

This time we choose p(n) and p(n+1) so that $\pi(2)$ and p(n+1) have the same alternative ranked at the top, and p(n) has a different alternative at the top. For any choice of q(2), at profile q(1), q(2), $\pi(3)$, ..., $\pi(n-1)$, p(n), p(n+1) the rule g cannot select the alternative ranked at the top of p(n+1) or else person 2 could manipulate g at profile q(1), $\pi(2)$, ..., $\pi(n-1)$, p(n), p(n+1). We continue in this fashion, at each step selecting the next individual i in line and replacing $\pi(i)$ with an arbitrary ordering q(i) on X. Having done so, we fix the preferences of all other individuals up to n-1 and let the preferences of persons n and n+1 vary. This defines a two-person rule, which must be strategy proof because g is, and it must be nonimposed, because g selects the top alternative of p(n) when p(n+1) = p(n). (That's just another way of saying that individual n is a dictator for f.) This new two-person rule must be dictatorial, and person n must be the dictator. When we have finished replacing the preference ordering of every i < n, one person at a time, we will have established that g always selects person n is a dictator for g.

Similarly, if $h(\alpha, \beta)$ always selects the alternative ranked at the top of β then person n + 1 will turn out to be a dictator for *g*.

We have proved the Gibbard-Satterthwaite Theorem for any number n of individuals and any number of alternatives exceeding two. For the sake of logical completeness we conclude with the case n = 1: If g is nonimposed then every alternative x is selected at some ordering of the sole individual. Then if x is on top of the individual's true preference ordering, strategy proofness implies that x must be selected under truthful disclosure, because the individual could otherwise report an ordering at which x is selected.

Source

The extension from two individuals to any larger number is based on Sen (2001).

Problem set

Each of the questions defines a social choice rule. In each case determine whether truthful revelation is a dominant strategy. If it is not, demonstrate that fact with a simple example. If truthful revelation is a dominant strategy then prove your claim by means of a simple, informal—but convincing—argument.

1. There are two feasible alternatives, *x* and *y*, and there are ten individuals. The rule selects the majority winner if there is one, and if there is a tie the rule selects the alternative that is preferred by person 1.

- **2.** There are two feasible alternatives, *x* and *y*, and there are *n* individuals. The rule selects *y* if all persons declare that they prefer *x* to *y*; otherwise the rule selects *x*.
- **3.** There are two feasible alternatives, *x* and *y*, and there are *n* individuals. The rule selects *y* if all persons declare that they prefer *y* to *x*; otherwise the rule selects *x*.
- **4.** There are two feasible alternatives, *x* and *y*, and there are six individuals. The rule selects the alternative that gets the most votes, but individuals 1, 2, and 3 are each allowed to cast three votes for their preferred alternatives, and individuals 4, 5, and 6 may each cast only two votes for their preferred alternatives.

4 THE REVELATION PRINCIPLE

A social choice rule must elicit information about individual preferences so that the outcome of the process will reflect these preferences in the appropriate way. If individuals have an incentive to misrepresent their preference the purpose of the rule is defeated. Is there any social choice procedure that is not vulnerable to this kind of manipulation? To increase the chance of an affirmative answer let us broaden the definition of a social choice rule to include *mechanisms*.

There are three or more alternatives. A mechanism requires individual *i* to announce some message m_i . This message could be a complete description of *i*'s preference scheme; it could be the name of some alternative; it could be both; it could be a list of numbers. Nothing is ruled out, but each particular mechanism will be based on a particular kind of message. In addition, the mechanism specifies which message is to be reported by *i* for each possible preference scheme. (For instance, the market system asks you to choose an affordable consumption plan at which your marginal rate of substitution is equal to the price ratio.) Let $\sigma_i(R)$ denote the mechanism specifies the outcome, or winner, for each possible configuration of messages transmitted by the voters. Let μ be the outcome function. The function μ specifies an outcome $\mu(m)$ in *X* for each profile $m = (m_1, m_2, \ldots, m_n)$ of individual messages.

DEFINITION: Social choice mechanism

A mechanism specifies the type of message, and the set of available messages, for each individual *i*, the behavioral rule σ_i for each *i*, and the outcome function μ to identify the winner. The social choice rules of Sections 1–3, which require individuals to report their preference orderings, are *direct revelation* mechanisms.

The referee observes *i*'s message m_i , but cannot tell whether m_i equals $\sigma_i(R)$ when R is *i*'s true preference scheme. That is because R cannot be observed. If it
4. The Revelation Principle

were verifiable by an outside observer there would be no need for a mechanism in the first place. The only way to ensure that individual *i* sets $m_i = \sigma_i(R)$ is to design the mechanism so that *i* always has an incentive to do so. We want this to be a *dominant strategy*. That is, for every possible preference scheme *R* that *i* might have and whatever messages the others report, there is no message m_i that *i* could send that would result in an outcome that ranks higher in the ordering *R* than the outcome that results when *i* reports $\sigma_i(R)$, given the messages of the others. A mechanism with this property (for each individual *i*) is said to be *strategy proof*.

DEFINITION: Strategy-proof mechanism

The mechanism is strategy proof if for each individual *i*, and each admissible preference scheme for that individual, submitting the message specified by σ_i is a dominant strategy.

Suppose that we have a strategy-proof mechanism μ . Then we can define a strategy-proof social choice rule g (i.e., direct revelation mechanism) by setting $g(\psi) = \mu(m)$ where $m_i = \sigma_i(\psi(i))$ for each i. Suppose that μ is nondictatorial. This means that for any individual i there is an assignment ψ of preferences to individuals such that $\mu(\sigma_1(\psi(1)), \sigma_2(\psi(2)), \ldots, \sigma_n(\psi(n)))$ is not the alternative ranked at the top of $\psi(i)$. But then for any individual i there is an assignment ψ of preferences such that $g(\psi)$ is not the alternative ranked at the top of $\psi(i)$. In other words, g is not dictatorial. Suppose that μ is nonimposed in the sense that for each feasible alternative x in X there is some ψ such that $\mu(\sigma_1(\psi(1)), \sigma_2(\psi(2)), \ldots, \sigma_n(\psi(n))) = x$. Then g is nonimposed, contradicting the Gibbard-Satterwthwaite Theorem. Therefore, every strategy-proof mechanism is either imposed or dictatorial. This observation is known as the revelation principle.

The revelation principle

The Gibbard-Satterwthwaite Theorem extends from direct revelation mechanisms to mechanisms in general.

1.	The	Economic Model	422
	1.1	A continuum of public projects	423
ć	9 1.2	Efficiency	424
ć	91.3	Competitive market equilibrium	428
	1.4	Voluntary contributions	431
ć	9 1.5	Average cost taxation	432
ć	9 1.6	Benefit taxation	433
		Problem set	436
2.	The	Pivotal Mechanism	440
	2.1	The model	441
	2.2	Two options	442
	2.3	Defects of the pivotal mechanism	446
	2.4	A continuum of options	449
	2.5	Relation to the Gibbard-Satterthwaite Theorem	450
		Problem set	451
3.	Gro	ves Mechanisms	453
	3.1	The model	454
	3.2	The mechanisms	456
4.	Effic	ciency and Incentive Compatibility	457
ć	ð 4.1	Dominant strategy equilibrium	458
	4.2	Nash equilibrium	461
		Problem set	465

This chapter continues the exploration of hidden characteristic problems, with attention confined to the problem of preference revelation in the presence of a good that can be consumed jointly and simultaneously by the whole community—a fireworks display, for instance.

Agent *A*'s action generates a *positive spillover* if some other agent *B* benefits as a result of that action. For example, if *A* removes weeds from *A*'s own property, then neighbor *B*'s grass will have fewer weeds because one source of seed has been eliminated. In this case, most of the benefit of *A*'s effort is reaped by *A*, so we say that the spillover is incomplete. However, if *C* produces a fireworks display then everyone else in town will have just as good a view of it as *C*. The spillover is complete in that case. When the agent creating the spillover benefit is not compensated for the positive effect on the welfare of others, we refer to it as an *externality*. Important examples include the containment of a virulent disease by a health organization, the retardation of global warming or ozone depletion by international treaty, and publication of information concerning public safety.

Our aim is to provide the individual decision maker with incentive to consider the benefit that others derive from his or her actions. The decision maker can be a single individual or household, or a region within a country, or even a country itself. When one country takes costly measures to reduce its output of carbon dioxide any resulting retardation of global warming is a benefit that is captured by every country. When a province or a state within a single country imposes restrictions on the firms within its borders to reduce the amount of sulphur dioxide dumped into the air, the benefits are enjoyed throughout the country—to a degree. Pollution in the air above one region can flow to other areas.

When the spillover is complete, as in the fireworks case, we refer to the commodity generating it as a *pure public good*. A public good is created when one individual or institution's action generates widespread benefit, and it is impossible (or very costly) for the agent creating the benefit to be compensated by those receiving it. Most cases that economists treat as pure public goods fall short of the ideal in one way or another, but the polar case is a useful laboratory device for investigating noncooperative behavior.

DEFINITION: Pure public good versus pure private good

A commodity is a pure *public* good if it is possible for every member of the community to consume every unit of the good that has been produced, and the utility derived by anyone is independent of the number of individuals who avail themselves of the opportunity to consume the good. It is a pure *private* good if the only person to benefit is the one doing the consuming.

By definition, any amount of a public good that is made available to one individual or group can be simultaneously enjoyed by everyone in the community,

although not everyone receives the same level of benefit. The higher is the community's level of consumption of the public good *X* the more input is devoted to its production, and hence the larger is the community's sacrifice of private goods due to the diversion of resources to the production of *X*. However, the production technology does not require the individual's sacrifice to be proportional to his or her benefit from consuming the public good. (An individual who makes no contribution at all to the financing of *X* is said to be a *free rider*, because that individual consumes the same amount of *X* as someone who did contribute.)

Truthful preference revelation is very hard to elicit because efficiency requires the amount of public good *X* produced to be a function of reported preferences, and hence so is the share of the financing of *X* contributed by each individual. Without very carefully designed incentives, individuals will be able to misrepresent their preference scheme in a way that significantly reduces their share of the cost of financing the public good—and hence leaves a lot of disposable income for purchasing private goods—without appreciably reducing the level of *X* available for consumption. This would give the individual a net increase in utility, relative to truthful revelation. (With a large number of individuals, the loss of one person's contribution to the financing of *X* will have a tiny effect on the amount of *X* produced.)

The previous chapter examined public sector decision making in a very abstract model. In this chapter we endow the model with much more structure by specifying individual utility functions with classical economic properties, a production function, and resource constraints.

1 THE ECONOMIC MODEL

The hypothesis of the Gibbard-Satterthwaite Theorem of Chapter 7 assumes that any logically possible ordering of the alternatives is a plausible preference scheme for any individual. In an explicitly economic model there are some restrictive properties that individual preferences exhibit. For example, if *x* offers everyone more of every good than *y* then we can rule out individual preferences that have *y* ranking above *x*. Also, indifference curves can often be assumed to be convex, which disqualifies many preference schemes. Perhaps such restrictions on the domain of admissible preferences. We explore this possibility in a simple model with one public good.

A public good is a commodity from which everyone in the community jointly benefits. A pure public good is an ideal case in which every amount produced is consumed in equal measure by all, even though varying levels of benefit are realized. Consider, for example, street lights. In this case the community is the group of residents on one street. Street lights reduce crime. Consider the placement of lights on a typical city street. Would the residents be served best by having one light in front of each house, one light for the entire street, or some intermediate number? The decision *should* depend on householders' preferences. Preferences will take into consideration the benefit in terms of crime reduction from the various plans and the nature and quantity of the goods and services that could be produced instead of a street lamp.

Having the government take the initiative for the project does not automatically solve the problem. The government must discover the extent of household preferences for the public good and for other goods and services that could be produced with the inputs required to manufacture the public good. This is by no means a simple matter. It is one virtue of the market mechanism that for a wide range or goods and services this process of accumulating essential information

Australia's dog fence is 3307 continuous miles of wire mesh, running essentially from one Australian coast to the other. The fence keeps the wild dingo dogs from killing the sheep. The dingoes are on the north side of the fence, and more than 120 million sheep are on the south side (O'Neil, 1997). A failure to repair a hole anywhere in the fence eventually puts all of the sheep in jeopardy. about household preferences is accomplished simply and neatly. (See Chapter 10, or the brief treatment in Section 1 of Chapter 3.)

In the case of any commodity for which the benefits are confined to one individual that is, a private good—a consumer does not derive any benefit unless the consumer pays for the good, and the higher the price, the higher the benefit that must be realized for the consumer to justify the purchase decision. However, when a good provides benefits to the wider community—not only the individual making

the purchase decision—a very high level of overall community benefit can be sacrificed when each individual determines that *his or her* benefit is not great enough to justify paying the purchase price.

1.1 A continuum of public projects

We investigate the possibility of achieving an efficient outcome in the presence of public goods by means of a simple model with two commodities, a pure public good *X* and a pure private good *Y*. Each individual *i*'s utility has the quasi-linear form $U_i(x, y_i) = B_i(x) + y_i$ where *x* is the level of output of the public good and y_i is individual *i*'s consumption of the private good. Therefore, if y_i changes to $y_i + \Delta y_i$ but *x* remains the same, the change in the individual's utility is

$$\Delta U_i = B_i(x) + y_i + \Delta y_i - [B_i(x) + y_i] = \Delta y_i.$$

In brief, if *x* does not change, then for each individual *i* we have $\Delta U_i = \Delta y_i$.

There are *n* individuals, and each individual *i* is endowed with ω_i units of the private good *Y* at the beginning of the period and zero units of the public good. The private good can either be consumed or used as an input in the production of the public good. (We could interpret *x* as an index number. Each value of *x* identifies a particular allocation of resources in every detail except the amount y_i of the private good *Y* delivered to each individual *i*.) Let g(x) denote the amount of good *Y* required to produce *x* units of the public good. The amount

of the private good *Y* available for consumption is $\theta - g(x)$, where θ specifies the amount of *Y* available initially. That is, $\theta = \omega_1 + \omega_2 + \cdots + \omega_n$.

DEFINITION: The public good model

There are *n* individuals, and each individual *i* has a utility function of the form $U_i = B_i(x) + y_i$, where *x* is the amount of the public good produced and y_i is *i*'s consumption of the private good. The production of *x* units of the public good requires g(x) units of the private good as input, leaving a total of $\theta - g(x)$ units of the private good available for consumption. (A total of θ units of the private good are available initially.) We often let *N* denote the set of all individuals.

∂1.2 Efficiency

Section 5.1 of Chapter 2 characterized efficiency with quasi-linear utility: An outcome is efficient if and only if it maximizes total utility. But the proof of that claim drives some individuals' consumption of the private good below zero in some situations. This section characterizes efficient outcomes subject to the constraint that no one's consumption of *Y* can fall below zero. That is, $y_i \ge 0$ must hold for each individual *i* at any feasible allocation. An allocation is said to be *interior* if all consumption levels are strictly positive. We show that an interior allocation is efficient if and only if it maximizes total utility.

DEFINITION: Interior allocation

If x > 0 and $y_i > 0$ for each individual *i* we say that the allocation is interior.

Total utility is $\sum_{i \in N} U_i = \sum_{i \in N} B_i(x) + \sum_{i \in N} y_i$. When we maximize the sum of individual utilities we have to respect the resource constraint, which is $\sum_{i \in N} y_i \le \theta - g(x)$. Efficiency obviously implies that this inequality will be satisfied as an equality, and hence $\sum_{i \in N} y_i = \theta - g(x)$. Therefore, we want to maximize

$$\Sigma_{i\in N} B_i(x) + \theta - g(x),$$

which is a function of a single variable x. Let f(x) denote this function. That is, f(x) is total individual utility after incorporating the resource constraint.

First-order condition for efficiency of an interior allocation when utility is quasi-linear, with positive marginal utility of X

If x > 0, and for each individual *i* we have $y_i > 0$ and $B'_i(x) > 0$, and the allocation (x, y) is efficient, then f'(x) = 0, where $f(x) = \sum_{i \in N} B_i(x) + \theta - g(x)$.

1. The Economic Model

Proof

 $U_i = B_i(x) + y_i$. We change *x* by an amount *dx*, and then y_i by an amount *dy_i*. Then

$$dU_i = B'_i(x) \, dx + dy_i$$

Suppose we can choose dy_i so that $\sum_{i \in N} dy_i = -g'(x) dx$ and $dU_i > 0$ for each *i*. Because -g'(x) dx is the change in the amount of *Y* available for consumption when *x* changes by dx, we wind up with a feasible allocation. Consequently, $dU_i > 0$ for each individual *i* implies that the original allocation is not efficient.

Note that $f'(x) = \sum_{i \in N} B'_i(x) - g'(x)$. Suppose that f'(x) > 0 at allocation (x, y). We show that each individual's utility can be increased. We have

$$\sum_{i \in N} B'_i(x) - g'(x) > 0,$$
[1]

and thus

$$1 - \frac{g'(x)}{B'_1(x) + B'_2(x) + \dots + B'_n(x)} > 0.$$
 [2]

(There are *n* households. If g'(x) > 0 then then the sum of the $B'_i(x)$ is positive. And certainly, the sum of the $B'_i(x)$ must be positive if each $B'_i(x)$ is positive. Therefore, we can divide [1] by the sum of the $B'_i(x)$ without changing the direction of the inequality.)

For arbitrary individual *i*, multiply both sides of [2] by $B'_i(x)$. Because $B'_i(x) > 0$ the direction of the inequality doesn't change. We get

$$B'_{i}(x) - \frac{B'_{1}(x)}{B'_{1}(x) + B'_{2}(x) + \dots + B'_{n}(x)}g'(x) > 0.$$

For dx > 0 we can multiply through by dx, yielding

$$B'_{i}(x) dx - \frac{B'_{i}(x)}{B'_{1}(x) + B'_{2}(x) + \dots + B'_{n}(x)} g'(x) dx > 0.$$
[3]

Now, let

$$t_i = \frac{B'_i(x)}{B'_1(x) + B'_2(x) + \dots + B'_n(x)}$$
[4]

be individual *i*'s share of the cost of increasing *x*. That is, we let $dy_i = -t_i \times g'(x) dx$. Note that [4] implies that $\sum_{i \in N} t_i = 1$. Therefore, $\sum_{i \in N} dy_i = -g'(x) dx$. Moreover, [3] implies that $dU_i = B'_i(x) dx + dy_i = B'_i(x) dx - t_ig'(x) dx$ is strictly positive. Therefore, we have constructed a new feasible allocation that gives everyone more utility than the original allocation at which f'(x) > 0. Therefore, the original allocation is not efficient. (Strictly speaking, we have actually proved that there is some dx > 0 sufficiently small such that the constructed allocation is feasible and gives everyone more utility than the original allocation.)

Now, suppose that $\sum_{i \in N} B'_i(x) - g'(x) < 0$. With t_i given by [4], we *reduce* x to x + dx and increase individual *i*'s consumption of Y by $-t_i \times g'(x) dx$. This will be possible for some dx < 0 sufficiently small in absolute value, because

x > 0. An argument parallel to the one of the previous paragraph shows that everyone's utility will increase if dx is sufficiently close to zero. The new level of consumption of *Y* is guaranteed to be positive, because this time we are reducing the amount of *Y* that is used in producing *X*, and this adds to an individual's private good consumption. This concludes the proof.

The cost shares specified by [4] are critical if we are to increase everyone's utility by moving from x toward x^* , the value of x at which f'(x) = 0. If instead we insist on equal cost shares, then those who derive little marginal benefit from X will suffer a net decline in utility when we increase x. And with equal shares, individuals who have a high marginal benefit from X will find their net utility falling when x is reduced. But if f'(x) is not zero there is *some* system of cost shares that will allow everyone to benefit from a change in the allocation in the appropriate direction. Making all individuals' cost shares proportional to their marginal benefits will always work.

Finally, we assume the *classical second-order condition*: decreasing marginal benefit from *X* and increasing marginal cost of producing *X*.

The classical second-order condition

Assume that $g''(x) \ge 0$ for all x, and for every individual i, $B''_i(x) \le 0$ holds for all x. In addition, either g''(x) > 0 for all x or else for some individual i, $B''_i(x) < 0$ holds for all x.

This condition implies that the value of *x* that maximizes total utility is unique, and that f'(x) = 0 must hold if *f* is maximized by an allocation for which x > 0 and $y_i > 0$ for each *i*.

Characterization of efficiency of an interior allocation when utility is quasi-linear

Assuming the classical second-order condition, an interior allocation (x, y) for which $B'_i(x) > 0$ holds for all x and each individual i is efficient if and only if it maximizes total utility. Moreover, there is a unique level x^* such that every efficient interior allocation satisfies $x = x^*$. We refer to x^* as the efficient level of x.

Proof

In *any* model, whether there is a divisible private good or not, and whether preferences are quasi-linear or not, any allocation that maximizes total utility is efficient. (If outcome F maximizes total utility then there can be no feasible outcome G that gives one person more utility than F and does not leave anyone else with less utility than F. If there were such an outcome G it would provide

more total utility than *F*, contradicting the fact that *F* maximizes total utility.) Therefore, it remains to prove that with a divisible private good and quasi-linear individual utility, the *only* interior allocations that are efficient are the allocations that maximize total utility.

We have seen that in this context efficiency implies that *x* maximizes $\sum_{i \in N} U_i = \sum_{i \in N} B_i(x) + \theta - g(x)$, and we let f(x) denote this function, as before. Note that $f'(x) = \sum_{i \in N} B'_i(x) - g'(x)$, and hence the classical second-order condition implies that f''(x) < 0 holds for all *x*. Therefore, *f* has a unique global maximum, say at x^* . In addition, f'' < 0 implies that f'(x) > 0 holds if $0 < x < x^*$. Therefore, $0 < x < x^*$ and $y_i > 0$ for all *i* implies that allocation (x, y) is not efficient via the first-order condition for efficiency. Similarly, f'' < 0 implies f'(x) < 0 if $x > x^*$, and thus $x > x^*$ and $y_i > 0$ for all *i* implies that (x, y) is not efficient, again by the first-order condition for efficient unless $x = x^*$ and $\sum_{i \in N} y_i = \theta - g(x)$.

We have finished our proof of the efficiency theorem for the classical secondorder condition. We have demonstrated that f'(x) = 0 is necessary and sufficient for efficiency, given x > 0 and $y_i > 0$ for each i and $\sum_{i \in N} y_i = \theta - g(x)$. Because $f(x) = \sum_{i \in N} B_i(x) + \theta - g(x)$, we have $f'(x) = \sum_{i \in N} B'_i(x) - g'(x) = 0$ at an efficient outcome. This is known as the *Samuelson efficiency condition*. The *sum* of the marginal utilities from X must equal the marginal cost of X at an efficient allocation.

Samuelson efficiency condition

$$\Sigma_{i\in N}B_i'(x)=g'(x).$$

If this condition is satisfied by a unique value of *x* we refer to that value as the Samuelson level (or efficient level) of the public good.

If the classical second-order condition is satisfied, x^* is the Samuelson level of the public good, and $\sum_{i \in N} y_i = \theta - g(x^*)$, then $\sum_{i \in N} U_i$ is maximized subject to the resource constraint and thus the allocation is efficient. The amount of Yavailable for consumption is $\theta - g(x^*)$, and if we redistribute it *total* utility is unchanged. (For instance, if we increase y_1 by δ units and reduce y_2 by δ units we don't affect total utility.) Because total utility remains the same after the redistribution, total utility is still maximized and hence the new allocation is efficient. (*Any* allocation that maximizes total utility is efficient in *any* model.) Therefore, assuming the classical second-order condition, an allocation at which all consumption variables are positive is efficient if and only if it provides exactly x^* units of the public good, where x^* satisfies the Samuelson condition, uses exactly $g(x^*)$ units of Y in producing the public good, and redistributes the remaining amount of Y to consumers.

Example 1.1: Two individuals and constant marginal cost

Individual 1's utility function is $U_1(x, y_1) = 10 - 10(x + 1)^{-1} + y_1$ and individual 2's utility function is $U_2(x, y_2) = 15 - 15(x + 1)^{-1} + y_2$. Each unit of the public good is produced by employing 1 unit of the private good as input. Thus, g(x) = x. A total of 120 units of the private good is available initially ($\theta = 120$):

$$U_1 + U_2 = 25 - \frac{25}{x+1} + y_1 + y_2 = 25 - \frac{25}{x+1} + 120 - x.$$

Therefore, we maximize $f(x) = 25 - 25(x+1)^{-1} + 120 - x$. We have $f'(x) = 25(x+1)^{-2} - 1$. The second derivative of f is $f'(x) = -50(x+1)^{-3}$, which is negative for all $x \ge 0$. Thus if we set f'(x) = 0 we get a unique global maximum.

The statement $25(x + 1)^{-2} - 1 = 0$ implies $25 = (x + 1)^2$ and hence $x^* = 4$. The efficient level of the public good is 4. Therefore, at an efficient outcome we have $y_1 + y_2 = 120 - 4 = 116$. An interior allocation is efficient if and only if x = 4 and $y_1 + y_2 = 116$.

∂1.3 Competitive market equilibrium

Why can't we leave the selection of a public project to the market system? Competitive markets will yield an efficient outcome *if* we create a market for each spillover effect. For instance, there would have to be one market in which the benefit conferred on agent 1 by agent 2's actions would be priced. However, most of these markets would be thin, with few agents. When markets are thin there will be a strong incentive to deviate from the competitive rules. What about having the public good itself traded in a conventional market? A simple example will show why this usually leaves us far from an efficient outcome.

Example 1.2: Competitive equilibrium with three agents

Let $U_1 = \ln(x + 1) + y_1$, $U_2 = 2\ln(x + 1) + y_2$, $U_3 = 3\ln(x + 1) + y_3$, and g(x) = x. Each agent is endowed with eight units of the private good. The Samuelson level of the public good solves

$$\frac{1}{x+1} + \frac{2}{x+1} + \frac{3}{x+1} = 1.$$

Hence, $x^* = 5$.

What is the competitive equilibrium for Example 1.2? Let P_X be the price of the public good, with P_Y denoting the price of private good. The production of x units of the public good will cost $P_Y \times x$ because x units of Y will be employed as input. The revenue from x units of the public good will be $P_X \times x$. Therefore, if a firm produces x units of the public good its profit will be

$$P_X \times x - P_Y \times x = (P_X - P_Y)x.$$

1. The Economic Model

On one hand, if $P_X > P_Y$ then profit can be made arbitrarily large by producing an arbitrarily large amount of *X*. The supply of the public good will exceed the demand, so we can't have an equilibrium at which P_X exceeds P_Y . On the other hand, if $P_Y > P_X$ then the firm will take a loss if *x* is positive, and the loss is larger the larger is *x*. When $P_Y > P_X$ the firm maximizes profit by setting x = 0. In that case, the demand for the public good will exceed supply if even one consumer has a positive demand for *X*. Therefore, competitive market equilibrium requires $P_X = P_Y$, and for convenience we let the price of each good be \$1.

For the economy of Example 1.2, when $P_X = 1 = P_Y$, and individual *i* purchases x_i units of the public good and consumes y_i units of the private good, we must have $x_i = 8 - y_i$. (The individual begins with 8 units of *Y* but sells $8 - y_i$ units at \$1 each. The money from the sale is used to buy x_i units of the public good.) Because $x_i = 8 - y_i$ we can replace y_i in individual *i*'s utility function of Example 1.2 with $8 - x_i$.

A competitive market equilibrium $(x^*, x_1^*, x_2^*, x_3^*, y_1^*, y_2^*, y_3^*)$ must satisfy the following five conditions:

$$\begin{aligned} x_1^* \max & \max(x_1 + x_2^* + x_3^* + 1) + 8 - x_1 & \text{subject to } 0 \le x_1 \le 8. \\ x_2^* \max & \max(x_2 + x_1^* + x_3^* + 1) + 8 - x_2 & \text{subject to } 0 \le x_2 \le 8. \\ x_3^* \max & \max(x_3 + x_1^* + x_2^* + 1) + 8 - x_3 & \text{subject to } 0 \le x_3 \le 8. \\ y_1^* & = 8 - x_1^*, \ y_2^* & = 8 - x_2^*, \ y_3^* & = 8 - x_3^*. \\ x^* & = x_1^* + x_2^* + x_3^*. \end{aligned}$$

If the inequality $x_i \le 8$ doesn't hold then y_i is negative because $y_i = 8 - x_i$. Note that the last two conditions imply that the total consumption of the private good equals 24 minus the amount of the private good needed to produce x^* units of the public good. To maximize individual *i*'s utility subject to $0 \le x_i \le 8$ we exploit the fact that if the marginal benefit to *i* of consuming another unit of *X* exceeds the marginal cost to *i* (which is 1) then *i* will demand more *X*. However, if the marginal benefit to *i* is less than *i*'s marginal cost then *i* will reduce x_i , unless it is already 0.

Set $x^* = 2$, $x_1^* = 0 = x_2^*$, and $x_3^* = 2$. If $y_1^* = 8 = y_2^*$ and $y_3^* = 6$, then the five conditions are satisfied, and we have a competitive equilibrium. Moreover, this is the only competitive equilibrium because the five necessary conditions for market equilibrium imply that

$$\frac{1}{x^*+1} \le 1$$
, $\frac{2}{x^*+1} \le 1$, and $\frac{3}{x^*+1} \le 1$,

with the first of these inequalities holding as an equality if $x_1^* > 0$, the second holding as an equality if $x_2^* > 0$, and the third holding as an equality if $x_3^* > 0$. Obviously, only the third can hold as an equality if all three are satisfied. Can we have all three holding as a strict inequality? If so, the five equilibrium conditions imply that $x_1^* = x_2^* = x_3^* = 0$ and hence $x^* = 0$ and

$$\frac{3}{x^*+1} > 1,$$

Table 8.1	Та	bl	е	8.	1
-----------	----	----	---	----	---

	Competitive equilibrium	Government outcome
$egin{array}{c} U_1 \ U_2 \ U_3 \end{array}$	ln 3 + 8 = 9.099 2 ln 3 + 8 = 10.197 3 ln 3 + 6 = 9.296	ln 6 + 7.5 = 9.292 2 ln 6 + 7 = 10.584 3 ln 6 + 4.5 = 9.875

which is a contradiction. Therefore, $3/(x^* + 1) = 1$ and hence $x^* = 2$. Then

$$\frac{1}{x^*+1} < 1$$
 and $\frac{2}{x^*+1} < 1$,

and thus $x_1^* = x_2^* = 0$.

The fact that $x_1^* = x_2^* = 0$ at equilibrium means that agents 1 and 2 are *free riders*. They each benefit from the 2 units of the public good financed by agent 3 without contributing themselves.

The competitive equilibrium is not efficient. *If* the government knows the agent utility functions, it could produce 5 units of the public good and impose taxes of 0.5 units of the private good on agent 1, and 1 unit on agent 2, and 3.5 units on agent 3. Table 8.1 shows each agent's utility at each allocation. Utility is higher for each agent at the new allocation than it is at the competitive equilibrium, confirming that the latter is not efficient. (Note that taxes of 0, 1, and 4 on individuals 1, 2, and 3 respectively would also serve to demonstrate the inefficiency of the competitive equilibrium.)

We have assumed that at a competitive equilibrium one agent cannot collect a payment from another for the benefit that the former confers on the latter by purchasing public goods. Therefore, all agents must pay the competitive price if they want to contribute to the production of the public good. Agents for

In the seventeenth century the English Crown gave a private firm the exclusive right to collect a port tax if it erected and maintained a lighthouse (Coase, 1974). However, even though the private provision of the public good was viable under this arrangement, the monopoly firm would exclude ships that did not derive enough benefit from using the port to justify paying the fee. If those firms were allowed to use the facility their payoff would increase, and the payoffs of other agents would be unchanged. Therefore, the monopoly equilibrium is inefficient-but not woefully so in this case.

which the marginal benefit of the public good is low will prefer to spend their income on other commodities, and enjoy—for free—whatever amount of the public good is financed by others.

Under special circumstances, the market system can provide close to the efficient level of the public good. In the case of a lighthouse there is a very narrow range of payoff functions and technologies that need to be considered, and for each of these there is a unique Samuelson number of lighthouses: Zero is too little and two is too much. One lighthouse is the socially optimal number. This means that only the question of viability remains to be settled. Any scheme that collects enough revenue to finance the operation of the lighthouse will

1. The Economic Model

do. Typically, the Samuelson output level is a function of preferences and technology, there is a wide range of output levels that would be efficient for some plausible configuration of individual preferences, and the market system will often be far from the Samuelson level.

1.4 Voluntary contributions

The previous subsection demonstrated that when a public good is allocated by the market system the outcome is usually inefficient. At some level of X below the efficient level, no individual will want to purchase an additional unit of the public good because the cost to the individual decision maker would exceed the benefit to that individual. But because of the spillover effects, the total benefit to the community of an additional unit of X will exceed the total cost. We run into the same problem if individuals are asked to contribute voluntarily to a fund that will be used to purchase the public good. If individuals determine how much to contribute by comparing the benefit they gets from another dollar spent on the public good to the utility they would have gained by using the dollar to buy more private goods for themselves, then the fund will not collect enough money to finance the efficient level of X.

We can prove our claim that voluntary contributions will not support the efficient level of X by reinterpreting Example 1.2 of the previous subsection: Instead of interpreting x_i as the amount of X purchased by individual i, we can view x_i as the amount of money that individual i voluntarily contributes to the fund. With either interpretation, the amount of the public good consumed by each person is the sum of the x_i , and individual i's consumption of the private good is the amount that i started with minus x_i . For the utility functions of Example 1.2, individuals 1 and 2 will contribute nothing and individual 3 will contribute \$2. Exactly two units of the public good will be made available but the efficient level is five.

Here is a somewhat different example to illustrate the inefficiency of a system of voluntary contributions.

Example 1.3: Cleaning up the neighborhood

A and *B* are neighbors. Each is bothered by the amount of debris (or carbon dioxide, CO_2 , if *A* and *B* are countries) that motorists discharge. If *A* supplies e_A units of effort to cleaning up then *A* and *B* will each derive $2e_A$ units of benefit. Similarly, if *B* supplies e_B units of effort to pollution abatement then *A* will receive $2e_B$ units of benefit from that activity, and so will *B*. We assume that each unit of effort expended by an agent reduces the agent's utility by 3 units. Therefore, the net payoff to *A* when *B* supplies e_B units of effort to cleanup and *A* supplies e_A units of effort is

$$2(e_A + e_B) - 3e_A = 2e_B - e_A,$$

and *B*'s net payoff is $2(e_A + e_B) - 3e_B = 2e_A - e_B$.

If $e_A = e_B = 1$, then the payoff to each is $2 \times (1 + 1) - 3 \times 1 = 1$. However, we have here a continuum version of the prisoner's dilemma game: Each player *i* chooses a number e_i between 0 and 1 inclusive, and for any choice of e_j by the opponent *j*, the response that maximizes player *i*'s payoff is $e_i = 0$, *treating* e_j *as a constant*. Clearly, $e_A = 0$ is a dominant strategy for *A*, and $e_B = 0$ is a dominant strategy for *B*. However, if $e_A = 0 = e_B$ then each player's payoff is 0, but each gets a payoff of 1 from the cooperative outcome $e_A = 1 = e_B$.

The pursuit of self-interest is self-defeating in this case. If the two individuals interact repeatedly over time then the cooperative outcome can be sustained

On Earth Day (April 22, 1990), celebrants in Central Park, New York City, left behind 150 tons of garbage. as an equilibrium. (See Sections 5.7 and 7.2 of Chapter 1.) However, Example 1.3 is intended as a metaphor for the interaction of *many* individuals. If one commuter decides to take public transportation instead of a car—to reduce air

pollution—that commuter cannot assume that it will motivate others to do the same.

Many experiments have been designed to test the Nash prediction of zero voluntary contributions. In fact, around half the efficient level of the public good is typically reached in these experiments, although the level diminishes somewhat when the subjects repeat the experiment.

∂1.5 Average cost taxation

Because an individual's preference scheme is private information, hidden from other agents and the government, if the outcome is guaranteed to be efficient then *any individual* has the ability to determine the level of output of the public good. That is a consequence of the Samuelson efficiency condition. Suppose, for instance, that individual 1 wants exactly 9 units of *X* to be produced. Suppose also that g'(9) = 15 and $B'_2(9) + B'_3(9) + \cdots + B'_n(9) = 13$. Then if individual 1 reports a benefit function B_1 such that $B'_1(9) = 2$ the Samuelson level of *X* is 9. There are lots of functions that would work—for instance, $B_1(x) = 12\sqrt{x}$. We refer to this ability of each individual to control the level of *X* as *unilateral decisiveness*.

DEFINITION: Unilateral decisiveness

For any individual i and any level x of public good output, given the utility functions reported by the other individuals, there is a utility function such that i can guarantee that the Samuelson efficiency condition is satisfied at x simply by reporting that utility function.

Why would we ever have an equilibrium—of any scheme for determining the quantity of *X* produced—if each individual can control the level of output? A successful allocation mechanism must give each individual an incentive to demand the level of output of *X* that satisfies the Samuelson condition with

1. The Economic Model

respect to the true individual preferences, even though the individual cannot know what those true preferences are, apart from his or her own. In short, individuals must have an incentive to report truthfully. This will not happen if the individual cost shares are equal, as we are about to see. Suppose there are *n* individuals, and each must sacrifice one-*n*th of the total amount of the private good necessary to produce the public good.

DEFINITION: Average cost taxation

Each individual *i* reports his or her benefit function B_i , and the Samuelson level *x* is then determined. Individual *i*'s consumption of the private good is reduced by g(x)/n units if *n* is the total number of individuals.

The problem with average cost taxation is that individuals who derive relatively little benefit from the public good will want relatively little of it to be produced, if they have to pay the same share of the cost as everyone else. Consequently, they will not report their true benefit functions. Similarly, those who get a high level of benefit will submit benefit functions that impart an upward bias to the output of the public good.

Example 1.4: Two individuals

Individual 1's utility function is $U_1(x, y_1) = 10 - 10(x + 1)^{-1} + y_1$, individual 2's utility function is $U_2(x, y_2) = 15 - 15(x + 1)^{-1} + y_2$, and g(x) = x. Each individual begins with 60 units of the private good ($\omega_1 = \omega_2 = 60$). This is the setup of Example 1.1. The efficient level of *X* is 4. To see if truthful revelation is a dominant strategy, we can exploit the unilateral decisiveness condition and see how much *X* each individual will "demand."

Maximize $U_i = \alpha - \alpha(x+1)^{-1} + 60 - \frac{1}{2}x$. The first derivative is $\alpha(x+1)^{-2} - \frac{1}{2}$. The second derivative is negative for all $x \ge 0$. When we set the first derivative equal to 0 we get $(x+1)^2 = 2\alpha$. Therefore, $x = \sqrt{2\alpha} - 1$. Individual 1, with $\alpha = 10$, will want 3.47 units of *X* to be produced and individual 2, with $\alpha = 15$, will want 4.48 units. Will both reveal their benefit functions truthfully? Suppose that individual 2 reports his true utility function. Because individual 1 wants $\sqrt{20} - 1$ units of *X* and $B'_2(\sqrt{20} - 1) = 15/20$ she only has to report the benefit function $B_1(x) = 5 - 5(x+1)^{-1}$, for which marginal benefit is 5/20 when $x = \sqrt{20} - 1$. Hence, individual 1 will not report truthfully.

∂ **1.6** Benefit taxation

Absent positive or negative spillovers, a *competitive* market equilibrium is efficient. All individuals choose a consumption plan at which their marginal rate of substitution for any pair if goods is equal to the ratio of the respective prices. Consequently, the price paid by consumers is proportional to the benefit that they derive from the good at the margin. With a public good in our model, we

can imitate the market system by imposing a tax on each individual that is proportional to the benefit that is derived from the public good at the margin.

Specifically, we require individuals to report their benefit functions B_i , and these are used to determine the level x^* of the public good that satisfies the Samuelson condition. This will require a total sacrifice of $g(x^*)$ units of the private good. Each individual *i* contributes the fraction t_i of that cost by having *i*'s consumption of the private good reduced by $t_i \times g(x^*)$. We want t_i to be proportional to *i*'s marginal benefit of *X* at x^* , and we also want the cost shares to sum to unity.

DEFINITION: Benefit taxation

Each individual *i* reports his or her benefit function B_i , and the Samuelson level x^* is then determined. Individual *i*'s consumption of the private good is reduced by $t_i \times g(x^*)$, where

$$t_i = \frac{B'_i(x^*)}{B'_1(x^*) + B'_2(x^*) + \dots + B'_n(x^*)}$$

The benefit tax mechanism is successful on one level: *Assuming* truthful revelation, each individual will want the efficient level of the public good to be produced. To confirm this we maximize $U_i = B_i(x) + \omega_i - t_i \times g(x)$. We have

$$\frac{dU_i}{dx} = B'_i(x) - \frac{B'_i(x^*)}{B'_1(x^*) + B'_2(x') + \dots + B'_n(x^*)} \times g'(x).$$

The Samuelson condition implies $B'_1(x^*) + B'_2(x^*) + \cdots + B'_n(x^*) = g'(x^*)$, so we can write

$$\frac{dU_i(x^*)}{dx} = B'_i(x^*) - B'_i(x^*),$$

which equals zero. Therefore, U_i is maximized at x^* , the efficient level of X, assuming that the benefit functions are reported truthfully. We will *not* get truthful revelation, however, because an individual can report a benefit function for which marginal benefit is zero at every value of x. In that case the individual's tax rate will be zero. The individual will get less benefit from the public good because the level of X will be lower than if he or she had reported truthfully, but in most cases that will be more than offset by the increased consumption of the private good due to a reduced tax burden.

Example 1.5: The benefit tax mechanism with three agents

Let $U_1 = x + y_1$, $U_2 = 2x + y_2$, $U_3 = 5x + y_3$, and $g(x) = \frac{1}{2}x^2$. Each agent is endowed with 24 units of the private good. Because g'(x) = x, the Samuelson level x^* of the public good solves 1 + 2 + 5 = x, and thus $x^* = 8$. If each *i* reports his or her benefit function truthfully, then 8 units of the public good are produced, and person 1's share of the cost is $\frac{1}{8}$. Therefore, $U_1 = 8 + 24 - \frac{1}{8} \times \frac{1}{2} \times 8^2 = 28$ under truthful revelation. To show that the three individuals will not all report truthfully at equilibrium we suppose that persons 2 and 3 do report their true benefit functions. Let's calculate U_1 when person 1 reports the benefit function " $B_1(x) = 0$ for all x." The resulting Samuelson value of x will solve 0 + 2 + 5 = x, and hence 7 units of X will be produced. In that case $U_1 = 7 + 24 - 0 \times \frac{1}{2} \times 7^2 = 31$, which is higher than 1's utility under truthful revelation.

One might object that individuals 2 and 3 can also play the game. That's true, but we have at least shown that truthful revelation is not a dominant strategy for each individual. But let's see what happens when everyone plays strategically. Perhaps a Nash equilibrium of the benefit tax mechanism will precipitate a level of *X* that is not too far from the actual efficient level.

Example 1.6: Nash equilibrium of the benefit tax mechanism

There are *n* individuals, each with the utility function $U_i = x + y_i$. We have $g(x) = \frac{1}{2}x^2$ and each agent is endowed with ω_i units of the private good. For computational convenience, we assume that each individual *i* is known to have a utility function of the form $U_i = \alpha_i x + y_i$, but the exact value of α_i is hidden information. Because g'(x) = x, the Samuelson level must satisfy $\alpha_1 + \alpha_2 + \cdots + \alpha_n = x$, where α_i is the benefit parameter reported by individual *i*. Therefore, $x^* = n$ under truthful revelation, and that is the efficient level of the public good. But we need to find the equilibrium level of *X* when each person plays strategically. Our calculations will be simplified by the fact that the individuals have identical utility functions, but we can't exploit that fact until *after* we maximize individual utility. Otherwise we will ascribe to a single person the ability to control the strategies of others.

Let α be the benefit parameter reported by person *i*, and we let β denote the sum of the benefit parameters reported by everyone else. Then $x = \alpha + \beta$. Therefore, individual *i* will choose α to maximize

$$U_i = \alpha + \beta + \omega_i - [\alpha/(\alpha + \beta)] \times \frac{1}{2}(\alpha + \beta)^2 = \alpha + \beta + \omega_i - \frac{1}{2}\alpha(\alpha + \beta)$$

because person *i's* tax rate is $\alpha/(\alpha + \beta)$. The first derivative of *i's* utility as a function α is $1 - \alpha - \frac{1}{2}\beta$ and the second derivative is negative. Therefore, we maximize U_i by setting $1 - \alpha - \frac{1}{2}\beta = 0$. *Now* we can exploit the fact that the individuals are identical and will make identical decisions. Because β is the sum of the n - 1 reported benefit parameters other than individual *i's* we have $\beta = (n - 1)\alpha$. Therefore

$$1-\alpha-\frac{1}{2}(n-1)\alpha=0,$$

the solution of which is $\alpha = 2/(n+1)$. Each individual's true benefit parameter is 1 but each will report the benefit parameter 2/(n+1) at equilibrium. As a result, only $n \times [2/(n+1)] = 2n/(n+1)$ units of the public good will be produced. Fewer than 2 units of the public good will be available at the benefit tax equilibrium, but the efficient level is *n*. For large *n* there is a vast difference between the benefit tax equilibrium and the efficient allocation.

We have tried the market system, average cost taxation, and the benefit tax mechanism. Are there any public decision schemes that perform better than the market system with respect to the allocation of public goods? The rest of this chapter is devoted to this question. Although the model that we use is extremely simple, it incorporates enough features of resource allocation with public goods to enable us to bring out all the strategic nuances and complications.

Sources

The efficiency condition for public goods is derived by Samuelson (1954). Anderson (2001) reviews the experimental evidence on the provision of public goods. Arrow (1970) demonstrated the competitive equilibrium is efficient *if* there is a market for each spillover effect.

Links

Cornes and Sandler (1996, Chapter 8) thoroughly explores the inefficiency of competitive equilibrium when there are positive spillovers. As the number of players increases in Example 1.3, the gap between the individual payoff when everyone cooperates and the individual payoff when each plays the dominant strategy also increases (Sandler, 1992; Cornes and Sandler, 1996, p. 163).

Problem set

The first three questions pertain to the following simple model of pure public goods production and consumption: X is the public good, and Y is a private good that can either be consumed or used in the production of the public good. There are three individuals. For the first three questions exactly 100 units of the private good are available initially.

- **1.** Let $U_1 = x + y_1$, $U_2 = 2x + y_2$, and $U_3 = 9x + y_3$. The production technology is represented by the real cost function $g(x) = x^2$. Determine the efficient level of output of the public good.
- **2.** Let $U_1 = 2\sqrt{x} + y_1$, $U_2 = 4\sqrt{x} + y_2$, and $U_3 = 6\sqrt{x} + y_3$. When 1 unit of *Y* is used as input, 1 unit of *X* is obtained as output. Determine the efficient level of output of the public good and characterize the set of efficient outcomes.
- **3.** Let $U_1 = 2\sqrt{x} + y_1$, $U_2 = 4\sqrt{x} + y_2$, and $U_3 = 18\sqrt{x} + y_3$. When 1 unit of *Y* is used as input exactly 1/3 of a unit of *X* is obtained as output.
 - **A.** Determine the efficient level of output of the public good and characterize the set of efficient outcomes.
 - **B.** Assume that the current allocation has x = 9, $y_1 = 18$, $y_2 = 15$, and $y_3 = 40$. The output level x = 9 is not consistent with efficiency. Show that some individuals will be made worse off by a move to the efficient level of *X* if the reduction in the consumption of the private good *Y* resulting from the increase in the production of *X* is shared equally by the individuals.

1. The Economic Model

- **C.** Assume that the current allocation has x = 9, $y_1 = 18$, $y_2 = 15$, and $y_3 = 40$ as in part *B*. The output level x = 9 is not consistent with efficiency. Show that everyone can be made better off by a move to the efficient level of *X* by some assignment of cost shares to the individuals. That is, the reduction in the consumption of the private good *Y* resulting from the increase in the production of *X* can be assigned to individuals in a way that leaves everyone better off.
- **D.** Assume that the current allocation has x = 9, $y_1 = 33$, $y_2 = 30$, and $y_3 = 10$. The output level x = 9 is not consistent with efficiency. Show that it is *not* possible to make everyone better off by a move to the efficient level if y_3 is not allowed to fall below 0.
- **E.** Assume that the current allocation has x = 9, $y_1 = 33$, $y_2 = 30$, and $y_3 = 10$ as in part *D*. The output level x = 9 is not consistent with efficiency. Show that it *is* possible to increase *x* by *some* amount in a way that leaves everyone better off than with x = 9.

For the next four questions (4–7) each of the three individuals has an endowment of 40 units of the private good and 0 units of the public good.

- **4.** Let $U_1 = 1\sqrt{x} + y_1$, $U_2 = 2\sqrt{x} + y_2$, and $U_3 = 3\sqrt{x} + y_3$. Each unit of the *X* produced requires 1 unit of *Y* as input.
 - **A.** Characterize the allocation that is defined by transferring all of the private good to person 3 and then choosing *x* to maximize 3's utility subject to the production constraint.
 - **B.** Is this allocation efficient? Explain.
 - **C.** Does this allocation satisfy the Samuelson condition? Support your answer with a simple proof.
 - **D.** Explain how an allocation can be efficient without satisfying the Samuelson condition.
- **5.** Each individual's preferences can be represented by the utility function $15 \ln(x + 1) + y_i$. When 1 unit of *Y* is used as input exactly 1/3 unit of *X* is obtained as output.
 - **A.** What is the marginal social cost of *X*?
 - **B.** Is an allocation for which x = 9 efficient if each individual has a positive amount of the private good? If so, explain why; if not, prove it with a numerical example.
 - **C.** Is an allocation for which x = 29 efficient? If so, explain why; if not, prove it with a numerical example.
- **6.** Let $U_1 = \alpha_1 \sqrt{x} + y_1$, $U_2 = \alpha_2 \sqrt{x} + y_2$, and $U_3 = \alpha_3 \ln(x+1) + y_3$. The production of 1 unit of *X* requires δ units of *Y* as input.
 - **A.** What is the marginal social cost of *X*?
 - **B.** Express the Samuelson condition as simply as you can, then further simplify the condition for the case $\delta = \alpha_3$.

- **7.** Let $U_1 = \sqrt{x} + y_1$, $U_2 = 4\sqrt{x} + y_2$, $U_3 = 5\ln(x+1) + y_3$. The real cost of production is given by the function g(x) = 5x.
 - **A.** What is the marginal social cost of *X*?
 - **B.** Characterize the set of efficient allocations.
- **8.** Let $U_1 = 6\sqrt{x} + y_1$, $U_2 = 12\sqrt{x} + y_2$, $U_3 = 18\sqrt{x} + y_3$, with $\omega_i = 200$ for each *i*. Let *P* denote the price of *X*. Normalize so that the price of *Y* is \$1. When 1 unit of *Y* is used as input exactly 1/3 unit of *X* is obtained as output.
 - **A.** What would be the equilibrium price of *X* in a private ownership market economy with both goods being allocated by the market system?
 - **B.** Will the outcome be efficient? If so, explain why; if not, prove it with a numerical example.
- **9.** Explain why the Samuelson condition would not be satisfied at equilibrium if the level of the pure public good were determined by demand and supply forces in a competitive market—that is, if it were allocated the way private goods are allocated in the market system.
- **10.** Let $U_i(x) = \ln(x+2) + y_i$ for each i = 1, 2, ..., n with $n \ge 3$. Let g(x) = x.
 - **A.** Find the efficient level of *x*.
 - **B.** Show that the situation in which each individual reports a benefit function with a constant marginal benefit of zero constitutes a Nash equilibrium if benefit taxation is employed to determine the output of the public good. Is this Nash equilibrium outcome efficient? (Extend the definition of benefit taxation so that when everyone's marginal benefit is constant at zero then no public goods are produced and no taxes are levied.)
- **11.** Let $B_i(x) = 2\sqrt{x}$ for each i = 1, 2, ..., n, with g(x) = x. Find the equilibrium level of x under benefit taxation when *each* individual reports the benefit function that gives him or her the maximum utility given the reports of others.
- **12.** There are two individuals, and each is endowed with 12 units of the private good. Let $U_1 = 6x + y_1$, and $U_2 = 6x + y_2$. If $g(x) = \frac{1}{2}x^2$ then the Samuelson condition is satisfied by x = 12. But 12 units of the public good require 72 units of the private good as input. The economy has only 24 units of the private good. What has gone wrong with our reasoning?
- **13.** Suppose that the benefit tax mechanism is used. Prove that for any specification of B_1 and g, person 1 maximizes utility by reporting truthfully *if* he or she is certain that every other individual *i* will report that $B_i(x)$ is constant—that is, that benefit does not increase when *x* increases.

Questions 14–18 concern the following special case of our model: $U_i = \alpha_i x + y_i$ for each *i*. The production of *x* units of the public good requires ${}^{1}\!/_2 x^2$ units of the private good as input. An allocation is determined by having each *i* report a benefit parameter β_i , producing $\beta_1 + \beta_2 + \cdots + \beta_n$ units of the public good and collecting a total of ${}^{1}\!/_2 (\beta_1 + \beta_2 + \cdots + \beta_n)^2$ units

1. The Economic Model

of the private goods from households. Specification of the mechanism is complete when we determine by how much each household's consumption of the private good is reduced to obtain the $1/2(\beta_1 + \beta_2 + \cdots + \beta_n)^2$ units that are needed for the production of the public good. (Note that we use α_i to denote individual *i*'s true benefit parameter, with β_i representing the reported benefit parameter.)

- **14.** Suppose that each individual *i*'s consumption of the private good is reduced by ${}^{1}\!/_{2} \beta_{i}(\beta_{1} + \beta_{2} + \dots + \beta_{n})$ units. Suppose that $\alpha_{2} + \alpha_{3} + \dots + \alpha_{n} = 2$. If $2 > \alpha_{1} > 0$ then person 1 is better off reporting a benefit parameter of 0 than reporting the true parameter α_{1} . But what is person 1's *best* strategy? What value of β_{1} , person 1's *reported* benefit parameter, maximizes U_{1} given $\alpha_{1} > 0$ and the fact that the benefit parameters reported by the others sum to 2?
- **15.** Suppose that each individual *i*'s consumption of the private good is reduced by ${}^{1}\!/_{2}\beta_{i}(\beta_{1} + \beta_{2} + \cdots + \beta_{n})$ units. Assume that $\alpha_{1} = 3$ and $\alpha_{i} = 1$ for all i > 1. Prove that we have a Nash equilibrium if person 1 reports a benefit parameter of 3 and everyone else reports a benefit parameter of 0.
- **16.** Assume that $\alpha_1 = \alpha_2 = \cdots = \alpha_n$. Suppose that each individual *i*'s consumption of the private good is reduced by

$$\frac{1}{2} \times \frac{\beta_1}{\beta_1 + \beta_2 + \dots + \beta_n} \times x^2$$

units if *x* units of the public good are produced. In Example 1.6 we saw that a Nash equilibrium resulted in the production of $2x^*/(n+1)$ units of the public good, where $x^* = \alpha_1 + \alpha_2 + \cdots + \alpha_n$ is the efficient level of *x*. Suppose that the center anticipates this and changes the rules so that

$$\frac{(n+1)(\beta_1+\beta_2+\cdots+\beta_n)}{2}$$

units of the public good are produced. Work out a Nash equilibrium for this mechanism. Is it efficient? Does the equilibrium take the society closer to, or farther away from, the original equilibrium at which $\beta_1 + \beta_2 + \cdots + \beta_n$ units of the public good are produced?

- **17.** Assume that the required input of the private good is obtained by proportional taxation, with household *i* paying the fraction t_i of the total cost of whatever level of *x* is provided. Assume also that *n* is odd and that $\alpha_1 > \alpha_2 > \alpha_3 > \cdots > \alpha_{n-1} > \alpha_n$. Find the level of the public good that would defeat every other level by a clear majority. Under what conditions would this majority equilibrium be efficient?
- **18.** Assume that n = 2, $U_1 = 4x + y_1$, $U_2 = 3x + y_2$, and $\omega_1 = \omega_2 = 20$. Each consumer has a one-half share in the ownership of the economy's only firm, which uses *Y* as input to produce *X*. Specifically, *x* units of the public good require $\frac{1}{2}x^2$ units of the private good as input. Suppose that the public good is allocated by the private market system, so that the public good is supplied by a price-taking, profit-maximizing firm to utility-maximizing individuals. Show that x = 4 at the competitive market equilibrium.

(Hint: Begin by determining the firm's supply as a function of the price *P* of its output. For convenience, fix the price of the private good at unity.)

- **19.** There are two individuals, with $\omega_1 = 200$ and $\omega_2 = 100$. The production of x units of the public good requires $1.5x^2$ units of the private good as input. The public good is produced by a single firm, and each individual gets half the firm's profit. Let $U_1 = 48\sqrt{x} + y_1$, and $U_2 = 30\sqrt{x} + y_2$. At a competitive equilibrium of the private ownership market economy we have the following: The price of X is 12, and the price of Y is 1. Individual 1 buys 4 units of X and individual 2 buys 0 units of X. Find the value of y_1 and y_2 at equilibrium *and confirm that we do indeed have an equilibrium*.
- **20.** Prove that the competitive market equilibrium of Section 1.3 really is an equilibrium by showing that all consumers are maximizing utility given their budget constraints, the firm is maximizing profit given the equilibrium price, and demand equals supply for each good.

2 THE PIVOTAL MECHANISM

A successful mechanism for determining the output of the public good must come to terms with the unilateral decisiveness principle. (See Section 1.5 for the definition.) The pivotal mechanism succeeds by making it costly for an individual to shift the level of *X* away from the value that maximizes total utility *under truthful revelation*. We see that the pivotal mechanism gives the individual the incentive to seek the value of *X* that maximizes total utility even though the individual does not know the benefit functions of the other agents!

What makes the pivotal mechanism work is that the cost to the individual of changing the level of X is made equal to the cost imposed on the rest of society by that change. Before defining this scheme we recall the role played by social cost pricing in Vickrey's second-price auction in allocating a single indivisible asset (Section 2 of Chapter 6). There are *n* individuals i = 1, 2, ..., n, and *i*'s reservation value for the asset is V_i . Suppose that $V_1 > V_i$ for all i > 1. Submitting a bid equal to one's reservation value is a dominant strategy, so that each V_i will be known to the auctioneer after the bids have been recorded. If the asset is a Picasso painting then the costs of production are zero, and we don't have to worry about taxing the community to underwrite its production. But giving the painting to individual *i* entails an opportunity cost because others are denied the utility that it would have provided. The cost to the community is V_1 when the painting is given to anyone but person 1, because V_1 is the maximum utility that can be generated by an alternative assignment. If person 1 does get the asset then the cost to the rest of the group is V_h , where h denotes the individual with the second-highest reservation value. Moreover, V_h is the price that will be paid by person 1 when everyone submits a bid equal to his or her true reservation value. No one else pays anything. In that case, we can say that the Vickrey auction requires all participants to pay a fee equal to the cost that their participation has imposed on the rest of society. Anyone whose reservation value is not the highest

2. The Pivotal Mechanism

pays nothing because the outcome would have been the same without his or her participation. Without the high-reservation-value person the asset would have gone to the one with the next highest reservation value so the second-highest reservation value is the amount that the high-reservation-value person pays. The pivotal mechanism for determining the output of a public good also imposes a charge on each individual equal to the cost that his or her participation imposes on the rest of society.

2.1 The model

As in Section 1, there are *n* individuals; each utility function has the quasilinear form $U_i(x, y_i) = B_i(x) + y_i$ (i = 1, 2, ..., n) where *x* is the level of output of the public good and y_i is individual *i*'s consumption of the private good. Each individual *i* is endowed with ω_i units of the private good *Y*, a unit of which can either be consumed or used as an input in the production of the public good, *X*. The production of *x* units of the public good requires g(x) units of *Y* as input. The amount of the private good *Y* available for consumption is $\theta - g(x)$, where θ specifies the amount of *Y* available initially. That is, $\theta = \omega_1 + \omega_2 + \cdots + \omega_n$.

Instead of explicitly tracking the level of the public good, we feature a *public project* that not only specifies the level of *X* but also determines how much of the cost of producing the public good is borne by each individual.

DEFINITION: Public project

A typical public project *F* specifies the menu x_F of public goods produced and, for each individual *i*, the amount $c_i(F)$ by which *i*'s consumption of the private good falls as a result of the fact that some of the private good is diverted from consumption to use as input in producing the public good.

By focusing on public projects we simplify the analysis in two ways. First, feasibility is implicit because we will assume that all the projects under consideration can actually be carried out with available resources and technology. Second, the utility of each individual from each project is determined because $U_i(F) \equiv B_i(x_F) + \omega_i - c_i(F)$ is the utility derived from project F by individual i. (If i's share of the cost were not specified then i's utility from a given level of Xwould be unknown.) Clearly, i will prefer project F to project G if and only if $U_i(F) > U_i(G)$. In words, i prefers F to G if and only if F yields more benefit to i net of the effect on i's consumption of the private good. Note that the project need not specify only quantities of output—it might include a description of how a particular industry is to be regulated.

It will not be necessary to explicitly specify how *i*'s private consumption is reduced by a particular project, but to the extent that resources *are* diverted to produce *X* there will have to be a reduction in the total community consumption of *Y*, and this will inevitably be borne by individuals, although there is a wide variety of ways in which the distribution of this burden can be determined. Whatever specific formula is used is captured by the cost functions c_i . The cost of

producing the public good could be shared equally. Alternatively, each person's cost share could be proportional to his or her wealth, and so on. Different projects could employ different cost sharing formulas. In fact the only difference between projects F and G might be the cost functions employed. That is, F and G might specify precisely the same set of governments activities, and hence the same level of government expenditure, but employ different tax formulas to collect the necessary revenue.

The public projects under consideration could generate *negative* individual benefit. For instance, *F* and *G* could be alternative proposals for disposing of nuclear waste. *F* might require the waste to be dumped in your neighborhood and *G* proposes dumping it in my neighborhood. All of the claims that we make for the pivotal mechanism are valid in cases where the public project reduces the utility of some individuals.

The pivotal mechanism—which is defined in the next subsection—will collect a surtax from each individual equal to the cost that the individual's participation imposes on the rest of society. This surtax is over and above the share of *i*'s cost of government activities determined by the cost function c_i . Because the surtax involves the private good, it has efficiency implications. Our initial goal is *project efficiency*, by which we mean the selection of the project that maximizes total utility *before* any surtaxes are imposed.

DEFINITION: Project efficiency

A mechanism for selecting a single public project from a given set of projects satisfies project efficiency if it always selects the project that maximizes total utility.

We are assuming quasi-linear preferences, so maximization of total utility is necessary and sufficient for efficiency. (We assume that $y_i > 0$ will hold for each *i*.)

We use $V_i(F)$ to denote *i*'s net benefit $B_i(x_F) - c_i(F)$ from project *F*. Because $U_i(F) = B_i(x_F) + \omega_i - c_i(F)$ we have $U_i(F) = V_i(F) + \omega_i$. Let $\sum_{i=1}^n V_i(F)$ denote the sum of the net benefit levels V_i from project *F*. That is,

$$\sum_{i=1}^{n} V_i(F) = V_1(F) + V_2(F) + \dots + V_n(F).$$

Then $\sum_{i=1}^{n} U_i(F) = \sum_{i=1}^{n} V_i(F) + \sum_{i=1}^{n} \omega_i$. Therefore, for any two projects *F* and *G* we have

$$\sum_{i=1}^{n} V_i(F) + \sum_{i=1}^{n} \omega_i > \sum_{i=1}^{n} V_i(G) + \sum_{i=1}^{n} \omega_i$$

if and only if $\sum_{i=1}^{n} V_i(F) > \sum_{i=1}^{n} V_i(G)$. Therefore, project efficiency is equivalent to maximizing $\sum_{i=1}^{n} V_i$.

2.2 Two options

Two public projects have been presented to the community for consideration. (Once we understand how the mechanism works in this setting we examine, in

2. The Pivotal Mechanism

Section 2.4, the choice of *x* from a continuum.) The two proposals are *F* and *G*, and each individual is asked to report the net benefit that she receives from each project. The project that generates the higher level of total utility is adopted and carried out.

DEFINITION: Project selection by the pivotal mechanism

Each *i* is asked to report $V_i(F)$ and $V_i(G)$. Project *F* is adopted if $\sum_{i=1}^n V_i(F)$ exceeds $\sum_{i=1}^n V_i(G)$, otherwise *G* is adopted.

Also each individual is required to pay a surtax equal to the cost that his or her participation has imposed on the rest of society. Typically, individual *i*'s surtax will be zero because the winning project would have generated more total utility even if *i*'s net benefit function had not been included.

DEFINITION: The surtax imposed by the pivotal mechanism

Suppose that $\sum_{i=1}^{n} V_i(F) > \sum_{i=1}^{n} V_i(G)$, and hence *F* is selected. If $\sum_{i=1}^{n} V_i(G) - V_j(G) > \sum_{i=1}^{n} V_i(F) - V_j(F)$ then individual *j* pays a surtax equal to the difference. That is, *j* pays a surtax of

 $\sum_{i=1}^{n} V_i(G) - V_i(G) - [\sum_{i=1}^{n} V_i(F) - V_i(F)].$

If $\sum_{i=1}^{n} V_i(F) - V_j(F) \ge \sum_{i=1}^{n} V_i(G) - V_j(G)$ then j's surtax is zero.

Note that $\sum_{i=1}^{n} V_i(F) - V_j(F)$ is the sum of everyone's net benefit except individual j's. The role of the surtax is to induce each individual to reveal his or her net benefit function V_i truthfully. When the pivotal mechanism is employed, *truthful revelation is a dominant strategy for each individual*, as we show.

Again, suppose that *F* is the winning project. That is, $\sum_{i=1}^{n} V_i(F) > \sum_{i=1}^{n} V_i(G)$. Individual *j* would not have to pay a surtax if $V_j(G) > V_j(F)$. Then even without *j*'s participation, total net benefit from *F* certainly exceeds total net benefit from *G* because that is the case even when we include V_j . Therefore, *j* does not pay a surtax. Nevertheless, the threat of a surtax will prevent this person from overstating the net benefit that he or she gets from *G* in order to precipitate the selection of *G*. Even when *F* is the winning project and $V_j(F) > V_j(G)$, individual *j* will not have to pay a surtax if there is more total net benefit from *F* than from *G* even without including V_j .

The only one who pays a surtax is the individual j who gets greater net utility from the wining project, and the other project would have been selected had V_j not been included. Even so, the surtax is not so large as to induce individual jto deviate from truthful revelation to the extent of understating j's net benefit from the winning project to precipitate the selection of the other project and thus escape the surtax. We test these claims with an example before providing a proof.

	Soren	Rosie	Edie
Project F	10	19	30
Project G	15	10	40
Surtax	0	0	4

Table 8.2. True Net Benefit Levels

Example 2.1: A community of three individuals

The first two rows of Table 8.2 give the net benefit V_i derived by each person from each project. If each individual reports truthfully, project *G* will be adopted because it yields a total net benefit of 65, versus 59 for *F*. Although Soren prefers *G* to *F* he does not pay a surtax because *G* would still win (by 50 to 49) without Soren's participation. Rosie does not pay a surtax because the outcome she prefers was not selected. But Edie pays a surtax because without her participation the outcome would have been *F*, which yields a total net benefit to the rest of the community of 29, versus 25 for *G*. The surtax that Edie pays is the difference, which is 4.

Clearly, Soren has no incentive to deviate from the truth: His preferred outcome is selected without him having to pay a surtax. He can only change the outcome by causing *F* to be selected, and then he would have to pay a surtax (of 1). Rosie could cause her preferred outcome *F* to be selected by overstating her net utility from *F* by 7 or more, but then Rosie would have to pay a surtax of (15 + 40) - (10 + 30) = 15. Rosie's net benefit is 10 when *G* is selected, and it would be 19 - 15 = 4 if she misrepresented his or her preference to ensure the victory of *F*. Clearly, this would not be to her advantage. Edie could avoid the surtax of 4 by understating her preference for *G* (or overstating her preference for *F*) but her net benefit when she tells the truth is 40 - 4 = 36, which is greater than her net benefit when she misrepresents her preference to avoid the surtax. (Her net benefit from *F* is only 30.)

In fact there is no situation in which anyone has an incentive to misrepresent his or her preference.

Pivotal mechanism theorem

Truthful revelation is a dominant strategy for each individual and each specification of the individual net benefit functions.

Proof

Without loss of generality, assume that *F* is selected when everyone reports truthfully. That is, $\sum_{i=1}^{n} V_i(F) \geq \sum_{i=1}^{n} V_i(G)$. Therefore, *if* individual *j* has to pay a surtax the payment will equal

$$\sum_{i=1}^{n} V_i(G) - V_i(G) - [\sum_{i=1}^{n} V_i(F) - V_i(F)].$$

2. The Pivotal Mechanism

We represent this surtax as $\sum_{i \neq j} V_i(G) - \sum_{i \neq j} V_i(F)$, the total net benefit from *G* of everyone in society except *j* minus the total net benefit from *F* derived by everyone in society except for *j*. We examine individual *j*'s decision in three cases.

Case 1. Individual *j* gets at least as much net benefit from *F* as from *G*, and *F* provides more total net benefit even when V_j is excluded.

In symbols, $V_j(F) \ge V_j(G)$ and $\sum_{i \ne j} V_i(F) \ge \sum_{i \ne j} V_i(G)$. There is no incentive for *j* to change the reported V_j to ensure the selection of *G*. That would lower *j*'s utility even without a surtax, although *j* would surely have to pay one if $\sum_{i \ne j} V_i(F) > \sum_{i \ne j} V_i(G)$.

Case 2. Individual *j* gets more net benefit from *F* than from *G*, but *G* provides more total net benefit when V_i is excluded.

That is, $V_j(F) > V_j(G)$ and $\sum_{i \neq j} V_i(G) \ge \sum_{i \neq j} V_i(F)$. This means that under truthful revelation individual *j* pays a surtax of $\sum_{i \neq j} V_i(G) - \sum_{i \neq j} V_i(F)$. Could individual *j* increase his or her utility by deviating from truthful revelation sufficiently to precipitate the selection of *G*? That would allow *j* to escape the surtax, but *j* would realize less net benefit from *G* than from *F*. Because $\sum_{i=1}^{n} V_i(F) \ge \sum_{i=1}^{n} V_i(G)$ we have

$$\sum_{i \neq j} V_i(F) + V_j(F) \ge \sum_{i \neq j} V_i(G) + V_j(G),$$

which implies

$$V_j(F) - [\Sigma_{i \neq j} V_i(G) - \Sigma_{i \neq j} V_i(F)] \ge V_j(G).$$

In words, the net benefit that j gets from F minus the tax surcharge is greater than j's net benefit from G. Therefore, j cannot improve on truthful revelation by precipitating the selection of G.

Case 3. Individual *j* gets more net benefit from *G* than from *F*.

We have $V_j(G) > V_j(F)$. Individual *j* would prefer to see *G* adopted but if *j* were to overstate the net benefit that he or she received from *G*, *j* would surely have to pay a surtax if *G* were selected. That is because *F* wins when *i* reports truthfully, even though $V_j(G) > V_j(F)$, and so it must be the case that $\sum_{i \neq j} V_i(F) > \sum_{i \neq j} V_i(G)$, and thus *j* would have to pay a surtax equal to $\sum_{i \neq j} V_i(F) - \sum_{i \neq j} V_i(G)$. Moreover,

$$\Sigma_{i \neq j} V_i(F) + V_j(F) \ge \Sigma_{i \neq j} V_i(G) + V_j(G)$$

implies

$$V_{i}(F) \geq V_{i}(G) - [\Sigma_{i \neq j} V_{i}(F) - \Sigma_{i \neq j} V_{i}(G)].$$

Therefore, j will not get more net benefit from G after paying the surtax than j gets from F. Therefore, j cannot improve on truthful revelation in this case.

We have demonstrated that truthful revelation is a dominant strategy for each individual. Social cost pricing provides the incentive for individuals to submit their true net benefit functions. In effect, the individual is induced to maximized total net benefit. In both cases 2 and 3 the individual has an incentive to ensure that *F* is selected because $V_j(F) - [\Sigma_{i \neq j} V_i(G) - \Sigma_{i \neq j} V_i(F)] \ge V_j(G)$ in case 2 and $V_j(F) \ge V_j(G) - [\Sigma_{i \neq j} V_i(F) - \Sigma_{i \neq j} V_i(G)]$ in case 3. Both are equivalent to $\Sigma_{i=1}^n V_i(F) \ge \Sigma_{i=1}^n V_i(G)$.

2.3 Defects of the pivotal mechanism

The surtaxes provide just the right incentive for individuals to report their benefit schedules truthfully. The surtax makes it costly for *i* to overstate *i*'s preference for his or her preferred outcome in an attempt to prevent the other from being selected. And because the surtax is independent of *i*'s report, *as long as it does not cause the outcome to change*, there is no incentive to understate one's benefit as there is with benefit taxation. But the surtaxes result in the accumulation of a government budget surplus. (The surplus is 4 in Example 2.1.)

This budget imbalance is inevitable if the incentive to report truthfully is to be maintained! The government could return the surplus to the community but that would alter the incentives and would in fact result in truthful revelation being inferior to some other strategy in some situations. We can demonstrate this by means of the net benefit functions of Example 2.1.

Example 2.2: Equal sharing of the surplus

Suppose that the budget surplus were shared equally by the three individuals. The true net benefit functions are as given in Table 8.2 of Example 2.1. If Rosie were to report $V_R(F) = 24$ and $V_R(G) = 10$ then G would still be selected, this time by 65 votes to 64. However, the surtax paid by Edie would increase by 5 to 9, and Soren would have to pay a surtax of 24 + 30 - (10 + 40) = 4. This means that Rosie would receive a one-third share of a \$13 surplus instead of a one-third share of a \$4 surplus. Her utility would increase from 10 + 4/3to 10 + 13/3 as a result. Therefore, misrepresentation pays in this case. If, however, she overestimates the support for G from the rest of the community and claims a net benefit of 26 from F (instead of 19, her true net benefit) then F would be selected (by 66 to 65) and her net utility would be 19 - (55 - 40) + 5 = 9, which is less than the net utility that she receives when G is selected, even without considering her share of the budget surplus when G is selected. (By reporting a net benefit of 26 she will obtain a net benefit from F of 19, but will pay a surtax of 15 although one-third of that will be returned to her when the budget surplus is paid out to the community.) Nevertheless, we have demonstrated that truthful revelation is no longer a dominant strategy.

2. The Pivotal Mechanism

Truthful revelation is a dominant strategy only when the government budget surplus is not rebated to the community. This means that the pivotal mechanism does not yield efficient outcomes even though it satisfies project efficiency: The other efficiency requirement, $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \omega_i - g(x)$, is violated because the surtaxes by individuals are payments over and above what is required

Table 8.3. True Net Benefit Levels

	Soren	Rosie	Edie
Project F	20	30	40
Project G	30	25	31
Surtax	0	1	5

to purchase the input needed to produce the public good. By definition, the government budget surplus is the sum of the individual surtaxes. Only g(x) units of *Y* are needed to produce *x* units of *X*, but $\sum_{i=1}^{n} y_i$ falls short of $\sum_{i=1}^{n} \omega_i - g(x)$ by the amount of the surplus, which is positive or zero. However, if the number of individuals is large then the surplus will be close to zero on average—because the probability that an individual's

participation changes the outcome will be low if the number of voters is large. By the same token, if the probability that an individual's participation will have an effect on the outcome is virtually zero then there is little incentive to participate in the first place. Thus, the budget surplus problem vanishes only when a more subtle incentive problem emerges.

Perhaps more serious is the fact that the pivotal mechanism can leave someone worse off than if the mechanism had not been used at all.

Example 2.3: Participation can be harmful

As usual, Table 8.3 reveals the true net benefit levels for the respective individuals. Suppose that F is the status quo, with no change in the government's provision of public goods, and G is the consequence of a government project to land on Mars. The status quo will be retained if the pivotal mechanism is used to elicit information about individual preference, but the surtaxes that are necessary to induce truthful revelation leave Edie worse off than if proposal G had not been considered. There is no change in the status quo, but it costs Edie one-eighth of her initial net benefit for the pivotal mechanism to determine that the status quo should prevail.

Another problem is that the mechanism is vulnerable to manipulation by *coalitions*—even coalitions of two people.

Example 2.4: Two-person manipulation of the pivotal mechanism

With the true net benefit functions of Table 8.4, if individuals reported truthfully then G would win, with 90 "votes" to 80, even though the majority prefers F to G. Although *project* efficiency is satisfied by the pivotal mechanism under truthful revelation, if Rosie and Edie collude and each agree to cast 300 "votes" for F

Soren	Rosie	Edie
20	30	30
40	25	25
10	0	0
	Soren 20 40 10	Soren Rosie 20 30 40 25 10 0

Table 8.4. True Net Benefit Levels

Table	8.5.	Reported	Net	Benefit	Levels

	Soren	Rosie	Edie
Project F	20	300	300
Project G	40	25	25
Surtax	0	0	0

then *F* will be selected without either having to pay a surtax. This is confirmed by Table 8.5. Rosie does not pay a surtax because *F* would win without her, and Edie does not pay a surtax for the same reason. Each gets a net benefit of 30 as a result of this ploy, and that is 20% higher than the net benefit that each receives with truthful revelation and the outcome *G*.

Although a single individual cannot manipulate the pivotal mechanism, it is extremely vulnerable to manipulation by a pair of individuals acting in concert. We can augment Example 2.4 by adding a large number n of individuals who have the same net benefit function as Soren. As long as Rosie and Edie each report a net benefit for F of more than 20(n+1) + 25 then F will be selected and neither of the collaborators will have to pay a surtax. Rosie and Edie could successfully collude even if there were millions of other voters.

We are unable to offer a single public decision mechanism that has almost all of the properties that are important for efficient and democratic public policies. In some cases the market mechanism may precipitate a better outcome than any

In the United States, the agencies responsible for liquidating failed property and casualty insurance companies recover only 33% of the book value of the companies' assets on average and only 41% of the liquid assets. Part of the problem is that the state regulatory agents whose job it is to liquidate the assets of an insolvent insurance company pay their own expenses first and do not have much incentive to maximize the proceeds from the sale of assets (Hall, 2000).

collective choice process, especially if the degree of spillover is moderate and the cost of exclusion is low. (The cost of exclusion is low if, as in the case of a lake with a single access road, individuals who don't pay a user fee can be prevented from enjoying the public good. In the case of mosquito control, the cost of exclusion is high.) However, if exclusion is costly and the degree of spillover is high, then the political process has a chance of outperforming the market system, but a lot will depend the incentives provided to key personnel: Public provision of the good will require appropriate effort by the relevant government officials. This effort

cannot be taken for granted because of another private information problem hidden action. Just as there is a possibility that the mechanic hired to repair your car will shirk and charge you for work that was not done, so government agents can contribute less than their best effort.

This is not to suggest that the private ownership market economy does not have its own hidden action problems. However, the preference revelation

2. The Pivotal Mechanism

problem for the allocation of private goods—commodities with very limited spillovers—disappears as the number of agents gets arbitrarily large, but it gets worse when public goods are added to the model.

2.4 A continuum of options

We return to the case where *x* is a real number, interpreted as the level of output of the public good. A project is a specification of the quantity *x* along with the amount $c_i(x)$ by which each individual *i*'s consumption of the private good is reduced to obtain the input of *Y* needed to produce *x*. Of course, $\sum_{i=1}^{n} c_i(x) =$ g(x), where g(x) is the amount of the private good *Y* required as input to produce of *x* units of the public good. We wish to elicit an individual's entire benefit function $B_i(x)$ so that the level of *x* that maximizes $\sum_{i=1}^{n} [B_i(x) + \omega_i - c_i(x)]$, or total utility, can be identified. Note that

 $\Sigma_{i=1}^{n} B_{i}(x) + \Sigma_{i=1}^{n} \omega_{i} - \Sigma_{i=1}^{n} c_{i}(x) = \Sigma_{i=1}^{n} B_{i}(x) - g(x) + \Sigma_{i=1}^{n} \omega_{i}(x),$

so total utility is maximized by the Samuelson level of x.

Social cost pricing is the key to the success of the pivotal mechanism, which requires each agent *i* to report *i*'s net benefit function $V_i(x) = B_i(x) - c_i(x)$ and then produces the efficient project. When *x* varies along a continuum, the calculation of the cost to the rest of society of an individual's participation is more complicated than in the two-option case, but the principal is the same. Note that $\sum_{i=1}^{n} V_i(x) = \sum_{i=1}^{n} B_i(x) - g(x)$, so that $\sum_{i=1}^{n} V_i(x)$ is maximized by the Samuelson level of *x*.

DEFINITION: The pivotal mechanism

If R_j is the net benefit function reported by agent j and x^* maximizes the sum of the reported net benefit functions, then x^* units of X are produced, each individual j's consumption of Y is reduced by $c_j(x^*)$, and, in addition, each individual j pays a surtax of

$$t_j = \sum_{i \neq j} R_i(x^j) - \sum_{i \neq j} R_i(x^*)$$

where x^{j} is the value of *x* that maximizes $\sum_{i \neq j} R_{i}(x)$.

When the reported net benefit functions are the true ones, the efficient (Samuelson) level of the public good is produced. Therefore, the pivotal mechanism satisfies project efficiency. (It would be fully efficient if it were not necessary to collect the surtaxes.) It is easy to see that t_j is the cost to the rest of society of individual *j*'s participation: The surtax t_j is equal to the maximum total utility that the rest of society could realize without consulting *j*'s preferences minus the total utility that the rest of society receives with *j*'s preferences included and with the resulting project x^* adopted.

It remains to show that truthful revelation is a dominant strategy for each individual in every possible situation. That is, whatever benefit functions are reported by others, arbitrary agent j cannot gain by reporting a net benefit

function that is not *j*'s true $V_j = B_j - c_j$. In other words, there is no function R_j that will precipitate a higher level of utility for *j* according to *j*'s true benefit function V_j than the utility level that results from reporting that true V_j .

To prove this we temporarily adopt the unilateral decisiveness principle of Section 1.5: For any value of x, an arbitrary individual can guarantee that x is selected as the level of the public good by reporting the appropriate net benefit function. Consider arbitrary individual j. Let R_1, R_2, \ldots, R_n be the reported benefit functions, with V_j denoting agent j's true benefit function. Given the mechanism's cost rules, which value of x maximizes j's utility?

$$U_{j} = V_{j}(x) + \omega_{j} - t_{j} = V_{j}(x) + \omega_{j} - \sum_{i \neq j} R_{i}(x^{j}) + \sum_{i \neq j} R_{i}(x)$$

= $V_{j}(x) + \sum_{i \neq j} R_{i}(x) + \omega_{j} - \sum_{i \neq j} R_{i}(x^{j}).$

Because $\omega_j - \sum_{i \neq j} R_i(x^j)$ is independent of *j*'s reported net benefit function, it can be treated as a constant from the standpoint of *j*'s maximization problem. Therefore, the value of *x* that maximizes U_j is also the value of *x* that

The pivotal mechanism has been tested in laboratory experiments. Surprisingly, the efficient outcome emerged only 70% of the time, even though truthful revelation is a dominant strategy (Attiyeh, Franciosi, and Isaac, 2000). This might be due to the mechanism's vulnerability to collusion. maximizes $V_j(x) + \sum_{i \neq j} R_i(x)$. What reported net benefit function by individual *j* will lead to the value of *x* that maximizes $V_j(x) + \sum_{i \neq j} R_i(x)$? If individual *j* submits $R_j(x)$ then the value of *x* selected will be the one that maximizes $R_j(x) + \sum_{i \neq j} R_i(x)$. Clearly, individual *j* can do no better than setting $R_j(x) = V_j(x)$. This is the case *even if* individual *j* could force any value of *x* to be produced. Therefore, without unilateral decisiveness individual *j* would

still want the same *x*. In other words, reporting the true $V_j(x)$ is a dominant strategy (The dominant strategy property follows from the fact that we have demonstrated that setting $R_j(x) = V_j(x)$ is a best response by *j* whatever the net benefit functions $R_i(x)$ reported by the other individual, whatever their motivation.)

2.5 Relation to the Gibbard-Satterthwaite Theorem

Truthful revelation is a dominant strategy for each individual when the pivotal mechanism is used to select a public project. Why does that not contradict the Gibbard-Satterthwaite Theorem of Chapter 7 (Section 2)? That theorem establishes that there does not exist a nondictatorial incentive scheme for inducing truthful revelation when there are more than two possible outcomes. The pivotal mechanism is even invulnerable to manipulation when there is a continuum of possible outcomes. And the pivotal mechanism is certainly not dictatorial. To confirm that, suppose that there are *n* individuals and person *j*'s net benefit is 10 from project *A* and 0 from all other projects. Each of the other individuals gets a net benefit of 20 from project *Z* and 0 from all other projects including project *A*. Then *Z* will be selected by the pivotal mechanism, but *j* prefers *A* to *Z*. Therefore, *j* is not a dictator.

The key difference between the framework of the Gibbard-Satterthwaite Theorem and that of the pivotal mechanism is that the latter assumes quasi-linear

2. The Pivotal Mechanism

preferences, with each person's utility being independent of the private goods consumption of other individuals. The Gibbard-Satterthwaite Theorem, however, requires a social choice rule to select an outcome for each logically possible ordering of the feasible alternatives by the individuals. Because the success of the pivotal mechanism depends on surtaxes, these additional payments must be part of the model. Once they are included we can no longer assume that any logically possible ordering is an admissible preference scheme for an individual. For one thing, the outcome that sees project *K* carried out with individual *j* paying a surtax of \$100 will never be preferred by *j* to the adoption of *K* without any additional payment by *j*. However, the proof of the Gibbard-Satterthwaite Theorem employs profiles in which the outcome "*K* plus a \$100 payment by *j*" is at the top of *j*'s preference ordering. In short, the proof of the Gibbard-Satterthwaite Theorem does not go through on the domain of preference profiles for which the pivotal mechanism induces truthful revelation.

Sources

The pivotal mechanism was discovered independently by Groves (1973) and Clarke (1971). (See also Tideman and Tullock, 1976.)

Links

Rob (1982) proved that the budget surplus is virtually zero with a large number of individuals. Green and Laffont (1979) have a simpler proof but they make stronger assumptions about the probability distribution of voter utilities.

Problem set

- 1. In the case of two available (i.e., feasible) options, does the pivotal mechanism select the option that is preferred by a majority ? Explain.
- **2.** Consider the case of three individuals, Soren, Rosie, and Edie, and two public projects *F* and *G*. Table 8.6 gives the benefit derived by each person from

Table 8.6. True Net Benefit Levels

	Soren	Rosie	Edie
Project F	10	20	30
Project G	15	10	40

each project—that is, the benefit net of the taxes assessed to command the resources necessary to construct the project.

A. What project would be undertaken if the pivotal mechanism were employed *and* individuals reported their net benefit figures truthfully? Calculate the surtax for each individual.

- **B.** Show that none of the three individuals has an incentive to misrepresent his or her benefit schedule in this setting of part *A*. What does it mean to say that truthful revelation of preference is a dominant strategy in this case?
- **C.** What is the size of the government budget surplus associated with your answer to question *A*?
- **3.** Repeat question 2 for the payoffs of Table 8.7.

Table 8.7. True Net Benefit Levels

	Soren	Rosie	Edie
Project F	4	30	24
Project G	10	20	30

4. Table 8.8 gives you the true net benefit levels for each of *three* public projects *F*, *G*, and *H* for each of three individuals. Repeat question 2 for this case.

Table 8.8. True Net Benefit Levels

	Soren	Rosie	Edie
Project F	5	18	33
Project G	10	25	18
Project H	15	10	28

- **5.** Prove that when the Samuelson level of output of the public good is produced the final allocation will be efficient if and only if the government budget is balanced.
- **6.** At the end of this section we proved that the budget surplus generated by the pivotal mechanism is zero or positive. What can you say about the likelihood of a strictly positive surplus?
- **7.** Suppose that the pivotal mechanism is augmented by having the surplus returned to the community by giving each individual a fraction of the surplus that is proportional to the net benefit that the individual gets from the project that is selected. *By means of a specific numerical example,* show that truthful revelation is no longer a dominant strategy.
- **8.** The pivotal mechanism is employed, and the individuals pay equal shares of the cost of funding public projects (apart from any supplementary tax employed to get them to reveal their true preferences). For the case $B_i(x) = \alpha_i x$ (i = 1, ..., n) with $\alpha_i > 0$ and $g(x) = \frac{1}{2} x^2$, show that the pivotal mechanism induces individual 1 to report his or her true α_1 .

3. Groves Mechanisms

9. The pivotal mechanism is employed, and the individuals pay equal shares of the cost of funding public projects (apart from any supplementary tax employed to get them to reveal their true preferences). There are three individuals, with $B_1(x) = 2\sqrt{x}$, $B_2(x) = 4\sqrt{x}$, and $B_3(x) = 12\sqrt{x}$. For the two cases, *A* and *B*, determine the outcome generated by the pivotal mechanism when g(x) = x. Specify the level of output of the public good, the share of the cost of financing that public good that is borne by each individual, and the surtax paid by each person. Compute the utility of each person at the pivotal equilibrium and compare it with the utility that each enjoys when no public goods are produced and everyone simply consume his or her own endowment. What does this comparison reveal about the properties of the pivotal mechanism?

A. Let
$$\omega_1 = \omega_2 = \omega_3 = 168$$
.

- **B.** Let $\omega_1 = 40$, $\omega_2 = 100$, and $\omega_3 = 60$.
- **10.** The pivotal mechanism is employed, g(x) = x, and the individuals pay equal shares of the cost of funding public projects (apart from any supplementary tax employed to get them to reveal their true preferences). For n = 3 and $B_1(x) = \alpha \sqrt{x}$, $B_2(x) = \delta \sqrt{x}$, $B_3(x) = \lambda \sqrt{x}$, and arbitrary ω_i show that person 2's pivotal surtax is zero if $\delta = (\alpha + \lambda)$. What is the underlying intuition?
- **11.** Consider the following modification of the pivotal mechanism: Use the same rule for determining the level of output of the public good but change the surtax formula for each individual *j* by dropping the term $\sum_{i \neq j} R_i(x^j)$. Prove that truthful revelation is still a dominant strategy. Assume that the Samuelson value of *x* is positive. Now prove that the government budget is always in deficit as a result of the modified tax formula.
- **12.** Show that the pivotal mechanism defined in Section 2.2 is actually a special case of the mechanism defined in Section 2.4. In other words, prove that when there are only two options the surtax formula of Section 2.4 agrees with that of Section 2.2.
- **13.** Rework the problem of allocating a single indivisible object to a group of *n* individuals by using the framework and notation of this section. Now, show that the allocation rule and surtax rule defined by the Vickrey auction coincides with the respective formulas of the pivotal mechanism.

3 GROVES MECHANISMS

Social cost pricing is the key to the success of the pivotal mechanism. Each individual pays a surtax equal to the cost that the individual's participation imposes on the rest of society. Even when no one is pivotal, and no surtaxes are collected, the *threat* of a surtax prevents individuals who are disappointed with the outcome from tilting it in another direction. But it is not the level of the surtax per se that provides the incentive for truthful revelation, it's the fact that the difference between the surtax paid by individual *j* when *j* reports truthfully and surtax paid when *j* deviates is equal to the change in the cost that *j* imposes

on the rest of society. This fact suggests a generalization of the pivotal mechanism by adding surtaxes (that may be positive or negative) that are independent of the agent's own report.

3.1 The model

There are *n* individuals, indexed by i = 1, 2, ..., n. There is a given (possibly infinite) set $Q = \{q, q', q'', ...\}$ of available public projects. Each individual *i* has a payoff function $U_i = V_i(q) - t_i$, where V_i is a real valued function on Q and t_i is a payment by agent *i* to the government. If t_i is negative then individual *i* receives $-t_i$ dollars from the government. We refer to V_i as *i*'s *net benefit function*—it is net of *i*'s share of the total cost of producing *q*. Our key assumption is that V_i is known only to individual *i*.

Example 3.1: Retardation of global warming

Each individual is one of the countries of the world, and each $q \in Q$ is a proposal to reduce global warming by x(q) percent by imposing an adjustment $\cot c_i(q)$ on each $i \in N$. Then $V_i(q) = B_i(q) - c_i(q)$, where $B_i(q)$ is the benefit that country *i* receives from *q*. Suppose, for instance, that the function B_i is the same for each $i \in N$ (or that each B_i is known). The cost functions c_i are unknown, however, and a central authority wants to learn each country's true cost function so that the burden of adjustment can be imposed on the low-cost countries. We will see that it is possible to design an incentive scheme to induce each agent to report truthfully, in spite of the apparent incentive to claim to be a high-cost country so that the burden will fall elsewhere.

In many applications there are only two projects under discussion, the status quo and a new proposal:

Example 3.2: The binary choice model

The agents can be countries, at one extreme, or individuals in a small community, at the other. $Q = \{0, 1\}$, where 0 represents the status quo and 1 is the new proposal. We can let V_i be a *number*—the difference between *i*'s benefit at the new proposal and at the status quo. If $\Sigma_i V_i > 0$ then $q^* = 1$ (only the new proposal is efficient), and if $\Sigma_i V_i < 0$ we have $q^* = 0$ (only the status quo is efficient).

An important special case of the general framework is the resource allocation model with pure public goods.

Example 3.3: The public goods model of Section 1

Each individual *i* is characterized by a utility function U_i representing *i*'s preferences, and an endowment ω_i of a private good. The utility function is defined over
3. Groves Mechanisms

the space of consumption plans, and agent *i*'s plan (x, y_i) specifies the amount $x \ge 0$ of a public good available to all and *i*'s consumption $y_i \ge 0$ of a single private good *Y*. We further simplify by assuming that utility is quasi-linear: For each $i \in N$ there is a real-valued function B_i defined on the set of nonnegative real numbers, such that

$$U_i(x, y_i) = B_i(x) + y_i.$$

We complete the specification of the fundamental data of the economy by assuming a real cost function g that specifies the amount g(x) of the private good required for the production of x units of the public good.

An *allocation* (*x*, *y*) identifies the level of the public good and the vector $y = (y_i, y_2, ..., y_n)$ of private good consumption levels. Allocation (*x*, *y*) is *feasible* if $x \ge 0$, $y \ge 0$, and

$$\Sigma_i y_i + g(x) \le \Sigma_i \omega_i.$$

This is called the *resource constraint*. In words, the amount of the private good allocated to agents plus the amount used up in producing the public good cannot exceed the sum of the endowments.

Each feasible allocation (x, y) corresponds to the project q for which x(q) = xand $c_i(q) = \omega_i - y_i$. Therefore, the set Q has been implicitly defined. The benefit function is $V_i(q) = B_i[x(q)] - c_i(q)$. The Samuelson condition, of course, is $B'_1(x) + B'_2(x) + \cdots + B_n(x) = g'(x)$.

Our framework can even be used to model the distribution of indivisible private goods. We illustrate with the case of a single indivisible asset.

Example 3.4: Allocation of a single indivisible asset

Q is the set of *n*-tuples $(q_1, q_2, ..., q_n)$ such that $q_i \in \{0, 1\}$ for each *i*, and $\sum_i q_i = 1$. (Agent *i* gets the asset if and only if $q_i = 1$.) If v_i is agent *i*'s reservation value for the asset, then $V_i(q) = v_i$ if $q_i = 1$, and $V_i(q) = 0$ if $q_i = 0$. If the "asset" is undesirable then v_i is negative. For instance, $q_i = 1$ means that garbage (or nuclear waste) is stored in region *i*.

As in the previous section, we strive for project efficiency, which means the selection of the project that maximizes the sum of the net benefit functions.

DEFINITION: Project efficiency

We say that q^* is project efficient if it is feasible (i.e., it belongs to Q) and

$$\sum_{i} V_i(q^*) \ge \sum_{i} V_i(q)$$

for all q in Q.

456 Public Goods and Preference Revelation

This definition is justified by Subsection 5.1 of Chapter 2: With quasi-linear preferences an outcome is efficient if and only if it maximizes total utility. The surtaxes that might be collected will render the entire outcome inefficient, but it will still satisfy project efficiency.

3.2 The mechanisms

We can generalize the surtaxes of the pivotal mechanism to generate the large family of Groves mechanisms. We preserve the property that the difference between the surtaxes paid by arbitrary individual *j* arising from alternative reported net benefit functions is equal to the difference between the respective costs imposed on the rest of society. That will allow us to apply the proof that truthful revelation is a dominant strategy for the pivotal mechanism to this much larger family. This more general approach employs *marginal* social cost pricing.

DEFINITION: Marginal social cost pricing

Marginal social cost pricing is used if alternative decisions D' and D'' each impose costs on the individual making the decision and the difference between that individual's cost arising from D' and the cost arising from D'' is equal to the total cost incurred by the rest of the group as a result of D' minus the total cost incurred by the rest of the group as a result of D''.

If the rest of the group derives benefit from *j*'s action, then the cost imposed on *j* will be negative. That is, individual *j* receives a payment, and marginal social cost pricing requires the difference between *j*'s stipend when *j* adopts D' and *j*'s stipend when *j* adopts D'' to be equal to the total benefit realized by the rest of the group as a result of D' minus the total benefit realized by the rest of the group as a result of D''.

A Groves mechanism selects the project that maximizes total net benefit according to the reported net benefit functions and then imposes surtaxes that embody marginal social cost pricing. That gives us a great deal of latitude, however, in determining the surtax formulas, and that is why we get a large family of mechanisms—one for each specification of the surtax formulas.

DEFINITION: Groves mechanism

Each individual *i* reports a net benefit function R_i , and the project q^* that maximizes $\Sigma_i R_i(q^*)$ is adopted. Each individual *j* pays a surtax

$$t_j = \Phi_j(R_1, R_2, \dots, R_{j-1}, R_{j+1}, \dots, R_n) - \sum_{i \neq j} R_i(q^*)$$

where Φ_j is independent of *j*'s report. Each specification of the Φ_j defines a different Groves mechanism.

4. Efficiency and Incentive Compatibility

We can let Φ_j depend on the messages of individuals other than *i* and the mechanism will still have the dominant strategy property. The key is that the social cost of a *change* in an agent's message will be reflected in a reduction in that agent's transfer.

It remains to show that truthful revelation is a dominant strategy for each individual in every possible situation. That is, whatever benefit functions are reported by others, arbitrary agent *j* cannot gain by reporting a net benefit function that is not *j*'s true $V_j = B_j(q) - c_j$. In other words, there is no function R_j that will precipitate a higher level of utility for *j* according to *j*'s true benefit function V_j than the utility level that results from reporting that true V_j .

Consider the decision of individual 1. Let $R_1, R_2, R_3, \ldots, R_n$ be the reported net benefit functions, with V_1 denoting agent 1's true net benefit function. Given the mechanism's cost rules,

$$U_1 = V_1(q) - t_1 = V_1(q) - \Phi_1(R_2, R_3, \dots, R_n) + \sum_{i>1} R_i(q)$$

Because $\Phi_1(R_2, R_3, \ldots, R_n)$ is independent of 1's reported net benefit function, it can be treated as a constant from the standpoint of 1's maximization problem. Therefore, the project *q* that maximizes U_1 is also the *q* that maximizes $V_1(q) + \sum_{i>1} R_i(q)$, and that can be achieved if agent 1 reports V_1 . Therefore, individual 1 cannot profit from misrepresenting his or her net benefit function. The same argument will work for the other individuals.

Source

This section is based on Groves (1973).

Links

For the public goods model of Section 1 (and Example 1.3) any mechanism for which truthful revelation is a dominant strategy is a Groves mechanism for some choice of the functions Φ_i . This was established by Green and Laffont (1977, 1979). Walker (1978) proved the same result for a much narrower family of preferences. (Walker's theorem is stronger because a mechanism is required to operate successfully over a narrower range of cases.) Danilov and Sotskov (2002, pp. 99–104), work out the extreme assumptions needed to guarantee the existence of a Groves mechanism that always leads to a balanced government budget.

4 EFFICIENCY AND INCENTIVE COMPATIBILITY

Because of the surtaxes that are collected in some situations, the pivotal mechanism does not guarantee budget balance, and hence it does not guarantee an efficient outcome, even though it satisfies *project* efficiency. Now we consider whether there exists *any* mechanism that induces truthful revelation and yields efficient outcomes without sometimes leaving an individual with less utility than if the mechanism had not been used at all. In Section 4.1, we prove that there is no such mechanism. In Section 4.2, we relax the incentive compatibility

458 Public Goods and Preference Revelation

requirement by merely asking that a Nash equilibrium exists and that it yield an efficient outcome. This time we encounter a satisfactory mechanism.

We seek a mechanism for determining the level of output *x* of a single public good, along with the consumption y_i of a single private good by each individual *i*. There are *n* individuals, and each *i* is endowed with ω_i units of the private good. (If the public good is not produced at all then individual *i* consumes $y_i = \omega_i$.) Individual *i*'s preferences are represented by a quasi-linear utility function $U_i(x) = B_i(x) + y_i$. The production of *x* units of the public good requires g(x) units of the private good to be used as input. Therefore, a feasible outcome (or allocation) must satisfy

$$y_1 + y_2 + \cdots + y_n \leq \omega_1 + \omega_2 + \cdots + \omega_n - g(x).$$

Efficiency can be satisfied by giving all of the private good to one individual, say *j*, and then choosing the output of the public good that maximizes *j*'s utility subject to the requirement that *j*'s consumption of the private good equal the total amount available initially minus what has to be used as input in the production of the public good. The outcome would be efficient because it maximizes *j*'s utility, and thus any change would lower it. In other words, it would be impossible to increase anyone's utility without making someone else worse off. This procedure would also satisfy our requirement that truthful revelation is a dominant strategy: Individual *j* clearly has no incentive to misrepresent his or her preference scheme. No one else can profit from misrepresentation either, because no one else's preferences have any influence on the outcome. A simple way of preventing a mechanism from going to such a dreadful extreme is to require it to select a level of output of the public good in a way that leaves all individuals at least as well off as they were before the mechanism was employed. We refer to this as the *participation constraint*.

DEFINITION: Participation constraint

For each individual *i*, utility at equilibrium must be at least as high as *i*'s utility when *i* consumes 0 units of the public good and ω_i units of the private good. That is,

$$B_i(x) + y_i \ge B_i(0) + \omega_i$$

where *x* is the output of the public good at equilibrium and y_i is *i*'s consumption of the private good at equilibrium.

The participation constraint disqualifies absurd mechanisms such as the dictatorial scheme of the previous paragraph.

∂4.1 Dominant strategy equilibrium

We simplify by assuming that there are only two individuals and that the production of *x* units of the public good requires x^2 units of the private good as input. In symbols, $g(x) = x^2$. We also assume that $\omega_1 = 2.5 = \omega_2$. In addition, we limit the benefit functions $B_i(x)$ to those of the form $\beta_i \ln(x + 1)$, where β_i can be

4. Efficiency and Incentive Compatibility

any positive constant. Note that the first derivative of *i*'s benefit function $B_i(x)$ is $\beta_i/(x + 1)$, which is positive for all $x \ge 0$. Applying Section 1.2 to the present case, if $y_1 > 0$ and $y_2 > 0$ then the outcome is efficient only if it maximizes total utility subject to $y_1 + y_2 = \omega_1 + \omega_2 - g(x) = 5 - x^2$. This gives us a very simple model, but even so we prove that there exists no mechanism for which truthful revelation is a dominant strategy, the participation constraint is satisfied, and the outcome is efficient.

We prove that the participation constraint does in fact imply $y_1 > 0$ and $y_2 > 0$, and thus to characterize the efficient outcomes we maximize

$$\beta_1 \ln(x+1) + \beta_2 \ln(x+1) + 5 - x^2.$$
 [5]

The first derivative of this function is

$$\frac{\beta_1}{x+1} + \frac{\beta_2}{x+1} - 2x,$$
 [6]

and the second derivative is therefore equal to

$$-\beta_1(x+1)^{-2} - \beta_2(x+1)^{-2} - 2$$

which is negative for all values of *x*. Therefore, setting [6] equal to 0 will lead to a unique maximum of [5].

We have $\beta_1/(x+1) + \beta_2/(x+1) - 2x = 0$ and thus $2x^2 + 2x - \beta_1 - \beta_2 = 0$, which yields

$$x^* = -0.5 + 0.5\sqrt{1 + 2\beta_1 + 2\beta_2},$$
[7]

the Samuelson level of *x*. (There is a second root for which x < 0, but it is of no interest in this case.) Therefore, an efficient allocation for which $y_1 > 0$ and $y_2 > 0$ must satisfy $x = x^*$ and $y_1 + y_2 = 5 - (x^*)^2$. And if those conditions are satisfied the outcome is efficient because it maximizes total utility.

We now have enough background to consider whether it is possible to design incentives in such a way that all individuals will report their benefit parameters β_i truthfully. We show that it is impossible to do so without violating efficiency or else making someone worse off than he or she was before the public good was provided.

Impossibility theorem:

If truthful revelation is a dominant strategy for a mechanism, it will either fail to deliver an efficient outcome for some specifications of the individual benefit functions or else violate the participation constraint for some specifications of the individual benefit functions.

Proof

We investigate the family of mechanisms that generate efficient outcomes and which satisfy the participation constraint. We show that every member of that family will provide opportunities for an individual to profit from misrepresenting his or her preference in some situations. We begin with the case $\beta_1 = \beta_2 = 2$. Then by [7]

$$x^* = -0.5 + 0.5\sqrt{1+4+4} = 1.$$

Suppose that the mechanism delivers *x* units of the public good and leaves individuals 1 and 2 with y_1 and y_2 , respectively, of the private good. Because we cannot yet rule out $y_1 = 0$ or $y_2 = 0$, at this stage we can only be sure that efficiency implies $x \le 1$, as we now show.

Because the second derivative of [5] is negative it has a unique global maximum at $x^* = 1$ of [7]. If x > 1 we can increase total utility by reducing x to $x^* = 1$. The private good thereby released can be divided equally between the two individuals to increase the utility of each. (Consult Section 1.2 if necessary.)

We have $x \le 1$ at an efficient outcome. Suppose in addition that $y_i = 0$. Then *i*'s utility cannot be greater than

$$2\ln(1+1) < 1.4 < 2.5 = \omega_i.$$

Therefore, the participation constraint implies that $y_1 > 0$ and $y_2 > 0$. It follows that efficiency implies $x = x^* = 1$. This requires 1 unit of the private good as input so we must have $y_1 + y_2 = 5 - 1 = 4$. Therefore, either $y_1 \le 2$ or $y_2 \le 2$. Without loss of generality, assume that $y_1 \le 2$.

We conclude the proof by showing that when individual 1's benefit parameter is $\beta_1 = 2$ that individual can get a higher level of utility by reporting $\beta_1 = 2/9$ than by reporting truthfully. Let x', y'_1 , and y'_2 denote the equilibrium when $\beta_1 = 2/9$ and $\beta_2 = 2$. Note that when $\beta_1 = 2/9$ and $\beta_2 = 2$, condition [7] becomes $x^* = {}^2/_3$. Therefore, efficiency implies $x' \le {}^2/_3$ in that case, because if $x' > x^*$ total utility can be increased by reducing x' to ${}^2/_3$. (The private good thereby released can be divided between the two individuals to increase the utility of each.) The participation constraint implies

$$\frac{2}{9}\ln\left(\frac{2}{3}+1\right) + y_1' \ge \frac{2}{9}\ln(x'+1) + y_1' \ge 2.5$$
[8]

when person 1 reports $\beta_1 = 2/9$. (The mechanism can do no better than work with the benefit functions that it is given, and thus it must provide an outcome that is efficient with respect to those functions.) Now, [8] implies

$$y_1' \ge 2.5 - \frac{2}{9} \ln\left(\frac{5}{3}\right) > 0.$$

But $y'_1 > 0$ and efficiency imply $x' = \frac{2}{3}$. (Because $\beta_2 = 2$ we have already established $y'_2 > 0$.) Finally, $x = \frac{2}{3}$ and $y_1 \ge 2.5 - (2/9) \ln (\frac{5}{3})$ imply

$$U_1 = 2\ln(x'+1) + y_1' \ge 2\ln\left(\frac{5}{3}\right) + 2.5 - \frac{2}{9}\ln\left(\frac{5}{3}\right) > 3.40.$$
 [9]

When person 1 truthfully reports $\beta_1 = 2$ we have $y_1 \le 2$ and $x \le 1$ and thus

$$U_1 = 2\ln(x+1) + y_1 \le 2\ln 2 + 2 < 3.39.$$
 [10]

Statement [10] establishes that person 1's utility is less than 3.39 when that person reports the true benefit parameter $\beta_1 = 2$, but [9] shows that his or her utility

4. Efficiency and Incentive Compatibility

will be higher than 3.40 if that person claims that his or her benefit parameter is $\beta_1 = 2/9$. Therefore, person 1 can manipulate when $\beta_1 = 2 = \beta_2$.

Note that our proof is valid for any mechanism, whether it asks individuals to report their benefit functions directly or uses a more subtle system of messages. No mechanism can satisfy all three properties because efficiency and the participation constraint imply that truthful revelation is not a dominant strategy for all specifications of individual preferences. This is unfortunate because the fact that an individual's preference information is hidden from others means that it must work well for a wide range of individual preferences—because the true preferences are unknown when the community chooses or designs a mechanism.

4.2 Nash equilibrium

A dominant strategy is a best response to *anything* that one's rivals might do. A strategy is a component of a Nash equilibrium if it is merely a best response to what others are *currently* doing. Typically, a Nash equilibrium is not a dominant strategy equilibrium. The latter is much more demanding. Therefore, we relax the requirement that truthful revelation be a dominant strategy for each individual and simply ask for a Nash equilibrium that is efficient. We also impose the participation constraint.

We assume that for each output level q of the public good the cost share $c_i(q)$ of each individual i is exogenously given. Then $V_i(q) = B_i(q) - c_i(q)$ is i's net benefit function. We assume that $c_i(0) = 0$ for each i. In words, if the public good is not produced, then no one is charged.

As in the case of the market process (studied in Section 1.3), each individual *i* has an opportunity to add an amount x_i to the amount of the public good available to all. But we now employ a mechanism that differs from the market mechanism in two important respects: First, the individual is allowed to reduce the output of the public good. In other words, x_i can be negative. At equilibrium, $x_1 + x_2 + \cdots + x_n$ units of the public good are provided. (There are *n* individuals, and in this section we assume that $n \ge 3$.) Second, individual *i*'s consumption of the private good is *not* reduced by x_i multiplied by the price of the public good. That doesn't lead to an efficient outcome, as we saw in Section 1.3. If $q = x_1 + x_2$ $x_2 + \cdots + x_n$ units of the public good are produced, individual *i*'s consumption of the private good is reduced by $c_i(q)$ but *i* receives a *transfer payment* of $x_{i+1} - x_{i-1}$ multiplied by the amount of the public good produced. An individual's transfer could be negative, in which case that individual pays that amount of money into a fund that is used to give positive transfers to others. We use the term transfer payment to emphasize that each dollar received by someone is a dollar paid by someone else.

DEFINITION: Walker's mechanism

Each individual *i* announces a real number x_i and $x_1 + x_2 + \cdots + x_n$ units of the public good are produced, with each individual's share of the cost determined in advance. We will refer to x_i as individual *i*'s *demand* for the

462 Public Goods and Preference Revelation

public good. Given the demands, each individual \boldsymbol{i} receives a transfer payment of

$$(x_{i+1}-x_{i-1}) \times (x_1+x_2+\cdots+x_n)$$
.

(If i = n then i + 1 is interpreted as individual 1, and if i = 1 then individual i - 1 is person n.)

We show that a Nash equilibrium of the Walker mechanism is efficient and that it satisfies the participation constraint. The latter is satisfied because one of the strategies available to individual *i* is to set x_i equal to the negative of the sum of the x_j over all $j \neq i$. That will result in x = 0, $y_i = \omega_i$, and $U_i = \omega_i$. If the individual adopts a different strategy, it must be because it will yield more utility. At a Nash equilibrium, where each individual is employing a best response strategy, individual *i*'s utility must be at least ω_i . Hence, the participation constraint is satisfied. For this subsection only we let *x* denote (x_1, x_2, \ldots, x_n) , the list (or vector) of demands. The total is denoted $\sigma(x)$. That is, $\sigma(x) = x_1 + x_2 + \cdots + x_n$ is the total amount of the public good produced if *x* constitutes a Nash equilibrium. Let $m_i(x)$ denote *i*'s transfer payment.

Now, let $x = (x_1, x_2, ..., x_n)$ be a Nash equilibrium. We show that the outcome is efficient. First, we show that we have budget balance, which means that the transfer payments sum to 0. Individual *i*'s transfer is $m_i(x) = (x_{i+1} - x_{i-1}) \times \sigma(x)$. Then for all x,

$$\frac{1}{\sigma(x)} \times [m_1(x) + m_2(x) + \dots + m_n(x)]$$

= $x_2 - x_n + x_3 - x_1 + \dots + x_1 - x_{n-1}$
= $x_2 + x_3 + x_4 + \dots + x_{n-1} + x_n + x_1$
 $- [x_n + x_1 + x_2 + \dots + x_{n-3} + x_{n-2} + x_{n-1}] = 0.$ [11]

The transfer payments do sum to 0.

We prove that the equilibrium is weakly efficient by showing that any outcome that gives everyone more utility than the Nash equilibrium x is not feasible. Consider the output level q of the public good, and transfer payments $s_1, s_2, \ldots s_n$. Suppose that for each individual i

$$V_i(q) + s_i > V_i(\sigma(x)) + m_i(x).$$

Individual *i* could have adopted a strategy that resulted in *q* units of the public good being provided. Specifically, given the demand x_j of each $j \neq i$, if *i* demands $d_i^i = q - [\sigma(x) - x_i]$ then we have $\sigma(d^i) = q$ if we set $d_j^i = x_j$ for all $j \neq i$. By definition of Nash equilibrium, individual *i* chose to demand x_i instead of d_i^i and hence we must have $V_i(\sigma(x)) - m_i(x) \ge V_i(\sigma(d^i)) - m_i(d^i) = V_i(q) - m_i(d^i)$. Then we have for each *i*,

$$V_i(q) + s_i > V_i(\sigma(x)) + m_i(x) \ge V_i(q) + m_i(d^i)$$

4. Efficiency and Incentive Compatibility

Therefore, $V_i(q) + s_i > V_i(q) + m_i(d^i)$. It follows that $s_i > m_i(d^i)$ for each *i*. (The output of the public good is the same in the two situations, so the preferred outcome must give individual *i* a higher transfer payment.) Therefore,

$$s_1 + s_2 + \dots + s_n > m_1(d^1) + m_2(d^2) + \dots + m_n(d^n).$$
 [12]

But we have $\sigma(d^i) = q$, $d^i_{i+1} = x_{i+1}$, and $d^i_{i-1} = x_{i-1}$ for each individual *i*. Therefore,

$$m_i(d^i) = (x_{i+1} - x_{i-1}) \times q_i$$

which implies that $m_1(d^1) + m_2(d^2) + \cdots + m_n(d^n) = 0$ by [11]. Therefore, [12] implies

$$s_1+s_2+\cdots+s_n>0.$$

The transfer payments $s_1, s_2, ..., s_n$ are not feasible because they have a positive sum. Anyone who "receives" a negative transfer actually pays that amount of money. Because $s_1 + s_2 + \cdots + s_n > 0$, the amount of money received exceeds the amount paid in. This means that the outcome that provides q units of the public good, along with transfer payments s_1, s_2, \ldots, s_n , is not feasible. We have shown that no feasible outcome can give everyone more utility than the Nash equilibrium.

Could there be a feasible outcome that gives, say, individual *j* more utility without leaving anyone else with less utility? No. If that were possible, then individual *j* could give a tiny amount of money to everyone else and we would then have a feasible outcome that gave everyone strictly more utility than the Nash equilibrium. But we have just proved that that is impossible. Therefore the Nash equilibrium is efficient.

∂Example 4.1: Three individuals

The production of *q* units of the public good requires *q* units of the private good as input. We have $c_i(q) = \frac{1}{3}q$ for each *i*. In words, each individual pays a third of the cost of the producing the public good. The respective utility functions are

 $U_1 = \ln(q+1) + y_1$, $U_2 = 2\ln(q+1) + y_2$, $U_3 = 3\ln(q+1) + y_3$.

The transfers are

$$m_1 = (x_2 - x_3)q$$
, $m_2 = (x_3 - x_1)q$, $m_3 = (x_1 - x_2)q$

The efficient level of the public good is value of *q* that maximizes

$$\ln(q+1) + 2\ln(q+1) + 3\ln(q+1) + \omega_1 + \omega_2 + \omega_3 - q_4$$

The first derivative is 6/(q+1) - 1, and when we set this equal to 0 we get q = 5, which is the efficient amount of the public good. Person *i* pays ${}^{1}\!/_{3} q$ as *i*'s share of the cost of producing *q*, and *i* receives a transfer of $t_{i} \times q$.

464 Public Goods and Preference Revelation

Therefore *i* pays a total of $({}^{1}/_{3} - t_{i})q$. At a Nash equilibrium of the Walker mechanism

$$x_{1} \text{ maximizes } \ln(x_{1} + x_{2} + x_{3} + 1) + \omega_{1} + \left[-\frac{1}{3} + x_{2} - x_{3}\right] \times [x_{1} + x_{2} + x_{3}],$$

$$x_{2} \text{ maximizes } 2\ln(x_{1} + x_{2} + x_{3} + 1) + \omega_{2} + \left[-\frac{1}{3} + x_{3} - x_{1}\right] \times [x_{1} + x_{2} + x_{3}],$$

$$x_{3} \text{ maximizes } 3\ln(x_{1} + x_{2} + x_{3} + 1) + \omega_{3} + \left[-\frac{1}{3} + x_{1} - x_{2}\right] \times [x_{1} + x_{2} + x_{3}].$$

The respective first order conditions are

$$\frac{1}{x_1 + x_2 + x_3 + 1} - \frac{1}{3} + x_2 - x_3 = 0,$$

$$\frac{2}{x_1 + x_2 + x_3 + 1} - \frac{1}{3} + x_3 - x_1 = 0,$$

$$\frac{3}{x_1 + x_2 + x_3 + 1} - \frac{1}{3} + x_1 - x_2 = 0.$$

We know that the Nash equilibrium is efficient, and thus $x_1 + x_2 + x_3 = 5$.

$$\frac{1}{6} - \frac{1}{3} + x_2 - x_3 = 0.$$

$$\frac{2}{6} - \frac{1}{3} + x_3 - x_1 = 0.$$

$$\frac{3}{6} - \frac{1}{3} + x_1 - x_2 = 0.$$

The second equation yields $x_3 = x_1$, and the third is $x_2 - x_1 = 1/6$. Because $x_1 + x_2 + x_3 = 5$ we need to solve

$$x_2 - x_1 = \frac{1}{6}$$
 and $2x_1 + x_2 = 5$.

The solution is $x_1 = 29/18$, $x_2 = 32/18$, and $x_3 = 29/18$. These values give us the transfers $m_1 = (1/6)q$, $m_2 = 0$, and $m_3 = -(1/6)q$. Now, substitute these transfers into the respective individual utility functions and then show that q = 5 maximizes each of the functions

$$\ln(q+1) + \omega_1 + \left(-\frac{1}{3} + \frac{1}{6}\right) \times q,$$

$$2\ln(q+1) + \omega_1 + \left(-\frac{1}{3} + 0\right) \times q,$$

$$3\ln(q+1) + \omega_1 + \left(-\frac{1}{3} - \frac{1}{6}\right) \times q.$$

The purpose of the transfers is to adjust each individual's marginal cost of acquiring an additional unit of the public good so that everyone wants the same amount of the public good. That guarantees that a Nash equilibrium exists. By having the transfers sum to zero we not only balance the government's budget, we precipitate an efficient outcome at equilibrium. However, although the average cost taxation mechanism of Section 1.5 always yields an efficient outcome

at equilibrium, a Nash equilibrium almost never exists, which in not the case for the Walker mechanism. That's because each individual's marginal cost of acquiring an additional unit of the public good is exogenously determined by average cost taxation—independently of anyone's preferences. We illustrate with a final example.

∂ Example 4.2: Average cost taxation

The setup is the same as for Example 4.1 except that we employ average cost taxation instead of the Walker mechanism. We have $c_i(q) = \frac{1}{3}q$ for each individual *i* but there are no transfers. If a Nash equilibrium exists and it leads to the production of *q* units of output, then *q* maximizes each of the three functions

$$U_1 = \ln(q+1) + \omega_1 - \frac{1}{3}q,$$

$$U_2 = 2\ln(q+1) + \omega_2 - \frac{1}{3}q,$$

and

$$U_3 = 3\ln(q+1) + \omega_3 - \frac{1}{3}q.$$

The respective first order conditions are:

$$\frac{1}{q+1} - \frac{1}{3} = 0$$
, $\frac{2}{q+1} - \frac{1}{3} = 0$, and $\frac{3}{q+1} - \frac{1}{3} = 0$.

These three conditions can't be satisfied simultaneously. Therefore, there is no Nash equilibrium.

Sources

The impossibility theorem of Section 4.1 was discovered and proved by Leonid Hurwicz (1972) for the case of pure private goods. It was Hurwicz who taught economists how to investigate the role of incentives in general equilibrium resource allocation. We actually employ the theorem presented by Roberts (1979) for the standard public goods model. The mechanism of Section 4.2 was devised by Walker (1981).

Links

See Walker (1980) for a more thorough investigation of the possibility of using dominant strategies to identify an efficient outcome. The Nash equilibria for Walker's mechanism are Lindahl equilibria, and the existence of a Lindahl equilibrium is proved in Foley (1970) and Milleron (1972). See Corchón (1996) and Repullo (1987) for a general treatment of Nash implementation. Chen (2006) discusses the performance of mechanisms in a laboratory setting.

Problem set

 Design a mechanism that satisfies the participation constraint and is invulnerable to manipulation by any individual. (Of course, it won't be the case that the equilibria are always efficient.)

466 Public Goods and Preference Revelation

- **2.** Design a mechanism that satisfies the participation constraint and that always yields equilibria that are efficient. (Of course, it has to be the case that at least one individual can manipulate the mechanism in some situations.)
- **3.** Rework Example 4.1 with $c_1(q) = \frac{1}{4}q$ instead of $c_1(q) = \frac{1}{3}q$, $c_2(q) = \frac{1}{4}q$ instead of $c_2(q) = \frac{1}{3}q$, and $c_3(q) = \frac{1}{2}q$ instead of $c_3(q) = \frac{1}{3}q$.
- **4.** Rework Example 4.2 with $U_1 = \beta_1 \ln(q+1)$, $U_2 = \beta_2 \ln(q+1)$, and $U_3 = \beta_3 \ln(q+1)$. Find a condition on the benefit parameters β_1 , β_2 , β_3 for which a Nash equilibrium of the average cost mechanism exists.
- **5.** Repeat question 3 but with $c_1(q) = \lambda_1 q$, $c_2(q) = \lambda_2 q$, and $c_3(q) = \lambda_3 q$, where λ_1, λ_2 , and λ_3 are given fractions that sum to 1.

Matching

1.	Stud	dents and Advisors	. 469
	1.1	One-to-one matching	469
	1.2	The main results	474
		Problem set	479
2.	Coll	ege Admissions	. 480
	2.1	Students apply to colleges	481
	2.2	Student placement	485
	2.3	Single-test placement	490
		Problem set	495
3.	Hos	pitals and Doctors	. 496
		Problem set	499
4.	Allo	cating Dormitory Rooms	. 499
	4.1	A commonly used scheme	500
	4.2	An efficient procedure	503
		Problem set	509
5.	Kidr	ney Transplants	. 510

This chapter examines allocation problems for which the scarce resources are available only in discrete units, such as dormitory rooms, and each "consumer" wants one unit and *only* one unit. The objective is to match students with rooms, or available kidneys with the patients on a waiting list for a transplant, and so forth. The entire family of matching problems can be subdivided in two different ways. We can classify according to the nature of preferences. On one hand, suppose that an economics department has a given number of students (i.e., majors) and professors, and the objective is to assign each student a professoradvisor. In this case both sides of the match have preferences: Students like some professors better than others, and the professors also have preferences over the students. On the other hand, there are matching problems for which only one side of the match has preferences: Students have preferences for dormitory rooms, but the rooms don't have preferences for students. A matching problem for which only one side has preferences is referred to as an *assignment problem*. The objective is to assign students to rooms, for instance.

The other way to classify matching problems is according to whether we can have an outcome in which some agent is matched more than once. In the case of college admissions, each student will be matched with at most one college, but each college is matched with more than one student. In that case, we say that we have an *admissions* problem. An admissions problem may or may not be an assignment problem. In the case of college admissions, the colleges do have preferences for students—if only for students with high test scores—and the students certainly have preferences for colleges. So college admissions is not an assignment problem.

If an allocation problem belongs neither to the admissions family nor the assignment family we say that it is a *marriage model*. In other words, a marriage model is one in which there are two types of agents: There is a set of *W* types and a set of *M* types. Each member of each type has preferences for the members of the other type, and each agent is matched with at most one member of the other type. The problem of matching students and advisors can be a marriage model. A student will have at most only one advisor, and if there is a department rule limiting a professor to at most one advisee then we have a marriage model. However, in the case of a department in which the professors know nothing about the students—they are freshmen, for instance—the professors will not have preferences for the students, so we have an assignment problem. And if there are more students than professors it is also an admissions problem if some professors are allowed to have more than one advisee. Our objective is not to classify allocation problems, however, but to *solve* them by designing satisfactory allocation procedures.

In general, there are two types of agents, *A*s and *B*s, and the objective is to match the *A*s with the *B*s and vice versa. There will be many ways to do this. If there are ten of each type then the marriage model has 36,288,000 different solutions—not including cases where at least two agents are unmatched. (We allow preferences that declare "I'd rather be unmatched than paired with *him*.") Some of the arrangements are efficient, but many will not be. Clearly, we want an outcome that at least passes the efficiency test. We also want an economical

1. Students and Advisors

procedure for matching the *A*s and *B*s. In particular, there will be very many ways of matching, and we don't want to employ an algorithm that consumes vast amounts of time and money working itself out. Of course, we also want to respect individual rights. We don't want to match agent α with agent β if α would prefer being unmatched to having β as a partner. Respecting rights opens the door to the possibility of manipulation. By declaring that she would rather be unmatched than paired with β , when that is not in fact true, agent α may force the procedure to give her a match that she prefers to the one that she would have had if she had reported her preferences truthfully. Therefore, we look for matching procedures that are incentive compatible, in the sense that they induce truthful revelation of preferences.

To simplify the discussion we assume throughout that all agents have preferences that are strict, in the sense that they are never indifferent between two agents of the opposite type; one will be strictly preferred to the other. Of course, in the case of assignment problems this applies only to the type with preferences. For the most part, we assume that all agents care only about their own match. In other words, agent α 's welfare is unaffected by matches that do not involve α .

1 STUDENTS AND ADVISORS

An economics department has a set M of majors, each of whom is referred to as type M. There is also a set W of professors, the type W agents. Some of the majors want to do an honors thesis, and that requires having a professor as an advisor. Each thesis writer needs only one advisor, and the department has a rule preventing a professor from advising more than one student. Each professor has encountered each student in one or more classes, so the professors have preferences for the students. And the students certainly have preferences for the professors. Hence, we have a marriage model, the subject of this section.

1.1 One-to-one matching

In this section we use the evocative language of matching men and women for expositional purposes, even though we actually are investigating institutional procedures for determining satisfactory matches.

DEFINITION: Marriage model

There are two types of agents, *W* and *M*. Each *W* will be matched with at most one *M*, and each *M* will be matched with at most one *W*. The *W*s have preferences for the *M*s and vice versa. A specific *matching* $\pi = \{(A, Z), (B, Y), \ldots\}$ is a set of pairs: Agent *A* belonging to *W* is matched with agent *Z* belonging to *M*, and so on. A specific member of a matching, such as (A, Z), will be termed a *match*. If an agent is not matched with anyone we say that the agent is *single*.

470 Matching

We are obviously using *W* and *M* as mnemonics for women and men. We do not recommend our matching procedure for real-world marriages. Our objective is to improve the matching procedure in cases such as the student-advisor relationship. This is problematic because, in general, there will be no obvious best match, as our first example reveals.

Example 1.1: Two agents of each type

The two *W* types are *A* and *B*, and the two *M* types are *Y* and *Z*. *A* prefers *Y* to *Z* and *B* prefers *Z* to *Y*. Therefore, it is possible to give each *W* her first choice. However, *Y* prefers *B* to *A* and *Z* prefers *A* to *B*. It is also possible to give each *M* his first choice. There are only two possible matchings, but neither gives *all* agents their first choice. On one hand, the matching $\pi = \{(A, Y), (B, Z)\}$ gives each *W* her first choice but each *M* gets his second choice. On the other hand, $\pi' = \{(A, Z), (B, Y)\}$ gives each *M* his first choice but each *M* gets her second choice.

When there are three (or more) agents of each type there will typically be matchings that are efficient but which are not stable in the sense that there will be an individual *W* and an individual *M* who would each prefer to be paired with the other rather than with the partner assigned by the matching in question.

DEFINITION: Stable matching

A matching π is stable if no agent prefers being single to his or her match at π , and if we cannot find two agents w and m of the opposite type such that w prefers a match to m to her situation at π , and m prefers a match to w to his situation at π . We say that w and m can *upset* the matching π if each does prefer the match (w, m) to his or her situation under π .

If π is stable and (w, m') is one of the matches specified by π , then neither w nor m' prefers being single to that match, and there is no other match (w', m) such that w prefers a match with m to the match with m' and m prefers a match with w to his match with w'.

All but two of our examples feature more than two members of each set, so we display the preferences in tabular form, with each column representing the preference ordering of the type that heads the column. The agent at the top will be most preferred, with the others listed in descending order of preference.

Example 1.2: An unstable matching

 $W = \{A, B, C\}$ and $M = \{X, Y, Z\}$ with the preferences as specified in Table 9.1. Agent *A* prefers *Y* to *X* and *X* to *Z* (and of course, *Y* to *Z*). The preferences of the others are interpreted similarly. The matching $\pi = \{(A, Y), (B, Z), (C, X)\}$ is not stable because *B* would rather have *X* than *Z*, and *X* would rather have *B* than *C*.

1. Students and Advisors

Table 9.1

A	В	С	X	Y	Z
Y	$egin{array}{c} X \ Z \ Y \end{array}$	X	B	B	A
X		Z	C	A	B
Z		Y	A	C	C

Every stable outcome is efficient, as we now prove: Suppose that $\pi = \{(w_1, m_1), (w_2, m_2), (w_3, m_3), \dots, (w_n, m_n)\}$ is a stable matching, but π' is another matching that w_i prefers to π . To establish that π is efficient, we need to show that some agent is worse off at π' than at π . Suppose that w_i is single at π . Then w_i must have a match at π' . Say w_i is matched with m_j at π' . If m_j is single at π and m_j prefers a match with w_i to being

single then π cannot be stable because then w_i and m_j both would prefer the match (w_i, m_j) to their situations under π . Because π *is* stable, if m_j is single at π then he prefers being single to a match with w_i and thus m_j is worse off at π' than at π . Suppose, then, that m_j is matched at π , to w_j . If m_j prefers w_i to w_j then π cannot be stable because then w_i and m_j each prefer π' to π . Because π is stable, it must be the case that m_j prefers w_j to w_i , and it follows that m_j is worse off under π' than under π . Finally, suppose that w_i is not single at π and is matched, say to m_i . Then w_i must be matched at π' , say to m_j . (We know that w_i prefers m_i to being single, because π is stable, and w_i prefers π' to π .) If m_j prefers being matched with w_i to his situation under π , then π wouldn't be stable. Therefore, stability of π implies that m_j is worse off at π' than at π . We have exhausted all the possibilities, so someone would be harmed by a move from π to π' , and hence π is efficient.

Every stable matching is efficient.

It is not the case that every efficient matching is stable, as we demonstrate with the next example.

Example 1.3: An efficient matching that is not stable

 $W = \{A, B, C\}$ and $M = \{X, Y, Z\}$ with the preferences as specified Table 9.2. Consider the matching $\pi = \{(A, X), (B, Y), (C, Z)\}$. It is not stable because both *A* and *Y* prefer the match (*A*, *Y*) to the match that they each have at π . However, it is efficient because if we begin with π and modify *A*'s match then *A* will be

Tabl	e 9.2
------	-------

A	В	С	X	Y	Z
Y X	Y	Ζ	А	A B	Ζ

worse off unless we replace (A, X) with (A, Y). But if we do that we make X worse off. If we modify B's match at π then we make B worse off, and if we modify C's match at π we make C worse off. Note that π remains efficient and unstable even if we add missing agents to any and all of the preference orderings, as long as the inserted agents are ranked below those already listed.

We have acknowledged the possibility that an agent may prefer to remain single rather than to be matched with a particular agent of the opposite type. This is captured by the column representing the agent's preference ordering by simply omitting the unacceptable agents. For the agent preferences as displayed in Table 9.2, *A* finds *Z* unacceptable, *B* finds both *X* and *Z* unacceptable, and so on.

DEFINITION: Acceptability

If agent α prefers being single to being paired with β we represent that fact by showing α 's preference ordering without listing β at all, and we say that β is unacceptable to α . We assume that no agent is indifferent between two *acceptable* agents of the opposite type.

It is easy to specify a matching algorithm that always generates a stable outcome. The deferred acceptance algorithm (DAA) is the one that we examine: We select one of the types, say W. Each W type proposes to the M type at the top of her preference ordering—the M that she most prefers, in other words. At the next stage each M provisionally accepts the W agent that he most prefers of all the acceptable agents that proposed to him and rejects all other proposals. At the third stage each rejected W type proposes to the next type M on her preference ordering. At that point there could be some Ms with two or more proposals. Each M type then rejects all proposals but the one from an acceptable W that ranks higher in his preference ordering than any other agent from which he has received a proposal. This means that he may reject an agent m at some stage then w cannot propose again to m at a later stage. The algorithm continues alternating proposing by Ws and accepting or rejecting by Ms until

- every *W* type either has a provisional acceptance or else has proposed to, and been rejected by, every *M* type that is acceptable to her, and
- every *M* has rejected all proposals from unacceptable *W*s and has rejected all proposals that rank lower in his preference ordering than some acceptable *W* from whom he has received a proposal.

At this point each *M* firmly accepts the one proposal that has not been rejected, if there is one, and remains single otherwise. This determines the matching. The algorithm is a *deferred* acceptance procedure because acceptances are not binding until the terminal round. (Of course, we could start from the beginning and have the *M* types doing the proposing and the *W*s accepting or rejecting.)

Example 1.4: Deferred acceptance with five agents of each type

Let $W = \{A, B, C, D, E\}$ and $M = \{S, T, X, Y, Z\}$ with the preferences as specified in Table 9.3. Note that each *W* is acceptable to some *M* and each *M* is acceptable to some *W*. Note also that *A* would only consent to a match with *Y*.

1. Students and Advisors

Table	9.3
-------	-----

A	В	С	D	Ε	S	Т	X	Y	Ζ
Y	Z T S Y X	S Z X T	X Y S	T S Z	D A C B E	B D C E A	E C A	A B C E D	D A C B

Every other *M* is unacceptable to *A*. For *D*, *E*, and *X* there are only three acceptable agents of the opposite type—agents who would be preferred to being single. Now, let's apply the DAA, *with the Ms doing the proposing*. After the first round of proposals, and rejections or provisional acceptances, the matches are

(A, Y), (B, T), and (D, S).

Note that both *S* and *Z* propose to *D* on the first round, and *D* responds by rejecting *Z* and accepting *S*. Note also that *X* has proposed to *E* but was rejected because *E* would prefer to be single. No *M* proposes to *C* at the first stage because no *M* has *C* at the top of his preference ordering. Because *X* and *Z* were rejected at the first stage, each now proposes again: *X* to his second choice, *C*, and *Z* to his second choice *A*. But *A* rejects *Z* because she would prefers *Y* to being single and being single to a match with *Z*. However, *C* accepts *X*, her only proposal, at this stage. The preliminary matches are now

$$(A, Y), (B, T), (C, X), \text{ and } (D, S).$$

Finally, *Z* proposes to his third choice, *C*, who now has proposals from both *X* and *Z*. Because *C* prefers *Z* to *X* she accepts *Z* and rejects *X*, so we now have

 $\pi^* = \{ (A, Y), (B, T), (C, Z), (D, S) \}.$

This is the outcome of the DAA. Both *E* and *X* will remain single. *X* will propose to *A* but will be rejected in favor of *Y*. *X* has now been rejected by all *W*s that are acceptable to him. It is clear that π^* is the terminal stage of the DAA.

It is easy to confirm that π^* of Example 1.4 is stable: *A* cannot find a preferred match because she prefers being single to being paired with anyone but *Y*. Similarly, *S*, *T*, and *Y* are each matched with the *W* that they most prefer, and so none of them has an incentive to deviate from π^* . Can we find a pair from $\{B, C, D, E, X, Z\}$ who would prefer each other to the partner assigned by π^* ? *B* is paired with *T*, and would prefer *Z* who prefers *C* to *B*. Similarly, *C* would only prefer *S* to the partner assigned by π^* , but *S* is assigned her first choice. Finally, can we find a pair from $\{D, E, X, Z\}$ who would prefer each other to the partner assigned by π^* ? But *S* is assigned her first choice. Finally, can we find a pair from $\{D, E, X, Z\}$ who would prefer each other to the partner assigned by π^* ? There are only four possibilities: (D, X), (D, Z), (E, X), and (E, Z). But *X* prefers being single to a match with *D*, and *D* prefers being single to a match

474 Matching

with *Z*. Of course *Z* prefers being single to a match with *E*, and *E* prefers being single to a match with *X*. Therefore, π^* is stable.

1.2 The main results

It is quite easy to prove that in *every* application of the DAA the outcome will be stable. Suppose that the Ws propose and the Ms accept or reject. Let π be the outcome of the DAA. Suppose π is *not* stable, and that w and m are not matched by π , but w prefers m to the match that she is assigned by π (or prefers m to being single if π leaves her single). Then w must have proposed to m at some stage in the operation of the DAA. Because (w, m) is not a match for π , then m must have rejected w, preferring some other agent w' to w. If (w', m) is a match determined by π then w and m cannot upset π . If (w', m) is not a match for π then m must have rejected w' at some stage, preferring some other agent w''. If (w'', m) is the match determined for m by π then we know that m prefers w'' to w' and w' to w. Therefore, *m* prefers w'' to *w*, and hence prefers his match at π to (*w*, *m*). Then w and m will not upset π . If (w'', m) is not the match determined for m by π then *m* prefers some other agent in *W* to w'', and so on. When we do finally arrive at the partner w^* determined for *m* by π then we will have a chain in which agent *m* prefers w^* to some other member of *W*, who is preferred to another member of W, and so on until we get to w'' who is preferred to w', who is preferred to w. But then w^* is preferred by m to w. Because (w^*, m) is a match for π , the pair consisting of w and m will not upset π . Therefore, π must be stable. Obviously, the same argument will establish that the outcome of the DAA is stable when the Ms do the proposing.

The outcome of the DAA is a stable matching.

The matching generated by the DAA will typically be different when the roles of proposer and acceptor are switched.

Example 1.5: The roles are switched

We apply the DAA to the agents and preferences of Example 1.1: The two *W* types are *A* and *B*, and the two *M* types are *Y* and *Z*. Agent *A* prefers *Y* to *Z* and *B* prefers *Z* to *Y*, and each prefers either *M* type to being single. Agent *Y* prefers *B* to *A*, and *Z* prefers *A* to *B*. Each *M* type prefers either *W* type to being single. If the *W*s propose, then the DAA yields $\pi^* = \{(A, Y), (B, Z)\}$ after one round. If the *M*s propose, then the DAA yields $\pi^{**} = \{(A, Z), (B, Y)\}$ after one round.

For Example 1.5 the DAA yields two different matchings, depending on which type does the proposing. If the *Ws* propose, then each *W* gets her most-preferred match, and if the *Ms* propose then each *M* gets his most-preferred match. We can generalize this: Suppose that the *Ms* do the proposing, and that each *W* is the first choice of some *M*, but no *W* ranks at the top of more than one *M's*

1. Students and Advisors

preference ordering. Then if each M is acceptable to each W, the DAA terminates at the first stage with each M being accepted by (and matched to) the W to whom he proposed. In that case, the DAA matches each M with his first choice.

It is not the case that for every marriage problem each proposer gets his or her most preferred match. For Example 1.4, with the Ms doing the proposing, only S, T, and Y are in that happy position. However, for every DAA outcome it is not possible to find another *stable* matching that makes any proposer better off. Here is the proof: For convenience, assume that the Ms do the proposing. Say that an arbitrary w in W is possible for a member m of M if there is a stable matching for which (w, m) is a match. We show that no member of M is rejected, at any stage in the operation of the DAA, by any W that is possible for him. That implies that each member *m* of *M* will be accepted by the member of *W* that ranks higher in *m*'s preference ordering than any other member of *W* that is possible for him—because an M type will propose in order of his preference. By definition, every match for *m* that belongs to some stable matching is possible for *m*. It follows that no member of *M* prefers any stable matching to the one that is selected by the DAA when the Ms do the proposing. It remains to prove that no member of M is ever rejected by a W that is possible for him if the DAA is employed and the Ms do the proposing.

Assume that we cut in on the operation of the DAA at some stage t such that no member of M has been rejected at any previous stage by a W type that is possible for him. (Note that a particular W that is possible for m must be acceptable to m) Suppose that at stage t, agent w rejects a proposal from m because w also has a proposal from some m' in M that w prefers to m. We show that w is not possible for *m*. (If *m* is unacceptable to *w* then *w* is clearly not possible for *m*.) We know that m' prefers w to any member of W that hasn't already rejected him. (And we also know that m' prefers a match with w to being single.) By assumption, m' has not been rejected in a previous round by any member of W that is possible for him. Therefore, m' prefers w to any W type that is possible for him. Therefore, no stable matching can contain (w, m) because if it did then m' would be matched with some w' that is possible for him (or else m' will be single), and we know that m' prefers w to w' (and prefers w to being single) and w prefers m' to m. Therefore, w is not possible for m. We have proved that any rejection received by a proposer comes from someone with whom the proposer is not matched in any stable outcome.

We can't claim that every proposer strictly prefers the outcome of the DAA to every other stable matching, because there is typically at least one proposer who has the same partner in more than one stable matching. But there is no stable matching at which some proposer gets a match that he or she strictly prefers to the one that results from the DAA.

Optimality theorem: If π^* is the outcome of the DAA, then there is no other stable matching that some proposer prefers to π^* .

Table 9.4

A	В	С	X	Y	Ζ
X	$Z \\ X \\ Y$	X	B	A	A
Z		Y	A	B	B
Y		Z	C	C	C

Now we turn to the question of incentive compatibility. Is it possible for one of the agents to profit by rejecting α and accepting α' when that agent actually prefers α to α' or proposing to β instead of β' when the agent prefers β' to β and β' has not yet rejected him or her? A simple way to address this question is to suppose that all agents submit their preference orderings to a referee who then uses the reported preferences to work out the

outcome of the DAA. Would submitting the true preference ordering be a dominant strategy?

DEFINITION: Strategy proofness

A matching algorithm is strategy proof if for each specification of the agent preferences truthful revelation is a dominant strategy for each agent.

Not only is the DAA not strategy proof, but there is *no* strategy-proof algorithm that always generates a stable matching, as we now prove.

Assume that there are three agents of each type, with $W = \{A, B, C\}$ and $M = \{X, Y, Z\}$. Consider the profile of agent preferences in Table 9.4. There are only two stable matchings in this case:

$$\pi = \{(A, Z), (B, X), (C, Y)\}$$
 and $\pi' = \{(A, X), (B, Z), (C, Y)\}$

We begin by confirming that π is stable. Note that both *X* and *Z* get their mostpreferred *W* type. Therefore, neither *X* nor *Z* can be a member of a *W*-*M* pair that upsets π . Agent *Y* prefers a match with *A* or *B* to the match (*C*, *Y*), but both *A* and *B* prefer π to a match with *Y*. Because each agent prefers a match with anyone of the opposite type to being single, π cannot be upset. Therefore, the matching π is stable. And π' is stable because *A* and *B* both get their most-preferred partners, and (*C*, *X*) is the only match preferred to π' by *C*, but *X* prefers π' to a match with *C*.

Next we show that π and π' are the *only* stable matchings for these preferences. Consider a different matching π'' . If π'' does not contain (*C*, *Y*) then it must contain either (*A*, *Y*) or (*B*, *Y*). (If *A* is single at π'' then π'' is not stable because *A* prefers *Y* to being single and *Y* prefers *A* to anything else. If *B* is single at π'' then π'' is not stable because *B* prefers *X* to being single and *X* prefers *B* to anything else.) But if π'' contains (*A*, *Y*) it is not stable because *A* prefers *Z* to *Y*, and *Z* prefers *A* to anything else. And if π'' contains (*B*, *Y*) it is not stable either because *B* prefers *X* to *Y*, and *X* prefers *B* to anything else. We have shown that π'' contains (*C*, *Y*), and thus there are only two ways to match the agents in {*A*, *B*} with the agents in {*X*, *Z*}, and one of these leads to π and the other leads to π' . (There can be no single agents in a matching that is stable for these preferences, because each agent prefers a match to anyone of the opposite type to being single.)

1. Students and Advisors

Note that *A* and *B* each prefer π' to π , and *X* and *Z* each prefer π to π' . Let **M** be any matching algorithm that generates a stable matching for each

Table 9.5

A	В	С	X	Y	Ζ
X	Z	X	B	A	A
Y	X	Y	A	B	B
Z	Y	Z	C	C	C

specification of agent preferences. Suppose that **M** results in π when applied to the preferences of Table 9.4. Now apply **M** to the preferences of Table 9.5. Note that Table 9.5 is the same as Table 9.4 except that we have changed the preferences of *A*. The only stable matching for Table 9.5 is π' , as we now show. Let π'' be a stable matching for Table 9.5. If π'' does not contain (*C*, *Y*) then it must contain either (*A*, *Y*) or (*B*, *Y*). But if π'' contains (*B*, *Y*) it is not stable because *B* prefers *X* to *Y*, and

X prefers *B* to anything else. Therefore, π'' contains (*A*, *Y*) if it doesn't contain (*C*, *Y*), and hence either

$$\pi'' = \{(A, Y), (B, X), (C, Z)\}$$
 or $\pi'' = \{(A, Y), (B, Z), (C, X)\}.$

The former is not stable for the preferences of Table 9.5 because *Z* prefers *B* to *C*, and *B* prefers *Z* to *X*. The latter is not stable because *X* prefers *A* to *C*, and *A* prefers *X* to *Y*. Therefore, π'' must contain (*C*, *Y*). Consequently, $\pi'' = \pi$ or $\pi'' = \pi'$. But π is not stable for Table 9.5 because *A* prefers *Y* to *Z*, and *Y* prefers *A* to *C*. It follows that π' is the only stable matching for Table 9.5. Therefore, **M** must lead to the selection of π' for the preferences of Table 9.5. Note that π matches *A* with *Z* and π' matches *A* with *X*, and *A* prefers *X* to *Z* according to *A*'s preference ranking in Table 9.4. Because the only difference between the two tables is the reported preference scheme of agent *A*, it follows that **M** is not strategy proof: When the true agent preferences are the ones given in Table 9.4, and π is selected, agent *A* can profit by misrepresenting her preference ordering. Specifically, if she were to report the preference scheme of an outcome that she prefers to the one that is generated by **M** under truthful revelation.

We know that π and π' are the only stable matchings for Table 9.4, and strategy proofness implies that M cannot select π . Therefore, if M is strategy

Table 9.6

A	В	С	X	Y	Z
X	Z	X	B	A	A
Z	X	Y	C	B	B
Y	Y	Z	A	C	C

proof it must generate matching π' for Table 9.4. Now, apply **M** to the preferences of Table 9.6. Table 9.6 is the same as Table 9.4 except with respect to the preferences of *X*. We show that π is the only stable matching for Table 9.6 by supposing that π'' is an arbitrary stable matching for Table 9.6. If π'' does not contain (*C*, *Y*) then it must contain either (*A*, *Y*) or (*B*, *Y*). But if π'' contains (*A*, *Y*) it is not stable because *A* prefers *Z* to *Y*, and *Z* prefers *A* to anything else. Therefore, π''

contains (B, Y) if it does not contain (C, Y), in which case either

$$\pi'' = \{(A, X), (B, Y), (C, Z)\}$$
 or $\pi'' = \{(A, Z), (B, Y), (C, X)\}.$

The former is not stable for Table 9.6 because *Z* prefers *B* to *C*, and *B* prefers *Z* to *Y*. The latter is not stable because *X* prefers *B* to *C*, and *B* prefers *X* to *Y*. Therefore, π'' must contain (*C*, *Y*), and hence $\pi'' = \pi$ or $\pi'' = \pi'$. The previous paragraph established that **M** must select π' at Table 9.4. But π' is not stable for Table 9.6 because *X* prefers *C* to *A*, and *C* prefers *X* to *Y*. Therefore, π is the only stable matching for Table 9.6, and thus **M** must select π for the preferences of Table 9.6. But π' matches *X* with *A* and π matches *X* with *B*, and *X* prefers *B* to *A* according to his preference ranking in Table 9.4. Tables 9.4 and 9.6 are identical except for the reported preference scheme of agent *X*. Therefore, **M** is not strategy proof because when the true agent preferences are the ones given in Table 9.4, agent *X* can profit by misrepresenting his preference ordering. Specifically, if he were to report the preference scheme of the forth column of Table 9.6 then he would precipitate the selection of an outcome that he prefers to the one that is selected under truthful revelation.

We have shown that **M** is not strategy proof. But **M** was an *arbitrary* procedure that always generated a stable matching for any specification of agent preferences. Therefore, we have shown that every strategy-proof procedure will precipitate an unstable outcome in some situations. Alternatively, if a procedure always delivers a stable matching under truthful revelation then truthful revelation cannot be a dominant strategy for every agent in every situation.

There is no strategy-proof allocation procedure for the marriage model that yields a stable outcome in every case.

The optimality theorem comes close to establishing that no *proposer* can profit by misrepresenting his or her preferences if a matching is determined by the DAA. There is no *stable* matching that a proposer prefers to the DAA outcome, so there is no deviation from truthful revelation by a proposer that leads to a stable matching that he or she prefers to the one that results from truthful revelation. However, when an agent contemplates a change in strategy, to see if it will precipitate a preferred outcome, we cannot expect the agent to restrict consideration only to strategies that lead to a stable matching. Nevertheless, it *is* the case that the DAA is immune to manipulation by proposers, whether the matching gives rise to a stable outcome or not.

If the DAA is employed then no proposer can benefit by deviating from truthful revelation.

If all agents were to submit their preference orderings to a referee, and the referee then applied the DAA to the stated preferences, resulting in the matching π^* , then submitting the true preference ordering is a dominant strategy for every proposer. We do not prove this because the optimality theorem reveals enough of the intuition behind the proof that the DAA is invulnerable to manipulation by any proposer.

We have only employed examples in which the number of *W* types equals the number of *M* types. All of our arguments can be extended to the general case by

1. Students and Advisors

adding "dummy" agents: If there are more *Ws* than *Ms* just add a sufficient number of *Ms* to make the two sets equal in size. Assign the degenerate preference scheme to these new *Ms*, so that they prefer being single to being matched with any *W*. Similarly, if the *Ms* outnumber the *Ws* we can add a sufficient number of dummy *Ws*. Why does this work? Because none of the original arguments depend on the absence of such agents. Also, no dummy agent will have a match at any stable outcome, so it's just as if the dummy agent didn't exist.

Sources

This section is based on Roth (1982). The seminal contribution is Gale and Shapley (1962).

Link

Roth and Sotomayor (1990) provides an extensives discussion of matching.

Problem set

- 1. Suppose there are three *Ws* and three *Ms*. Specify the preferences however you like, except that each agent finds everyone of the opposite type to be acceptable, and there is at least one inefficient outcome. Identify the efficient outcomes and the inefficient ones.
- **2.** Suppose that type *W* agent *A* prefers *X* to every other type *M* agent, and *X* prefers *A* to every other type *W* agent. Explain why every stable matching has to pair *A* with *X*.
- **3.** Prove that the unstable matching $\pi = \{(A, Y), (B, Z), (C, X)\}$ of Example 1.2 is efficient.
- **4.** Apply the DAA to Example 1.4, but with the *W*s doing the proposing.
- **5.** Assume that there are three *W*s and three *M*s. Specify the preferences of each agent so that the DAA generates the same match when the *M*s propose as when the *W*s propose.
- **6.** For the special case of two agents of each type, prove that truthful revelation is not a dominant strategy for each agent if the DAA is used to arrive at a matching. (*Hint*: Agents can declare that they would rather be single than matched with a particular agent of the opposite type, even when that is not true.)
- **7.** For each of the preferences of Tables 9.1, 9.2, and 9.3 determine the matching precipitated by the DAA when the *Ms* propose.
- **8.** In proving that there is no matching algorithm that always yields a stable outcome and for which truthful revelation is a dominant strategy for every agent for every specification of individual preferences, we employed a situation with exactly three agents of each type. Assume that every agent would prefer to be matched with anyone of the opposite type to being single, and extend the proof to the general case, with m > 3 type W agents and n > 3 type M agents.

2 COLLEGE ADMISSIONS

The problem of matching students to colleges is a generalized marriage problem. Each side has preferences, but colleges are matched with more than one student, although each student can be matched with at most one college. There are m colleges, named C_1, C_2, \ldots, C_m . College C_j has a capacity (or quota) q_j . It will admit fewer than q_j students if it receives fewer than q_j applications from students that meet its admissions criteria, but it cannot admit more than q_j . There are n student applicants, and we let S denote the set of all such individuals. A matching is an assignment of students to college such that no student is assigned to more than one college and no college has a student intake that exceeds its capacity. This section can be read independently of the others, but familiarity with the previous section is recommended.

DEFINITION: The college admissions problem

 C_1, C_2, \ldots, C_m is the set of colleges, and each college C_j has a capacity q_j . *S* is the set of students. Each college has a preference ordering for students, and each student has a preference ordering for colleges. *A matching* $\pi = \{(C_1, S_1), (C_2, S_2), \ldots, (C_m, S_m)\}$ assigns a set S_j of students to college C_j (for $j = 1, 2, \ldots, m$) such that no student is assigned to more than one college and S_j has no more than q_j members for each college C_j . The pair (C_j, S_j) itself is called a *match*.

If a college would not accept a particular student under any circumstances, then we say that the student in question is unacceptable to the college. Similarly, if a student would not be willing to attend a particular college, even if it meant not attending college at all, then that college is unacceptable to that student.

DEFINITION: Acceptability

If a college C_j would not admit student *s* even if its student population fell short of its capacity, then we say that *s* is unacceptable to C_j , and we incorporate that fact into the college's preference ordering by omitting student *s* from the ranking. Similarly, if a student *s* would prefer being unmatched to attending college C_j we say that C_j is unacceptable to *s*, and C_j will be omitted from the student's preference ranking.

We adapt the notion of stability to the college admissions problem in the obvious way: A matching π is stable if there is no student-college pair such that the student would prefer this college to the one to which the student was assigned by π , and the college would prefer to admit that student to one of the students assigned to it by π .

DEFINITION: Stable matching

A matching $\pi = \{(C_1, S_1), (C_2, S_2), \dots, (C_m, S_m)\}$ is stable if (i) every college is acceptable to every student assigned to that college, and every student is acceptable to the college to which the student is assigned, and (ii) we cannot find a college C_j and a student *s* such that *s* prefers college C_j to the college C_i to which *s* is assigned by π (or prefers C_j to being unmatched if *s* unmatched by π) and college C_j prefers student *s* to one of the students in S_j (or *s* is acceptable to C_j if S_j has fewer than q_j students). If *s* does prefer C_j to the college to which *s* is assigned by π and C_j prefers student *s* to one of the students in S_j we say that C_j and *s* can *upset* π . Similarly, we say that student *s* can upset π if *s* prefers being unmatched to the college to which *s* is assigned by π . College C_j can upset π if it is assigned an unacceptable student.

Note that if *s* prefers C_j to the college C_i to which *s* is assigned by π , then C_j is different from C_i and thus *s* does not belong to S_j .

Suppose that the preferences are such that there is a student *s* and a college C_j such that *s* prefers C_j to every other college, and C_j prefers at most $q_j - 1$ students to *s*. Then every stable matching π must assign *s* to C_j . Otherwise, *s* and C_j could upset π . That fact sometimes makes it easier to identify stable matchings.

Example 2.1: Finding the stable matchings

 $S = \{A, B, C\}$ is the set of students and the colleges are *X*, *Y*, and *Z*. Each college has room for exactly one student. (In other words, $q_X = q_Y = q_Z = 1$.) The preferences are specified Table 9.7. Every stable matching must include (*B*, *X*) because *X* is *B*'s top choice and *B* is *X*'s top choice. That leaves *A* and *C* to be assigned to colleges *Y* and *Z*, and there are only two ways to do that, given the unit capacity of each college. Therefore,

 $\pi^{S} = \{(A, Y), (B, X), (C, Z)\}, \text{ and } \pi^{C} = \{(A, Z), (B, X), (C, Y)\}$

are the only matchings that could be stable. And π^{S} is in fact stable, because *A* and *B* each get their most-preferred match, and although *C* prefers *X* to *Z*, college *X* does not prefer *C* to *B*. The matching π^{C} is also stable, because *X* and *Z* is each assigned its most prefered student, and although *Y* prefers *B* to *C*, student *B* prefers *X* to *Y*. Note that π^{C} favors the colleges and π^{S} favors the students, in a sense to be made precise in the next section.

2.1 Students apply to colleges

A simple extension of the deferred acceptance algorithm of the previous section allows us to define an algorithm that always yields a stable matching: The *student optimal deferred acceptance* (SODA) algorithm requires the students to do the

Table 9.7

A	В	С	X	Y	Z
Y	X	X	B	B	A
X	Z	Z	C	C	B
Z	Y	Y	A	A	C

"proposing," and of course the colleges respond by accepting or rejecting. In standard terminology, students apply for admission to the colleges. Acceptance is deferred in that it is not binding until the algorithm has terminated. (We will prove that SODA delivers a matching that no student prefers to any other stable matching, hence the term "student optimal.") At the outset, each student applies to the student's most-preferred college. Then each college C_i provisionally accepts the q_i students from its applicant pool that rank highest in that college's preference ordering. (Remember: Unacceptable students are not included in a college's preference ranking.) If fewer than q_i acceptable students have applied then C_i provisionally admits all of the applicants that it finds acceptable. Applicants that are not provisionally admitted are rejected. Once a student is rejected by a college the student cannot reapply to that college at a subsequent stage. At the next stage, all rejected students apply to the college that ranks next highest in their preference orderings-except in the case of a student who has been rejected by all colleges that the student finds acceptable, in which case he or she drops out of the running and will be unmatched. Each college will now have a new, augmented, set of applicants, and it provisionally accepts the q_i students that rank highest in its preference ordering. (A college's augmented set of applicants includes all students who have been provisionally accepted by that college in a previous round and who have not been subsequently rejected, along with all new applicants to the college. If fewer than q_i acceptable students have applied then C_i provisionally admits all of the applicants that it finds acceptable.) The process continues in this fashion until every student is on some college's list of provisionally accepted applicants or has been rejected by every college that is acceptable to her-that is, every college that the student finds preferable to being unmatched. At this point each student that is currently provisionally accepted by a college is firmly accepted by that college, resulting in a stable matching.

Now we prove that the outcome of the SODA algorithm is in fact stable for every specification of student and college preferences. First we establish that for arbitrary college C_j the number of provisional admissions by C_j cannot decrease over time; it can only increase. Let S_j^t be the set of provisional admissions by C_j at stage t. If S_j^t has exactly q_j members, then the set of provisional admissions will be different at the next stage only if some members of S_j^t are rejected by C_j and replaced with new applicants who are preferred by C_j . This substitution keeps the number of provisional admissions by C_j at q_j .

Let π be the matching generated by SODA. Suppose that student *s* prefers college C_i to the college to which *s* was assigned by π . If π does not fill C_i to

2. College Admissions

capacity then when C_j rejects *s* at some stage it must be because *s* is unacceptable to C_j . (By the previous paragraph, if C_j becomes full at some stage it remains full.) Then *s* and C_j cannot upset π in that case. Suppose then *s* is rejected by C_j at stage *t*, and S_j^t has exactly q_j members. The previous paragraph establishes that at each subsequent stage there will be exactly q_j provisional acceptances by college C_j . Suppose that $m \ge t$. Either $S_j^{m+1} = S_j^m$ (because none of the new applicants is preferred to any member of S_j^m or there are no new applicants), or else some new applicant *s''* is preferred by C_j to some student *s'* in S_j^m . Then C_j will provisionally accept the q_j highest-ranking members of A_j^m , where A_j^m consists of all the students in S_j^m plus all the new applicants that C_j finds acceptable. Because some new applicant *s''* is preferred by C_j to some member of S_j^m , the q_j highest-ranking members of A_j^m will not include the lowest-ranking member of S_j^m (lowest ranking in terms of the preferences of C_j). Therefore, C_j will prefer the lowest-ranked member of S_j^{m+1} to the lowest-ranked member of S_j^m .

We have established the following: If S_j^m contains exactly q_j members then either $S_j^{m+1} = S_j^m$ or else the lowest-ranked member of S_j^{m+1} is strictly preferred to the lowest-ranked member of S_j^m by college C_j . Now, student *s* was rejected by C_j at stage *t* when S_j^t had q_j members, so college C_j strictly prefers the lowestranked member of S_j^t to student *s*. Let S_j^t , S_j^{t+1} , ..., S_j^T be the subsequence of provisional acceptances by college C_j from stage *t* to the final stage *T*. For each such stage *m*, either $S_j^{m+1} = S_j^m$ or else college C_j strictly prefers the lowestranking member of S_j^{m+1} to the lowest-ranking member of S_j^m . We know that C_j prefers the lowest-ranking member of S_j^t to student *s*. It follows that the lowestranking member of S_j^T is strictly preferred to student *s* by college C_j . Because S_j^T is the list of actual (final) acceptances by C_j , student *s* and college C_j will not upset π , which must be stable.

The outcome of the SODA algorithm is a stable matching.

In the previous section we discovered that a stable matching emerges whichever type does the proposing. Similarly, we could adapt the DAA algorithm of the previous section so that students don't apply to colleges, but the colleges offer admission to the students and the students respond by accepting or rejecting. We'll refer to this as the *college optimal deferred acceptance* (CODA) algorithm. When a college is rejected by one or more students it has invited to enroll it then moves down its preference ranking, inviting new students to enroll, and so on. When the process terminates we will have a stable matching. The CODA algorithm is closer to the procedure used in the United States for recruiting student athletes.

In the case of the marriage model (Section 1) we saw that the DAA gives the proposers an advantage in the sense that no proposer ever strictly prefers another stable matching to the one that emerges from the DAA. For the college admissions problem there is no stable matching that any *student* prefers to the one that emerges from the SODA algorithm. The proof is fairly simple:

We say that college C_i is *possible* for student *s* if there is some stable matching (not necessarily the one precipitated by the SODA algorithm) at which *s* is assigned to C_i . (Note that if C_i is possible for *s* then *s* must be acceptable to C_i

and vice versa.) We show that if the SODA algorithm assigns *s* to college C_i (with the students "proposing") then *s* does not prefer any possible college to C_i . It follows that *s* does not prefer any other stable matching to the one precipitated by the SODA algorithm.

Suppose that up to stage t of SODA no student has been rejected by a college that is possible for that student. Suppose that s is rejected by C_i at stage t. If s is unacceptable to C_i then C_i is not possible for s. Suppose, then, that s is acceptable to C_i . Then C_i must be at capacity q_i when it rejects s, and any student s' in S_i^t , the set of students that have provisional admission from C_i at stage t of SODA, must be preferred to s by C_i . We show that C_i is not possible for s. Because no student has been rejected by a possible college prior to stage t, every s' in S_i^t prefers C_i to every other college that is possible for that student. Now, let π' be any matching. If C_i is not at capacity under π' then π' can't be stable because C_i prefers each of the q_i members of S_i^t to having an empty slot, and all of the members of S_i^t prefer C_i to any other college assigned to them by any stable matching (and to being unmatched). Therefore, if π' is stable then C_i is at capacity under π' . If s is assigned to C_i by π' there will be a member s' of S_i^t *not* assigned to C_i by π' , and s' prefers C_i to any other match determined for s'by any stable matching, and C_i prefers s' to s. Therefore, if π' is stable it cannot assign s to C_i , and hence C_i is not possible for s.

Optimality theorem: If π^* is the outcome of the SODA algorithm, then there is no other stable matching that some student prefers to π^* .

Example 2.2: The SODA outcome

We begin with the agents and the preferences of Table 9.7 of Example 2.1. The stable matchings are $\pi^{S} = \{(A, Y), (B, X), (C, Z)\}$ and $\pi^{C} = \{(A, Z), (B, X), (C, Y)\}$. It is easy to see that SODA results in π^{S} , which is stable, by the argument of Example 2.1: In the first stage *A* applies to college *Y*, and *B* and *C* both apply to *X*. Then *X* accepts *B* and rejects *C*, because *X* prefers *B* to *C*. At the next stage *C* applies to *Z*, her second-ranked college. *Z* accepts *C*, its only applicant. We now have *A* provisionally accepted by *Y*, *B* provisionally accepted by *X*, and *C* provisionally accepted by *Z*. The algorithm terminates with these becoming binding acceptances, resulting in π^{S} .

Because the marriage model is a special case of the college admissions problem, it follows from the discussion of strategy proofness in the previous section that there is no college admissions procedure that always yields a stable outcome and which cannot be manipulated by a student or college deviating from its true preference ordering. However, it's unsatisfactory to have the proof based on a system with only three students and three colleges. To generalize, begin with the three students *A*, *B*, and *C* and the three colleges *X*, *Y*, and *Z* employed in the proof in Section 1. Now add any number *m* of colleges C_1, C_2, \ldots, C_m with

2. College Admissions

respective capacities q_1, q_2, \ldots, q_m that can be as large as you like. We also add n students s_1, s_2, \ldots, s_n such that $n = q_1 + q_2 + \cdots + q_m$. Have each new student prefer each new college C_j to X and Y and to Z. Have each new college C_j prefer each new student s_i to A and to B and to C. Therefore, every stable matching will assign each new student to one of the new colleges, which are now full. It remains to determine how A, B, and C are to be assigned to X, Y, and Z. At this point we continue with the proof from Section 1.

For any college admissions procedure that always yields a stable outcome, there will be some configuration of preferences for students and colleges such that some student or college can benefit by deviating from its true preferences.

The optimality theorem demonstrates that no *student* can profit by misrepresenting his or her preferences *if* the outcome is determined by SODA if the deviation from truthful revelation by the student results in another stable matching. However, this is not the same as proving that no deviation from truthful revelation by a student can precipitate a matching, whether stable or not, to the one that emerges from truthful revelation. We merely state without proof the more general result.

If the SODA algorithm is employed then no student can benefit by deviating from truthful revelation.

If all students were to submit their preference orderings to a referee, and the referee then applied the SODA algorithm to the stated preferences and to the college preferences, then submitting the true preference ordering is a dominant strategy for every student.

2.2 Student placement

In the previous subsection we adapted the DAA (of Section 1) to the college admissions problem, for which one of the types can be matched with more

Many business schools (including those of Columbia, Berkeley, Northwestern, Michigan, and Yale) use bidding to assign students to courses. All students are given an endowment of 1000 points, which they allocate across courses. Because a student will bid low on a course that the student ranks high if he or she believes others will bid even lower, the bids cannot be used to infer student preferences. If students were to submit their preference orderings in addition to their bids, the SODA algorithm could be used to supplement the allocation process and improve the matching (Sönmez and Ünver, 2003).

than one member of the opposite type, by having the students "propose." That is, the students took the initiative by applying to colleges. Now we consider the *college optimal* deferred acceptance (CODA) algorithm in which the colleges do the "proposing." That is, a college takes the initiative by offering admission to the students that rank at the top of its preference ordering, until it has reached capacity. Each student then responds by rejecting every college except the one that the student most prefers out of all the schools that have offered admission. Some colleges will then have openings and they will then offer admission to the students, in order of college preference, from whom they have not previously received

rejections. Students with multiple admission offers will then choose the most preferred offer and reject the others, and so on.

We are now in a position to explain why we call SODA student optimal and refer to CODA as college optimal. No student prefers any stable matching to the outcome of SODA, and no college prefers any stable matching to the outcome of CODA, as the next example illustrates.

Example 2.3: The CODA outcome

We return to the set up of Table 9.7 of Example 2.1. The stable matchings are

$$\pi^{S} = \{(A, Y), (B, X), (C, Z)\}$$
 and $\pi^{C} = \{(A, Z), (B, X), (C, Y)\},\$

and we saw that SODA precipitates π^S . Now we show that CODA precipitates π^C . In the first stage, colleges *X* and *Y* both offer admission to *B*, while *Z* offers admission to *A*. *Z* is accepted by *A*, but *Y* is rejected by *B* in favor of *X*. At the next stage college *Y* offers admission to *C* and will not be rejected because *C* has no other offers. In that case, *A* is assigned to college *Z*, *B* is assigned to *X*, and *C* is assigned to *Y*, resulting in π^C . Student *B* is indifferent between π^C and π^S because they give *B* the same match. However, *A* and *C* each prefer π^S to π^C , while colleges *Y* and *Z* each prefer π^S to π^S . College *X* is indifferent between the two because its student intake is the same in either case.

We now investigate a special application of CODA, called the *student placement model*, for which the college preferences are based on test scores only. In their final year of high school students take a number of different tests. For instance, there may be only two kinds of tests—quantitative and verbal. But there may be more. Each college uses one of the tests to rank the students. For this model, no student is unacceptable to any college, but some students may prefer not to attend college rather than go to one of the schools from which the student is offered to admission. In the first round of the CODA algorithm each college C_j offers admission ("proposes") to the q_j students that rank highest in the test that it employs. Then each student accepts the college the student most prefers from the set of colleges from which the student has received an offer of admission. At the next stage, a college offers admission to the next r students in its test score ranking, where r is the number of students who rejected that college's admission offer at the previous stage. The algorithm terminates when

- every college *C_j* either has been accepted by *q_j* students or else has offered admission to every student, and
- every student has rejected all offers of admission from colleges that the student finds unacceptable and has rejected all offers of admission that rank lower than some acceptable college from which the student has received an offer of admission.

2. College Admissions

At this point each student is actually admitted to the college that the student has accepted.

DEFINITION: The student placement problem

There are *m* colleges, $C_1, C_2, ..., C_m$. Each college C_j has a capacity q_j . *S* is the set of students. Each student *s* writes *m* different tests, and we let $\sigma_j(s)$ denote the score of student *s* on test *j*. Then college C_j prefers student *s* to student *s'* if and only if $\sigma_j(s)$ is higher than $\sigma_j(s')$.

If there are hundreds (perhaps thousands) of colleges, we are not supposing that each student has to write hundreds of tests before applying to college. Several (perhaps a hundred or more) colleges can use the same test to rank the students. Some of the "tests" can even be weighted averages of several conventional tests.

Example 2.4: Four colleges and two tests

The four colleges are *E* (engineering school), *H* (humanities only), *L* (liberal arts), and *M* (premedical). The two tests are *quantitative* and *verbal*. Let q(s) and v(s) be the respective scores of student *s* on the quantitative and verbal tests. *A* reasonable supposition is that

 $\sigma_E(s) = q(s), \quad \sigma_H(s) = v(s), \quad \sigma_L(s) = 0.5q(s) + 0.5v(s), \text{ and}$ $\sigma_M(s) = 0.75q(s) + 0.25v(s).$

Turkey uses a generalized version of the SCA, in which a college's preference ordering is indeed the ranking of students according to the test score used by that college. This algorithm has several defects: It can render an inefficient outcome, and it is vulnerable to manipulation by students misrepresenting their preferences. It can also penalize students for improving their test scores. These deficiencies are overcome by the SODA algorithm with students applying to colleges and the colleges responding by accepting or rejecting (Balinski and Sönmez, 1999). If the colleges rank students according to the test scores, then the colleges really don't need to participate in the placement process. The state (or some other agency) can act as proxy for each college, assigning the q_j students with the top scores on test j to college C_j , and then have each student respond by rejecting the admissions offers from all but one college—the one that the student most prefers from the colleges to which he or she has been assigned, and so on. For reasons that become apparent later in this section, we call this placement version of the CODA algorithm the *serial choice algorithm* (SCA).

Does a student always have an incentive to get the highest possible test score when the SCA

is used? No! The following is an example in which a reduction in a student's test score results in the student being reassigned to a college that he or she prefers to the original assignment.

Example 2.5: A lower test score can mean a better college

There are two colleges, *X* and *Y*, and two students, *A* and *B*. Each college has room for only one student. Student *A* prefers *X* to *Y* and *B* prefers *Y* to *X*. There are two tests, quantitative and verbal. College *X* uses the quantitative score and *Y* uses the verbal score. The original test scores are displayed in Table 9.8. In the first round student *B* is offered admission to *X* (because *B*'s quantitative score is higher than *A*'s), and *A* is offered admission to *Y* (because *A*'s verbal score is higher than *B*'s). Because *A* does not receive an offer of admission from a preferred college, *A* accepts *Y*, and *B* accepts *X* because *B* does not have any other offer. That, then, is the final assignment.

Suppose, however, that *A* had done less well on the tests, resulting in the scores reported in Table 9.9. Note that *B*'s scores have not changed. With this second set of scores, student *B* is offered admission to both *X* and *Y*, because *B* has a higher score than *A* on both the quantitative and verbal test. Both colleges

Table 9	9.8
---------	-----

|--|

Student	Quantitative score	Verbal score	Student	Verbal score	Quantitative score
А	80	90	Α	90	70
В	90	80	В	80	90

have reached capacity, so there are no initial offers to *A*. Then *B* responds by choosing *B*'s preferred college, which is *Y*. Then *X* now has a vacancy, and it makes an offer of admission to the student with the next highest quantitative score, and that of course is *A*. Student *A* accepts *X*'s offer, and the algorithm terminates with *A* going to *X* and *B* going to *Y*. With the original test scores *A* was matched with *Y*. Clearly, student *A* prefers the match that results when *A* gets lower test scores.

Note that the SCA doesn't *wind up* assigning a student to two different colleges. A multiple assignment can occur only at an intermediate stage. Once it does occur, the student in question will reject all but the assignment that he or she prefers to all others.

We have serious reservations about a matching procedure that punishes students for improving their test scores. We are also concerned about *fairness*. We say that a student placement algorithm is unfair if some student prefers some college to the one to which the student was admitted *and* the student has a higher score on the test used by the preferred college than someone who was actually admitted to that college. It is also unfair if the student was not assigned to any college by the algorithm but has a higher test score for a college that is acceptable to her than someone admitted to that college.

DEFINITION: Fairness

A matching π is fair if there is no student *s* and no college C_j such that *s* prefers C_j to the college to which *s* is assigned by π (or prefers C_j to being unmatched if π does not assign *s* to any college) and *s* has a higher score on the test used by C_j than some student actually admitted to C_j . We say that the student placement algorithm itself is fair if for every specification of student preferences and test scores the resulting matching is fair.

Unfortunately, fairness and efficiency are sometimes incompatible, regardless of the algorithm employed. We demonstrate this by means of a simple example.

Example 2.6: Fairness and efficiency cannot be satisfied simultaneously

There are three students, *A*, *B*, and *C*, and two colleges, *X* and *Y*, each with room for only one student. (We know that one of the students will not be able to attend college.) Student *A* prefers *Y* to *X*, and the other two prefer *X* to *Y* as shown in Table 9.10.

College *X* uses a quantitative test score and *Y* uses a verbal test score. The student scores are displayed Table 9.11. We show that the only fair matching π has *A* enrolled in *X* and *C* enrolled in *Y*. The matching is not fair if *A* isn't enrolled in any college because *A* has the highest quantitative score, and thus will

Table 9.10

А	В	С
Y X	$X \\ Y$	X Y

have a higher quantitative score than whoever is enrolled in college *X*. Similarly, the matching can't be fair if *C* is not enrolled in any college because *C* has the highest verbal score. Therefore, a fair matching either assigns *A* to *X* and *C* to *Y*, or else *A* to *Y* and *C* to *X*. But the latter isn't fair because *B* has a higher quantitative score than *C*. The matching π that assigns *A* to *X* and *C* to *Y* is the only fair one in this case. However, π is not efficient because *A* prefers *Y* to *X* and *C* prefers *X* to *Y*. Therefore, if *A* and *C* switch assignments they will each be better off than at π . Student *B* is unaffected by the switch, so we will have made two individuals better off without harming the other person, and hence π is inefficient. If an algorithm precipitates the lone fair matching in this situation it will not deliver an efficient outcome, and if it delivers an efficient outcome it can't be fair.

The CODA algorithm selects the fair matching for Example 2.6: X will offer admission to A, the high scorer on the quantitative test, and Y will offer admission to C, the high scorer on the verbal test. Because each of these students has only one offer, each accepts and the algorithm terminates with A being assigned to X and C assigned to Y. This is also the outcome of the SODA algorithm, as we now show: In the first round A applies to Y and B and

Table 9.11

Student	Quantitative score	Verbal score
А	90	80
В	80	70
С	70	90

C each apply to *X*. College *X* will accept *B* and reject *C* because *B* has a higher quantitative score than *C*. Then *C* will apply to *Y* on the second round, and *Y* will now have applications from both *A* and *C*. Because *C* has a higher verbal score than *A*, student *A* will be rejected by *Y*, which will provisionally accept *C*. Student *A* will now apply to *X* and be accepted because *A* has a higher quantitative score than *B*, its other applicant. Because *B*

is rejected by *X*, *B* will apply to *Y* and be rejected because *C* has a higher verbal score than *B*. The algorithm now terminates because *B* has been rejected by both schools, and *A* and *C* have both been rejected by their top-ranked schools. SODA assigns *A* to *X* and *C* to *Y*, the only fair matching for this problem.

It's easy to see why SODA always delivers a fair outcome. If student *s* prefers college C_j to C_i then *s* will apply to C_j before C_i . If *s* has a higher score in the test used by C_j than student *s'* then C_j would never reject *s* before rejecting *s'*. In other words, if C_j rejects *s* then it will also reject *s'* before the SODA algorithm terminates. As we have seen (Examples 2.2 and 2.3 for instance), SODA and CODA typically will not precipitate the same outcome, but neither guarantees an efficient outcome, as Example 2.6 and the subsequent paragraph establish. However, in any situation the fair outcome π^s generated by SODA will weakly dominate every other fair matching π from an efficiency standpoint. That is, every student will either prefer the school assigned to him or her by π^s to the one assigned to him or her by π , or else the student will be indifferent because the student has the same match in both cases.

2.3 Single-test placement

An interesting special case of the student placement problem has the colleges all using the same test score. We call this the *single-test placement problem*. In that case there will be a single stable matching if we use the test scores to generate a college's ranking of students. *A* simple way of finding the stable matching is by means of the *serial choice algorithm*: The student with the highest test score announces his or her most-preferred college, and the student is permanently assigned to that college. The student with the next highest ranking is assigned to that student's most-preferred college, and so on until a college is full—that is, has reached its capacity. That college is dropped from the list of available schools. The student with the highest score, of all those students who have not yet been matched with a college, then chooses his or her most-preferred
2. College Admissions

college from the set of schools that have not yet reached capacity. The algorithm proceeds in this fashion, with the students choosing in order of test score and with a college being removed from the list of available schools as soon as it is full.

DEFINITION: The serial choice algorithm (SCA)

The *n* students are ranked $s_1, s_2, ..., s_n$ according to their scores on the single test σ . That is, $\sigma(s_1) > \sigma(s_2) > \cdots > \sigma(s_{n-1}) > \sigma(s_n)$. *A* is the set of all colleges, and A^t is the set of available colleges at stage *t*. At stage *t* student s^t choose the college that s^t most prefers from A^t . We have $A^1 = A$, with $A^t = A^{t-1}$ if no college in A^{t-1} reached its capacity as a result of the choice of s_{t-1} . If college C_j did reach its capacity q_j at stage t - 1 then we remove C_j from A^{t-1} to arrive at A^t . The algorithm terminates when each of the *n* students has chosen or each college has reached its capacity, whichever comes first. Each student who has had an opportunity to choose is then matched with the college that he or she selected.

Clearly, in this single-test model it never benefits students to lower their scores. That would only result in a lower ranking, and a later choice, perhaps from a smaller list of schools. Certainly, there would be no college added to the list available by the time the student had a chance to select. The order of choice by the higher ranking students will not have changed, so they will make the same selections, and hence the set of schools available to a student will either shrink or stay the same if the student's test score falls. Given that a lower score will give a student lower priority but will not otherwise affect the order in which people choose, a student cannot profit from a lower score. Therefore, the student's incentive is to get the highest score possible. Now, given the test scores, can there be any benefit to students from misrepresenting their preferences when the SCA is employed? No. In this case, misrepresenting one's preference ranking of colleges can't change the set of schools from which a student is able to choose, because it can't affect the order in which students choose. Therefore, misrepresentation of preference will either have no affect on a student's welfare, because the school that ranks highest in the available set still ranks highest according to the false preference ordering, or else it results in the student being assigned to a less desirable college-less desirable in terms of the student's true preferences. We have proved that the SCA is strategy proof. However, this proof is based on the assumption of a single test used by all colleges.

When applied to the student placement model with a single test score the serial choice algorithm is strategy proof, and it never punishes students for improving their test scores.

We conclude this section by proving that the student placement problem in which all colleges use the same test score has only one stable matching, and that it is the one generated by the SCA. We already know that the matchings precipitated by the either SODA or CODA algorithms are stable. It follows that the SODA and CODA algorithms yield the same outcome for this special model.

The matching π^* generated by the serial choice algorithm applied to the student placement model with a single test score is the only stable matching for that model. Also, π^* is generated by both the SODA and CODA algorithms.

We begin by proving that π^* is stable, and then we show that it is the only stable matching. Let s_1, s_2, \ldots, s_n be the ranking of students by test score, with the lower numbered students getting higher scores. Because the SCA gives s_1 his or her most-preferred college out of the set of all colleges, no student-college pair involving s_1 can upset π^* . Suppose that for the first t students s_1, s_2, \ldots, s_t there is no college that can join with one of these t students to upset π^* . Could student s_{t+1} join with some college to upset π^* ? No. None of the colleges that receive one of the first t students would prefer to have s_{t+1} instead of one of the students with a higher score. Of the colleges still available at the time s_{t+1} had an opportunity to choose, none would be preferred by s_{t+1} to the one assigned by π^* because the SCA allowed s_{t+1} to select the college that he or she most preferred from among those available. Therefore, no college-student pair involving s_{t+1} could upset π^* because either s_{t+1} would prefer the college assigned by π^* or the college would prefer each of the students assigned to it by π^* to s_{t+1} . We have proved that s_1 cannot be part of a student-college pair that upsets π^* , and that if none of the first *t* students can be part of a student-college pair that upsets π^* , then that also must hold for the first t + 1 students—that is, the t + 1 students with the highest scores. It follows that no student s_t , for any value of t, can upset π^* and hence π^* is stable.

Next we show that π^* is the only stable matching. We assume that π is an arbitrary stable matching and prove that $\pi = \pi^*$: Obviously, stability implies that s_1 gets his or her most-preferred college—call it c^* . That follows from the fact that s_1 has the highest score and hence is at the top of the preference ranking of every college. (If π does not assign c^* to s_1 then the pair consisting of s_1 and c^* can upset π .) Now suppose that we can prove that any stable matching π assigns student s_m to his or her most preferred college in A^m , for m = 1, 2, ..., t. (The set A^m of colleges available to s_m is the set determined by the SCA, as defined previously.) Then stability of π implies that s_{t+1} can only be assigned to a college in A^{t+1} . (If s_{t+1} is assigned to C_i not in A^{t+1} then some higher ranking student s_i must be assigned to C_i by π^* but not by π , in which case the student-college pair consisting of s_i and C_i can upset π , contrary to the supposition that π is stable.) Because stability of π implies that π can only send student s_{t+1} to a college in A^{t+1} (as we have just shown) and π^* sends s_{t+1} to his or her most preferred college in A^{t+1} , any other matching consistent with all of the top t students getting their most-preferred colleges from their available sets (as determined by the SCA)

2. College Admissions

would assign s_{t+1} to a college that was less desirable to him or her than the college C_j that s_{t+1} is admitted to under π^* and would leave C_j with a student with a lower test score than s_{t+1} , in which case s_{t+1} and C_j would upset π . Therefore, π must also assign s_{t+1} to the college in A^{t+1} that s_{t+1} most prefers.

Here's what we have so far: If π is stable then it must assign s_1 to s_1 's mostpreferred college, and if for any *t* the matching π assigns s_m to s_m 's most-preferred college in the set A^m for m = 1, 2, ..., t (where A^m is specified in the definition of the SCA) then if π is stable it must assign s_{t+1} to his or her most-preferred college in A^{t+1} . It follows that a stable matching must assign students to colleges in precisely the way that the SCA does. That is $\pi = \pi^*$.

We have proved that π^* , the matching generated by the SCA, is the unique stable matching for the student placement model with a single test. Because both SODA and CODA generate stable matchings, we then conclude that SODA, CODA, and SCA precipitate the same outcome. However, it is a lot easier to prove directly that the SODA and CODA algorithms result in π^* than to prove stability in general, so we conclude this section by proving the SODA and CODA algorithms result in π^* for the student placement model with a single test score.

First, we prove that SODA generates π^* . Clearly s_1 will wind up with s_1 's mostpreferred college because he or she will apply to that college in stage 1 and will be accepted. However many other students apply to that college in subsequent rounds, the acceptance of s_1 will never be withdrawn because every college ranks s_1 first in its "preference" ordering. Now we prove that if SODA assigns each of the students s_1, s_2, \ldots, s_t to the same college as π^* does then SODA will also assign student s_{t+1} to the same college that π^* does. It is clear that the colleges in A^{t+1} (which is defined by the SCA), and only those colleges, will have unused capacity. Moreover, any college that is full after stage t of the SCA will, under the SODA algorithm, eventually have enough applications from students with higher test scores than s_{t+1} , and so will reject s_{t+1} if and when he or she applies. Therefore, the only colleges that would accept s_{t+1} and never reject that student are the colleges in A^{t+1} . Under the SODA algorithm student s_{t+1} will eventually apply to s_{t+1} 's most-preferred college in A^{t+1} because he or she will eventually be rejected if he or she applies to any other college. This student will be accepted by his or her most-preferred college in A^{t+1} , and never be rejected, because any student with a higher test score than s_{t+1} will be assigned to the same college by SODA as by π^* , and hence the college in A^{t+1} that is most preferred by s_{t+1} will have room for s_{t+1} even after s_1, s_2, \ldots, s_t have each been assigned to the same college as π^* . Therefore, SODA assigns s_{t+1} to the same college as the SCA. Consequently, we can show that for every t, student s_t will be assigned the same college by the SODA algorithm as by the SCA. Therefore, the two algorithms generate the same matching.

Finally, we prove that the CODA algorithm generates π^* . At stage 1 of CODA each college C_j offers admission to the q_j students with the highest test scores. That means that s_1 is offered admission by every college, and s_1 will accept his or her most preferred school. Clearly, the CODA algorithm will

In Boston, Cambridge, Denver, Minneapolis, Seattle, St. Petersburg-Tampa, and other U.S. metropolitan areas, students are matched with public schools by a variant of the "Boston mechanism." For each school, public policies (instead of test scores) are used to rank the students. In round 1 each school C_i accepts the top q_i students (in C_i 's ranking) who listed C_i as first choice, where q_i is school capacity. In round 2, all students not yet assigned are admitted by their secondranked school if that school has a place available after accepting higher ranked students who listed that school first or second. And so on. This mechanism is unfair, in the sense that it can assign student s to school C_i even though s prefers C_i and C_i has admitted a student who is below s in the ranking of students for C_i . The mechanism can also lead to an inefficient outcome and gives students a strong incentive to misrepresent their preferences (Abdulkadiroğlu and Sönmez, 2003.) Proponents of the mechanism state that about 70% of students are typically assigned their first choice. However, preferences are often misrepresented, and only about 30% of the students are placed in the school that ranks at the top of their *true* preference ordering (Chen and Sönmez, 2003).

assign s_1 to the same college as π^* . Also, s_2 is offered admission by every college (except those with room for only one student), so s_2 will accept his or her most-preferred school in A^2 , where as usual A^t is specified in the definition of the SCA. The school C_i that s_2 accepts will be s_2 's ultimate match because s2 will never receive an offer from a school that he or she prefers: Either s_2 gets the school that ranks at the top of his or her preference ordering, or the school at the top of s_2 's ordering is the one that is most preferred by s_1 and that school has room for only one student. Suppose that we can prove that CODA assigns each of the students s_1, s_2, \ldots, s_t to the same college as π^* . Then as we showed in the previous paragraph the only schools that will not be full after we place all of the t students with the highest scores in the college to which they are assigned by π^* are the ones in A^{t+1} . These colleges, and only these colleges, will offer admission to s_{t+1} . (The other colleges can fill their positions with higher ranking students who will never withdraw their acceptance.) Then s_{t+1} will accept the offer from the college in A^{t+1} that s_{t+1} prefers to all the other colleges in A^{t+1} , and that offer will never be withdrawn because any student with a higher score will already have been offered admission by the college accepted by s_{t+1} or by a college that he or she prefers to the one accepted by s_{t+1} . It is clear that we can proceed from t = 1 to t = 2 to t = 3 and so forth until we have

established that each student is assigned the same college by the CODA algorithm as by the SCA.

Sources

This section is based on Gale and Shapley (1962) and Balinski and Sönmez (1999). Examples 2.5 and 2.6 are from Balinski and Sönmez (1999), who prove, among other things, that the fair outcome generated by SODA will weakly dominate every other fair matching from an efficiency standpoint.

Links

See Ergin and Sönmez (2004) for more on the "Boston mechanism." See Roth and Sotomayor (1990) and Roth and Xing (1994) for additional discussions of matching in general.

2. College Admissions

Table	9.12
-------	------

A	В	С	D	Ε	F
X	X	Z	Z	Y	X
Y	Y	Y	Y	X	Z
Z	Z	X	X	Z	Y

Problem set

1. There are six students, *A*, *B*, *C*, *D*, *E*, and *F*, and three colleges, *X*, *Y*, and *Z*, each with room for two students. The student preferences are given in Table 9.12 and the college preferences are given in Table 9.13. Work out the matching generated by SODA. Demonstrate that it is stable.

Table 9.13

X	Y	Ζ
В	Α	Α
Α	F	В
С	С	С
D	E	F
Ε	D	D
F	В	Ε

- **2.** Work out the matching generated by CODA for the data of question 1. Demonstrate that it is stable.
- **3.** There are six students, *A*, *B*, *C*, *D*, *E*, and *F*, and three colleges, *X*, *Y*, and *Z*, each with room for two students. The test scores are displayed in Table 9.14 and the student preferences are given in Table 9.15. Work out the matching determined by the CODA if college *X* were to use the quantitative score and *Y* and *Z* each used the verbal test score.

Tab	le	9.	14	

Student	Quantitative score	Verbal score
А	90	80
В	85	90
С	80	70
D	75	60
Ε	70	75
F	65	85

Table 9.15

A	В	С	D	Ε	F
X	X	Z	Z	Y	X
Y	Y	Y	Y	X	Z
Z	Z	X	X	Z	Y

4. Assume the data of question 3, but this time assume that colleges *X* and *Y* use the quantitative test score, and *Z* uses the verbal score. Work out the matching determined by the CODA.

3 HOSPITALS AND DOCTORS

An obviously important matching problem is the pairing of physicians graduating from medical school with hospitals employing them as residents. This is a variant of the college admissions problem because each doctor will wind up at only one hospital, but most hospitals will employ more than one resident. The medical labor market in the United States has used an effective matching algorithm for half a century, although it has been patched from time to time in response to pressure of various kinds. The algorithm requires each hospital to rank the doctors in order of preference, and each doctor to rank the hospitals in order of preference.

The modern U.S. internship program was instituted in the early 1900s when the hospital demand for graduating doctors was significantly larger than the supply. Until 1952 there was no algorithm and no central clearing house. Hospitals negotiated directly with medical students, offering internships to some and being either accepted or rejected by each student receiving an offer. Competition for newly minted physicians resulted in hospitals approaching medical students earlier and earlier in their programs. By 1944 appointments were being made two years before graduation. This meant that matches between hospitals and doctors were based on far less information than if the hospital waited until late in the graduating year to approach a student. But no individual hospital had an incentive to wait until then because other hospitals would have already contracted with the best doctors. Because of the strong incentive to beat the competition to a student's door, efforts by the medical community to halt the process did not avail. Adding to the overall dissatisfaction with this decentralized market process was the fact that it was not uncommon for doctors to renounce a contract when a better opportunity presented itself at a later date.

The notion of a centralized matching procedure became increasingly attractive, and in the academic year 1951–2 the algorithm made its first appearance. Hospitals submitted their preference orderings of graduating medical students to a centralized clearing house, and the students submitted their rankings of hospitals. One important criterion was *stability*: When the final matching is determined there should be no hospital and no doctor who would each prefer being matched with the other to the assignment determined by the algorithm.

3. Hospitals and Doctors

Tab	le	9.	1	6
-----	----	----	---	---

D	E	F	Н	Ι	J
H	I	H	E	D	F
I	J	J	F	E	E
J	H	I	D	F	D

DEFINITION: Stable matching

Matching π is stable if we cannot find two matches (H, D) and (H', D') specified by π such that H prefers D' to D and D' prefers H to H'. We say that H and D' upset the matching π if both do prefer the match (H, D') to their situations under π .

A complete ranking by either hospital or doctor would be prohibitively costly, so the algorithm works with partial rankings. However, we illustrate with a small number of doctors and hospitals, so we employ complete rankings. We begin by illustrating the trial algorithm that was proposed by the *National Intern Matching Program* (NIMP).

Example 3.1: Three hospitals and three doctors

The three doctors are D, E, and F, and the three hospitals are H, I, and J. We assume that each hospital wishes to hire one and only one resident. Table 9.16 gives the preference rankings of each. Doctor D prefers H to I and I to J (and of course, H to J). The columns for E and F are interpreted similarly. Hospital H ranks doctor E first, then F, and then D. And so on. The trial algorithm begins by searching for a hospital-resident pair such that each gives the other first-place rank. There are no such pairs for this table, so we proceed to the second stage in which we search for a hospital that ranks a doctor second when that same doctor ranks that hospital first. There are two such matches: E and I, and F and H. Then doctor E is assigned to hospital I, and F is assigned to H. By default, J and D are paired.

Note that hospital *J* is *D*'s last choice. *D* would have fared better to have (untruthfully) ranked *I* in first place, as we now show.

Example 3.2: Preference misrepresentation by a doctor

Assume that the true preferences are as specified in Table 9.16 of Example 3.1, but the reported preferences are those of Table 9.17. This is the same as Table 9.16 except for the fact that D has switched the ordering of H and I. Because the trial algorithm begins by searching for a hospital-resident pair such that each

Table 9.17

D	Ε	F	Н	Ι	J
I H	I I	H I	E	D E	F
J	у Н	ј I	D	F	D

gives the other first-place rank, we now have *D* matched with *I*, and *D* prefers *I* to hospital *J* to which *D* is assigned when *D* reports his or her true preference ranking.

Because the trial algorithm sometimes gives physicians incentive to misrepresent their first choice of hospital, it was revised in response to student objections. The rule that was actually employed in the first year of centralization begins by searching for a hospital-resident pair for which each ranks the other first, as before. If there are no such pairs, the algorithm searches for a match between a *hospital's* first choice and a *doctor's* second choice.

Example 3.3: Example 3.1 with the revised algorithm

The true preferences are the ones given in Table 9.16 of Example 3.1. There are no hospital-resident pairs such that each ranks the other first. But hospital I ranks D first and D ranks I second, so D and I are paired. Similarly, hospital J ranks F first, and F ranks J second. Therefore F is assigned to J, leaving E and H to be paired by default. The algorithm assigns E to H, D to I, and F to J. None of the doctors has an incentive to misrepresent his or her first choice: Doctor D prefers H to I, but already ranks H in first place and winds up at hospital I nonetheless. Similarly, F ranks J in second place and is not assigned to H, F's top-ranked hospital. How about doctor E? E does get H, but it is E's bottomranked alternative. If E were to rank H at the top then E would still be assigned to H because then H and E would be the only hospital-doctor pair for which each ranked the other first. If E ranked J at the top then there would still be no hospital-doctor pair for which each ranked the other first. And the next step would still be to assign D to I and F to J because hospital I ranks D first and D ranks I second, and hospital J ranks F first, and F ranks J second. This leaves E and H to be paired by default, as with truthful revelation. Note also that there is no hospital-resident pair that could upset the assignment by striking a mutually advantageous contract, with each preferring its new partner to the one assigned by the revised algorithm. In other words, the revised algorithm precipitates a stable outcome. In this case, stability is confirmed simply by observing that each hospital is assigned the physician that it most prefers, so no hospital wants to be rematched-with anyone.

4. Allocating Dormitory Rooms

The algorithm was modified from time to time to repair defects that came to light. In the past thirty years, particularly, the increasing frequency of doctors married to doctors has necessitated adjustments in the matching algorithm. We do not specify any of these adjustments, but it would not be too far from the truth to say that the NIMP used a variant of the CODA algorithm from 1951 to 1998. Of course, the hospitals played the role of the colleges. (See Section 2 for a definition and discussion of CODA.)

In the mid-1990s there was another crisis of confidence in the matching process. One of the factors related to changes in the financing of health care. Specifically, students felt that the NIMP algorithm favored the hospitals to an unreasonable degree. A new algorithm was commissioned in 1995 and used for the first time in 1998. It would be fair to say that the new algorithm is closer to the SODA procedure (of Section 2) than to CODA.

Source

This section is based on Roth (1984).

Links

See Roth (1990) for a discussion of the market for medical school graduates in the United Kingdom. Roth and Peranson (1999) discuss recent changes in the matching algorithm from several important perspectives.

Problem set

- 1. Determine all the stable matchings for the preferences of Example 3.1.
- **2.** Assume that there are three doctors and three hospitals, with each hospital intent on hiring exactly one doctor. Use the CODA algorithm of Section 2 to represent the revised hospital-doctor matching. Show that it precipitates a stable matching for every specification of hospital and doctor preferences.
- **3.** Assume that there are three doctors and three hospitals, with each hospital intent on hiring exactly one doctor. Use the CODA algorithm of Section 2 to represent the revised hospital-doctor matching. Prove that physicians never have incentive to rank their true most-preferred hospital lower than first.

4 ALLOCATING DORMITORY ROOMS

Assignment problems, in which only one type has preferences, are often called *one-sided matching* problems. Obviously, the dormitory allocation problem qualifies. There are *n* students and *n* dormitory rooms, and each student has a preference ordering over the set of rooms. To simplify the discussion, we assume that no student is indifferent between two different rooms.

The allocation problem is typically applied to a situation in which some of the students already have rooms. If the student has to choose at some point between retaining his or her present room for the next academic year or participating in the new allocation of rooms and possibly winding up with one that ranks lower

than the one given up then the student has to make a decision in the presence of uncertainty. Therefore, individual preferences will be represented initially in terms of utility functions, so that we can compute and compare the expected utility (EU) of different strategies.

There are *n* students and *n* rooms. We let *S* denote the set of students and *R* denote the set of rooms. Each student *s* in *S* has a utility function U_s defined on the set of rooms. Of course, for any two rooms *r* and *r'* in *R* we have $U_s(r) > U_s(r')$ if and only if *s* prefers *r* to *r'*. Some of the rooms already have occupants, and we let π_0 denote the set of matches (s, r) such that room *r* is already occupied by student *s*. We can refer to π_0 as the status quo. Typically, π_0 is not complete, in the sense that there will be many *s* in *S* such that (r, s) does not belong to π_0 for any *r* in *R*. In that case, we think of *s* as a new student—new to the dormitory allocation system, that is.

DEFINITION: The room assignment problem

S is the set of *n* students and *R* is the set of *n* rooms. Each student *s* has a utility function U_s defined over the rooms. Student *s* prefers strategy σ to strategy σ' if and only if the expected utility from σ exceeds the expected utility from σ' . The (partial) matching π_0 , assigning some students to rooms, is the status quo. If (s, r) belongs to π_0 it means that *r* is already occupied by *s*.

If (s, r) belong to π_0 then the room allocation scheme that we employ may or may not give *s* the right to keep *r*, and if *s* does relinquish room *r* the allocation scheme may or may not guarantee that *s* will wind up with a room that is at least as good as *r*. That is, the room allocation scheme may or may not satisfy the *participation constraint*.

DEFINITION: Participation constraint

The room allocation scheme satisfies the participation constraint if for every (s, r) in π_0 the expected utility realized by *s* when *s* participates in the scheme is at least as high as $U_s(r)$.

4.1 A commonly used scheme

The first allocation scheme that we consider gives any student who already occupies a room the right to hold on to that room. That means that the students can choose not to participate in the allocation scheme, in which case they are assigned their current rooms again. In that case we say that the allocation scheme has no guarantees. A commonly used mechanism is *serial choice with no guarantee*: Before this allocation scheme begins operating, all students who already occupy rooms have to declare whether they want to participate. If they do not participate then they are assigned the rooms they already occupy. The students who do not presently occupy rooms or who have given up the rooms they occupy are given numbers. The lowest-numbered student chooses a room

4. Allocating Dormitory Rooms

Table 9	.1	8
---------	----	---

Room	U_A	U_B	U _C
X	3	4	3
Y	4	3	4
Ζ	1	1	1

from the set of available rooms, and the one that is chosen is removed from the list of available rooms. The second student chooses from the set of currently available rooms, and that choice is removed from the list. Then the third student makes a choice, and so on.

DEFINITION: Serial choice with no guarantee (SCNG)

If (s, r) belongs to π_0 then *s* must decide whether to keep room *r* or participate in the allocation scheme. Let R_1 denote the set of rooms that are either not currently occupied or are presently occupied by a student who has agreed to give up his or her room and participate in the scheme. S_1 is the set of students who do not have a room or who have given up the room that they had. The members of S_1 are numbered, s_1, s_2, \ldots, s_m . Student s_1 has highest priority, and s_2 is next, and so on. Student s_1 chooses his or her most-preferred room from R_1 , and that room is removed from the list, resulting in the available set R_2 . Then s_2 chooses from R_2 , and the room selected is withdrawn, resulting in the available set R_3 . At stage *t* student s_t chooses from R_t , where R_t is obtained from R_{t-1} by removing the room chosen by s_{t-1} .

The numbering of students can be determined entirely by chance, with each having the same probability of being selected to go first. This random allocation can be modified by giving seniors and/or those with a high GPA a higher probability of receiving a low number. In any case an assignment of priority numbers can result in an inefficient assignment if a student has the right to hang onto a room that the student already has but is not guaranteed an equally good room if he or she gives it up, as our first example demonstrates.

Example 4.1: Inefficiency of the SCNG with three students and three rooms

We allocate rooms using the SCNG algorithm. The three students are *A*, *B*, and *C*. The rooms are *X*, *Y*, and *Z*. Student *A* currently occupies *X*, and the other two rooms are unoccupied. The utility functions are given by Table 9.18. We have $U_A(X) = 3$, $U_A(Y) = 4$, $U_A(Z) = 3$, and so forth. *A* has to decide whether to hold onto *X* or participate in the allocation scheme. We assume that the priority ordering is determined by a draw from a uniform probability distribution. In plain words, three pieces of paper are placed in a hat, each one bearing a name

of a student. A slip of paper is removed at random, and the student whose name is on the slip is first, and the second student is the one whose name is drawn next. There are six different possible orderings, namely *AB*, *AC*, *BA*, *BC*, *CA*, and *CB*. (*AB* means that *A* is first and *B* is second.) Each ordering has an equal chance of being the one determining priority, so the probability of each ordering is 1/6. If *A* holds on to room *X* then *A*'s utility will be 3, because $U_A(X) = 3$. If *A* participates in the allocation scheme, then *A* will choose room *Y*, with $U_A(Y) = 4$ if the ordering of students is *AB* or *AC*, because *A* will go first and choose *Y*. With *BA* or *BC* student *B* goes first and will take *X*. With *BA* student *A* gets the next choice and will choose *Y*. With *BC*, student *B* chooses *X* and *C* chooses *Y*, so *A* is left with *Z*. With *CA* or *CB* student *C* will choose *Y* first, and *A* will get *X* with *CA* and will get *Z* with *CB*. Student *A* will get *Y* in three of the cases, will get *Z* in two of the cases, and will get *X* in only one case. Therefore, *A*'s expected utility is

$$\frac{3}{6} \times U_A(Y) + \frac{2}{6} \times U_A(Z) + \frac{1}{6} \times U_A(X) = \frac{3}{6} \times 4 + \frac{2}{6} \times 1 + \frac{1}{6} \times 3 = \frac{17}{6}.$$

The expected utility from participating in the lottery is less than the utility of holding onto *X*, so *A* will retain *X*. Therefore, only *B* and *C* participate, and the available rooms are *Y* and *Z*. With probability $\frac{1}{2}$ *B* will go first, in which case *B* chooses *Y*. With probability $\frac{1}{2}$ *C* will go first and will get *Y*. With probability $\frac{1}{2}$ the room assignment will be π under which *A* will get *X*, *B* will get *Y*, and *C* will get *Z*, resulting in the utility profile of the first row of Table 9.19. But if assignment π' gives *Y* to *A*, *X* to *B*, and *Z* to *C* then the utility profile would be that of the last row of Table 9.19. Clearly, assignment π is not efficient because π' gives *A* and *B* more utility and *C* the same utility.

Versions of serial choice with random determination of the order of choice are employed at many U.S. universities, including Michigan, Princeton, Rochester, and Stanford (in their graduate residences) and Carnegie-Mellon, Duke, Harvard, Northwestern, University of Pennsylvania, William and Mary, and Yale (in their undergraduate dormitories). Many different randomization procedures are employed, with variations even across undergraduate colleges at Yale. At Duke, Harvard, Northwestern, Pennsylvania, and William and Mary incumbent students are permitted to opt out of the allocation process and retain their current rooms (Abdulkadiroğlu and Sönmez, 1999).

If there are no existing tenants, as would be the case with a freshman dorm, then the SCNG procedure always leads to an efficient outcome. It's easy to see why. Student s_1 gets his or her most-preferred room, so any change in s_1 's assignment will reduce his or her utility. Given that we can't change s_1 's assignment and that s₂ gets s₂'s first choice of all the rooms available after s_1 makes a selection, we can't change the assignment of rooms to s_1 and s_2 without making one of those students worse off. In fact, suppose that π is the assignment that results from the SCNG and that π' is some other assignment. Let t be the smallest integer such that the room assigned to s_t by π' is different from the room assigned to s_t by π . Because the rooms allocated to students $s_1, s_2, \ldots, s_{t-1}$ are the same for π' as for π , the set of rooms R_t still available for s_t is the same for π' as for π .

4. Allocating Dormitory Rooms

Assignment	A's utility	B's utility	C's utility
$\pi \pi'$	3	3	1
	4	4	1

Table 9.19

But π gives s_t his or her most-preferred room from R_t and π' gives s_t a different room. Therefore, π' leaves s_t worse off than under π . Therefore, we can't change π without making someone worse off, and hence π is efficient.

If none of the rooms is already occupied then the outcome of the SCNG is efficient.

It is obvious that no student can profit by misrepresenting his or her preferences if the SCNG is employed and there are no existing tenants—that is, none of the rooms is already occupied. We formally test for strategy proofness by supposing that all students submit their utility functions to a referee who then uses these functions to work out the outcome of the SCNG. Would submitting the true utility function be a dominant strategy?

DEFINITION: Strategy proofness

An algorithm for determining an assignment of rooms is strategy proof if for each specification of the true student utility functions, reporting truthfully is a dominant strategy for each student.

Student s_1 gets his or her most-preferred room under truthful revelation, so s_1 has no incentive to misrepresent his or her preferences. In general, student s_t gets whichever room in R_t maximizes the utility function that s_t reports, and *the utility function that* s_t reports *has no effect on the priority ranking of students*. Therefore, the dominant strategy for s_t is to report his or her true utility function and get the room in R_t that s_t actually likes best. We have proved that truthful revelation is a dominant strategy for each student if there are no existing tenants. (Note that it suffices for all students to report their preference orderings of rooms. A utility representation is not necessary. But the students themselves make expected utility calculations to see if they should participate if they currently occupy a room.)

If none of the rooms is already occupied then the SCNG is strategy proof.

4.2 An efficient procedure

The possibility of winding up with an inferior room if you give up the one that you currently hold can lead to inefficiency. Therefore, we modify the SCNG

504 Matching

algorithm by removing the potential penalty for relinquishing a room. We begin with the assumption that only one of the rooms is occupied before the matching process begins. The new allocation scheme, called the *serial choice with guarantee* (SCG1) algorithm, proceeds in the same fashion as SCNG, except that *all* students are numbered, perhaps by lottery, and all participate, and if at some stage someone claims a room that you occupy you get to take that student's place in line, which means that it is now your turn to choose. You can reclaim your own room or take a better one if one is available. It is the possibility of getting a better room if someone takes yours, and the right to reclaim your room if nothing better is available, that guarantees that the outcome will be efficient.

DEFINITION: Serial choice with guarantee (SCG1)

All students are numbered. Let $s_1, s_2, ..., s_n$ be the resulting priority ranking. Let R_1 denote the set of *all* rooms. Let R_t be the set of rooms that have not been selected prior to stage *t*. The algorithm proceeds according to the SCNG, with student s_t choosing from R_t at stage *t*, until we reach a stage *t* at which s_t claims a room that is already occupied, say by student *s*. At this point the algorithm is suspended while the ordering of students is changed, with *s* replacing s_t as the next student to choose. The priority ranking is otherwise unaffected. The algorithm now resumes with *s* choosing from R_t .

The SCG1 circumvents SCNG's potential inefficiency when there is only one room that is currently occupied, as we now demonstrate.

Example 4.2: Efficiency of SCG1 for the utility functions of Example 4.1

AB, *AC*, *BA*, *BC*, *CA*, and *CB* are the six different possible orderings of all three students. The utility functions are the ones given in Table 9.18 of Example 4.1, and *A* currently occupies room *X*. If the priority ranking is *AB* or *AC*, then *A* will go first and choose *Y*. If *B* goes second *B* will choose *X*, and *C* will choose *X* if *C* goes second. With *BA* or *BC* student *B* goes first and will take *X*. Because *X* is occupied by *A* the ordering of students is changed from *BA* to *AB* or from *BC* to *AB*. In either case *A* now goes first and choose *Y*. Student *B* will go second and choose *X*. With *CA* or *CB* student *C* will choose *Y* first and *A* will get *X* with *CA*. With *CB* student *B* claims *X* after *C* chooses *Y*. But when *B* claims *X* the ordering is changed, with *A* choosing after *C* and selecting *X*. Table 9.20 summarizes. Note that *A*'s expected utility with the SCG1 is

$$\frac{4}{6} \times U_A(Y) + \frac{2}{6} \times U_A(X) = \frac{4}{6} \times 4 + \frac{2}{6} \times 3 = \frac{22}{6},$$

which is greater than $U_A(X)$, the utility from opting out and holding on to *A*'s original room.

4. Allocating Dormitory Rooms

Each of the possible outcomes of the SCG1 is efficient. In four of the cases student A gets the room (Y) that maximizes A's utility. Therefore, any change in A's room assignment will make A worse off. And both B and C prefer X to Z so any change in their assignments will result in someone being moved

Priority ranking	A	В	С
AB	Y	X	Z
AC	Y	Z	X
BA	Y	X	Z
BC	Y	X	Z
CA	X	Z	Y
CB	X	Z	Y

Table 9.20

from *X* to *Z*, and that person will be made worse off. The last two assignments in Table 9.20 give *C* his or her most preferred alternative, *Y*. Again, the other two each prefer *X* to *Z* so any change in their assignments will move one of them from *X* to *Z*, making that person worse off. Therefore, none of the assignments can be changed without making someone worse off: The matching determined by SCG1 is efficient.

Even if a student already in a room were allowed to opt out, the SCG1 would precipitate an efficient outcome because of the guarantee that participation would not diminish a student's utility. That guarantee increases participation and forestalls any inefficiency. Rather than proving this we generalize the SCG1 algorithm to the case where two or more rooms are already occupied and then establish the efficiency of the more general version.

If there are two or more rooms that are already occupied we have to adjust the SCG1 algorithm to deal with possible *loops*. A loop is a subset R' of rooms and a sequence $s_1, s_2, ..., s_t$ of students such that the first choice of s_1 from R' is the room occupied by s_2 , the first choice of s_2 from R' is the room occupied by s_3 , and so on, with the first choice of student s_{t-1} from R' being the room occupied by s_t and the first choice of s_t from R' being the room occupied by s_1 . When that happens we give all of the t students in the loop their chosen rooms and then remove all t students and all t of the chosen rooms from the process, continuing with the algorithm applied to the rooms that have yet to be assigned and the students who do not yet have rooms.

DEFINITION: Serial choice with guarantees (SCG) in general

Let $s_1, s_2, ..., s_n$ denote the priority ranking (of all students), with R_1 being the set of all rooms. Let R_t be the set of rooms that have not been selected prior to stage *t*. (Rooms that had already been occupied when the algorithm

started will belong to R_t if and only if they have not yet been newly assigned to someone by the algorithm.) The student s' with the highest rank in the priority ordering who has not yet received a room chooses from R_t . If the chosen room was originally unoccupied then it is assigned to s' and removed from the set of available rooms resulting in R_{t+1} , the new set of available rooms. If s' does choose a room that had already been occupied, say by s'', the algorithm searches for a loop. If there is no loop, the ordering of students is changed, with s'' now having highest priority and s' second. The algorithm now resumes with s'' choosing from R_t . If a loop *is* found then all students in the loop are assigned their chosen rooms, and all those rooms are removed from R_t to constitute the set R_{t+1} of rooms available at the next stage.

The outcome of the SCG is strategy proof because at each stage the student whose turn it is to choose gets the room that ranks higher in his or her reported preference ordering than any other available room, and no student can affect the order in which he or she chooses or the set of rooms available by changing his or her reported preference ordering. Therefore, choosing according to one's true preference ordering is a dominant strategy. This remains true for all students that are part of a loop. They get their most-preferred rooms from the set of rooms available when the loop is identified. What about the case of a student *s*' who slips one position in the queue because *s*' has chosen the room occupied by student *s*'' but these students are not part of a loop? Instead of *s*' choosing followed by *s*'', under truthful revelation student *s*'' will choose, followed by *s*'. If *s*' forestalls the switch in the ranking by choosing a room that is *not s*''s most preferred (in the set of available rooms) *s*' will do no better than get the room ranked second in *s*''s preference ordering, and *s*' is guaranteed at least that if *s*'' chooses first.

As we did in the earlier sections, we formally investigate incentive compatibility by supposing that all students give their preference ordering to a referee, who then uses the reported preferences to work out the matching determined by the algorithm in question. Would submitting the true preference ordering be a dominant strategy?

DEFINITION: Strategy proofness

An algorithm for determining an assignment of rooms to students is strategy proof if for each specification of the student preferences truthful revelation is a dominant strategy for each agent.

We have informally demonstrated that the SCG passes this test:

The SCG is strategy proof, and it always generates an efficient outcome, regardless of the number of rooms already occupied at the time it is employed.

4. Allocating Dormitory Rooms

We begin by proving rigorously that for every specification of individual preferences for rooms, truthful revelation is a dominant strategy for every student when the SCG is employed. Consider arbitrary student σ_t , where $\sigma_1, \sigma_2, \ldots, \sigma_n$ is the ordering in which the students actually chose as a result of applying the SCG to the original priority ranking s_1, s_2, \ldots, s_n . Let P_t be the set of available rooms when arbitrary student σ_t has an opportunity to choose. All other rooms have been assigned by the time σ_t has his or her turn. Clearly, P_1 is the set of all rooms, so σ_1 (who might not be s_1) gets the room that is at the top of σ_1 's true preference ordering. Student σ_1 has no incentive to reveal a preference ordering different from his or her true one-that is, to choose a room that is not most preferred according to σ_1 's true preference scheme. Similarly, student σ_t has no incentive to choose anything but his or her most-preferred alternative in P_t . Student σ_t can't do anything to change the set of available rooms when it is his or her turn to choose. If student s chooses a room r' that was already occupied by a student s' before the SCG algorithm began, and s' has not yet been assigned a room and there is no loop, then s' gets the turn of student s, who then follows s'. If s' elects to keep his or her room r' then it is removed from the list available to s. Then s will choose room r, the one in P_t that ranks above every other member of P_t except r'. Had s not chosen r' initially then s's choice from P_t could not have ranked higher than r. In other words, student s has no incentive to reveal a preference for rooms other than s's true preference. Finally, suppose that σ_t is the beginning of a loop. Then the SCG gives σ_t his or her most-preferred room in P_t if σ_t reveals truthfully. Student σ_t could have made a different choice, to prevent the loop from forming, but σ_t still would have been in the position of choosing from P_t . Because σ_t 's declared preference can have no effect on the set of rooms from which σ_t is allowed to choose, that student cannot profit by deviating from truthful revelation. Can a student benefit by deviating from truthful revelation to create a loop? No. That will not make any room available that would not have been available under truthful revelation. We have established that the SCG induces truthful revelation by all students, whatever their priority rankings.

Now, we turn to the efficiency claim. Let s_1, s_2, \ldots, s_n and $\sigma_1, \sigma_2, \ldots, \sigma_n$ and P_t be defined as in the previous paragraph. Let π^* be the matching precipitated by SCG. We show that we cannot alter π^* without making one of the students worse off. Clearly, student σ_1 gets the room at the very top of σ_1 's preference ordering, whether or not σ_1 and s_1 are the same. Therefore, we can't alter π^* without making σ_1 worse off. Suppose that we have proved that we cannot alter π^* without making one of the students $\sigma_1, \sigma_2, \ldots, \sigma_t$ worse off. We now show how it follows from that supposition that we cannot alter π^* without making one of the students $\sigma_1, \sigma_2, \ldots, \sigma_t, \sigma_{t+1}$ worse off. If we are not to make one of the first t + 1 students worse off then we know that we can't change the room assigned by π^* to one of the first t students. Therefore, if we are not to make one of the first t + 1 students worse off, the set of rooms available for σ_{t+1} is precisely the same as the set available under the SCG. The previous paragraph showed that σ_{t+1} gets his or her most-preferred room from that set, and thus we can't change the room assigned by π^* to σ_{t+1} without making that person worse off,

Table 9.21

A	В	С	D	Ε
W V	Ζ	W Y	W Z X	X Y W V Z

unless we give σ_{t+1} one of the rooms assigned by π^* to one of the first *t* students. We have proved that if we don't change the assignment to one of the first *t* students then we can't change the assignment to σ_{t+1} without making one of the first t + 1 students worse off. And we know that we can't change the assignment to σ_1 without making σ_1 worse off. (Note: σ_t and s_t may not be the same.) It follows that we can't change the room assigned by π^* to *any* student without making some student worse off. Therefore, π^* is efficient.

Example 4.3: Application of the SCG algorithm

There are five students, *A*, *B*, *C*, *D*, and *E*, and five rooms, *V*, *W*, *X*, *Y*, and *Z*. Three of the rooms are presently occupied: *W* by student *B*, *X* by student *C*, and *Y* by student *D*. All five students will participate in the allocation scheme because all three tenants are guaranteed rooms that are at least as attractive as the ones they currently occupy. Assume that the priority ordering is *ABCDE*. A partial list of the students's preferences is presented in Table 9.21. We have given only a partial preference ordering for the first four students because we don't need any more information to work out the details of the SCG in this case.

Student *A* goes first and chooses *W*, which is already occupied by *B*, so the priority immediately changes to *BACDE*. Then *B* goes first and chooses *Z*, which is not occupied, so room *Z* is assigned to *B* and removed from the list of available rooms. *A* then choose *W*, which is *not* occupied by someone waiting for a turn to choose, so *W* is assigned to *A* and removed from the list of available rooms. The available rooms are now *V*, *X*, and *Y* with *C*, *D*, and *E*, in that order, waiting to choose. Student *C* chooses *Y*, which is occupied by *D*, and *D* would choose *X*, which is occupied by *C*. We have a loop, and thus room *Y* is assigned to *C* and *X* is assigned to *D*. That leaves room *V* to be assigned to student *E*. The resulting matching π is displayed as Table 9.22. Is this matching efficient? *A* and *B* both wind up with the rooms that they prefer to every other room. Therefore,



A	В	С	D	Ε	
W	Ζ	Y	X	V	

if we change the matching in a way that changes the room assigned to A we will make A worse off, or if we assign a different room to B we will make B worse off. Therefore, in searching for a matching that makes at least one person better off than π and no one worse off we have to give W to A and Z to B. That leaves rooms V, X, and Y to be assigned to C, D, and E. The SCG assignment π gives C

4. Allocating Dormitory Rooms

the room that *C* most prefers, subject to the constraint that *A* and *B* have to get *W* and *Z*, respectively. Therefore, if we change the room assignment of *A*, *B*, or *C* then we will make one of those three worse off. Similarly, the SCG assignment π gives *D* the room that *D* most prefers, subject to the constraint that *A* and *B* have to get *W* and *Z*, respectively. Therefore, if we change the room assignment of *A*, *B*, *C*, or *D* then we will make one of those students worse off. It follows that if we do not assign room *V* to student *E* then we will inevitably leave *A*, *B*, *C*, or *D* worse off than they are at the SCG matching π . Therefore, if π is our starting point, we cannot make one of the five students better off without leaving at least one of them worse off. Therefore, π is efficient.

Finally, we consider whether one of the students could have been assigned a better room (according to his or her own preferences) by making a different choice than the one that we ascribed to the student in the first part of the example. Because A and B both wind up with the rooms that they prefer to every other room, neither could have profited by deviating from truthful revelation. There is nothing that C can do to change the set of rooms available when it is C's turn to choose, and because the SCG gives C his or her most-preferred room from that set individual C cannot profit by misrepresenting his or her preferences. The same can be said of D. Student E goes last, and there is nothing that E can do to get an earlier choice. Therefore, student E cannot profit by misrepresenting E's preferences.

Source

The section is based on Abdulkadiroğlu and Sönmez (1999). Example 4.1 is from that article.

Links

Zhou (1990) proves that there is no strategy-proof allocation scheme that treats individuals symmetrically and always yields an efficient outcome. Chen and Sönmez (2002) test SCNG and SCG. The latter is significantly more likely to result in an efficient allocation in laboratory settings with human subjects. See also Chen and Sönmez (2004).

Problem set

- 1. We proved that if there are no existing tenants then the outcome of the SCNG is always efficient. The proof was quite simple. Why doesn't it work when some of the rooms *are* already occupied?
- **2.** For each possible priority ordering determine the matching generated by the SCNG for Example 4.1 if none of the three rooms is originally occupied—the case of a freshman dorm, for instance.
- **3.** There are five students, *A*, *B*, *C*, *D*, and *E*, and five rooms, *V*, *W*, *X*, *Y*, and *Z*. Agent *A* already occupies room *Y*, and no other room has a tenant before the new room assignments are determined. The order in which students choose is determined by a random draw from a uniform probability distribution, and

then the SNGN algorithm is employed. Specify a utility function for student *A* and preference orderings for the other students so that *A* decides not to participate in the lottery and the resulting outcome of the SCNG is inefficient for at least one priority ranking.

- **4.** Prove that the student who chooses first is *not* a dictator (in the sense in which that term is used for the Gibbard-Satterthwaite Theorem of Chapter 7) if the SCG is used.
- **5.** For the case of five students and five rooms, specify the student preferences so that the SCG assigns all their most preferred rooms, regardless of how many rooms are occupied and who occupies them. Explain your answer.
- **6.** Work out the matching for the SCG when the priority ranking is *EDCBA* and three of the rooms are already occupied—*W* by *B*, *X* by *C*, and *Y* by *D*. The preferences of the five students for the five rooms are given as Table 9.23.

Table 9	9.23
---------	------

A	В	С	D	Ε
W	Z	W	W	X
V	X	Y	Z	Y
Z	Y	V	X	W
X	W	X	Y	V
Y	V	Z	V	Z

7. Work out the matching for the SCG when the priority ranking is *ACEBD* and three of the rooms are already occupied—*W* by *B*, *X* by *C*, and *Y* by *D*. The preferences of the five students are given as Table 9.23.

5 KIDNEY TRANSPLANTS

End stage renal disease, also called chronic kidney disorder, can be treated with dialysis or by transplanting another kidney into the patient. Transplant is the better treatment method because it offers a greater probability of success. Most transplanted kidneys are taken from cadavers, but a significant number are provided by live donors. Of course, transplants depend on the availability of kidneys from either a cadaver or a live donor. There is a substantial gap between the demand and the supply, and the gap is growing.

In 2002 almost 55,000 patients in the United States were on the waiting list for a transplant, and about 8,500 of these received a kidney from a cadaver and about 6,200 received one from a live donor. Every year about 6% of the people on the waiting list die before a kidney becomes available, and between 1% and 2% become too ill to qualify for a transplant. A particular kidney can be transplanted into a particular patient only if two criteria are satisfied: the blood types of recipient and donor have to be compatible (that's not the same as saying

5. Kidney Transplants

that the types have to be *identical*), and the recipient must not have antibodies that would attack the donor's proteins.

Suppose (temporarily) that no cadaver kidneys are available. We begin the analysis of matching patient and kidney donor with a set $S = \{s, s', s'', \ldots\}$ of n patients requiring a transplant. Each patient s has a willing donor r(s) whose kidney may or may not be a good match for s. Think of $R = \{r(s), r(s'), r(s''), \ldots\}$ as a set of kidneys. The suitability of the match between s and r(s) is not binary: There are impossible matches, bad matches, good matches, better matches, and ideal matches. We can think of a patient's "preferences" as a list of the kidneys in order of the suitability of their match with that patient.

We now have a parallel to the room allocation problem of the previous section. The kidneys play the role of the rooms waiting to be attached to a student, with the patients in the role of the students. We take our cue from the serial choice with guarantees (SCG) algorithm to determine the matches between patients and kidneys. Why do we need to have the guarantees? Because each patient s_i already has a willing donor r_i , and to participate in the scheme a patient would have to be assured of receiving a kidney that is at least a good match for the patient as the one that the donor—typically a relative—is able to provide. If r_i has some chance of being successfully transplanted in s_i then s_i is treated as a student with an existing room. The extension of the SCG algorithm will precipitate a matching that is stable and efficient and gives all patients incentive to reveal their preferences truthfully.

The algorithm begins by searching for a loop involving only the patients who come paired with donors. There is a loop if some k of the patients can be

Tabl	e 9	.24
------	-----	-----

A	В	С
Y	Z	X
X	Y	Z
Z	X	Y

numbered s_1, s_2, \ldots, s_k so that $r(s_2)$ is the kidney most preferred by $s_1, r(s_3)$ is the kidney most preferred by $s_2, r(s_4)$ is the kidney most preferred by s_3 , and so on, with $r(s_1)$ being the kidney most preferred by s_k . If there is a loop, all patients in the loop are given their most-preferred kidneys, and then all of these patients and kidneys are removed from further consideration. The algorithm then searches for another loop, and so on, until all loops have been identified and removed from further consideration.

Example 5.1: No bilateral exchanges are possible

There are three patients, A, B, and C. All have siblings willing to donate kidneys to them, but not directly interested in sacrificing kidneys for anyone else. Unfortunately, A's sibling X is a bad match for A, patient B's sibling Y is a bad match for B, and C's sibling Z is a bad match for C. Suppose also that X is a bad match for B, Y is a bad match for C, and Z is a bad match for A. Assume that "bad match" means virtually a zero chance of a successful transplant. None of the patients can be matched with his or her sibling. We can't even find two pairs of siblings who can exchange kidneys with any chance of success. Suppose, however, that if X gives a kidney to C, person Y gives a kidney to A, and Z gives a kidney to B then each patient gets a viable kidney. This assignment would be achieved by the SCG with any priority ordering, as we now show: Table 9.24 gives the preferences of the three patients. We have a single loop with *A* choosing "the room occupied by *B*" and *B* choosing "the room occupied by *C*" and *C* choosing "the room occupied by *A*." The algorithm deals with that loop by giving all their choices. In other words, *X*'s kidney is transplanted into *C*, *Y*'s kidney is transplanted into *A*, and *Z*'s kidney is transplanted into *B*.

Unless we make the bizarre assumption that each donor's objective is to have his or her kidney implanted in a sibling, rather than to extend the family member's life, a loop accomplishes the goal of each donor and patient, even though no patient *s* receives the kidney of the original donor r(s).

Now let's complicate the model by adding *m* patients on the waiting list for a cadaver kidney. Their priority on the waiting list is expressed by w_1, w_2, \ldots, w_m , with w_1 first in line, w_2 second, and so on. The patients on the cadaver waiting list have "preferences" for the live donor kidneys $r(s), r(s'), r(s''), \ldots, r(\hat{s})$ just as the patients in *S* do. Suppose, to take an extreme case, that r_1 is incompatible with each of the patients s_1, s_2, \ldots, s_n but r_1 is at the top of w_1 's preference ordering. Then by extension of the SCG algorithm to the kidney allocation problem, r_1 is assigned to w_1 but then s_1 assumes w_1 's place at the head of the waiting list for a cadaver kidney. Let's generalize this.

The algorithm searches for a loop (as defined in the previous section) involving the patients in *S* and their donors's kidneys. All members of each loop are assigned their "chosen" kidneys, and both are removed from the allocation problem for the next stage. When all the loops have been dealt with the algorithm searches for a *w*-chain, which involves the patients in *S*, their donor partners's kidneys, and the cadaver waiting list. There is a *w*-chain if we can number some *k* of the patients in *S* so that $r(s_2)$ is the kidney most preferred by s_1 , $r(s_3)$ is the kidney most preferred by s_2 , $r(s_4)$ is the kidney most preferred by s_3 , and so on, with none of the kidneys suitable for s_k , for whom the waiting list is the preferred option. At this point s_i is assigned the kidney $r(s_{i+1})$ for i = 1, 2, ..., k - 1, and s_k receives high priority on the cadaver waiting list. The kidney $r(s_1)$ is either offered to the cadaver waiting list or to one of the patients in *S*. When this is done, the algorithm searches for loops again. The procedure continues in this fashion until no patient is left.

Source

This section is based on Roth, Sönmez, and Ünver. (2004).

Links

Su and Zenios (2003) and Votruba (2003) have other perspectives on kidney allocation. Blair and Kaserman (1991) argue that the supply of cadaver kidneys could be substantially increased by very simple (and ethical) changes in public policies and law.

10

General Competitive Equilibrium

1.	Com	npetition, Property Rights, and Prosperity	514
	1.1	Competition and reputation	515
	1.2	Responsiveness of competitive markets	516
	1.3	Why not China?	518
	1.4	To make a long story short Problem set	519 523
2.	The	Arrow-Debreu Economy	523
	2.1	The model	524
	2.2	Welfare theorem for an exchange economy	529
	2.3	The welfare theorem in the general model	531
	2.4	Externalities	533
		Problem set	537
3.	Non	Problem set	538 542
4.	Effic	ciency and Incentive Compatibility	543
	4.1	Dominant strategy equilibrium	544
	4.2	Nash equilibrium	548
	4.3	Subgame-perfect Nash Equilibrium	549
	4.4	Price taking	554
		Problem set	556
5.	Com	nmon Property Resources	556
		Problem set	560

514 General Competitive Equilibrium

This chapter examines a hidden characteristic problem of great significance: We take an economy-wide perspective and ask if there is a mechanism that will elicit private information about individual preferences and firm production recipes in a way that allows an efficient allocation of private goods and services to be identified and implemented. We assume away all other hidden information problems. In particular, every consumer is assumed to know the quality of every firm's output, every employer knows the abilities of every prospective employee, every lender knows the probability of default of every creditor, and so on. Every manager can be relied on to maximize profit. In fact there is no shirking by anyone.

The economy still has an impressive challenge—to induce truthful revelation of the remaining hidden information, specifically the preferences and production functions. In fact, three of the five sections even assume away this hidden information problem, highlighting instead the transmission of information. Recall that an outcome is efficient if there is no other arrangement of production and consumption activities that makes one person better off without lowering the utility of anyone else. Identification of an efficient outcome would seem to require an enormous amount of information about all of the private characteristics. Therefore, even with most of the hidden information problems assumed away, identification of an efficient outcome by the market system is a remarkable accomplishment. Marginal social cost pricing is the key.

In a private ownership market economy the consumer maximizes utility subject to a budget constraint, and this results in a chosen consumption plan at which the marginal rate of substitution (MRS) for each pair of commodities equals the price ratio for those goods. Therefore, the price ratio transmits information about one consumer's MRS to every other consumer and every firm. The budget constraint gives consumers the incentive to employ that information in their utility maximization calculations, as Subsection 1.4 demonstrates. Profit maximization by a firm in a competitive environment results in the equilibrium price of a commodity that is also equal to the marginal cost of producing that good. Therefore, prices transmit information about consumer preferences and firm production recipes—*marginal* information.

In spite of the fact that prices transmit only marginal information, in Sections 2 and 3 we demonstrate that this marginal information suffices to allow an efficient outcome to be identified in a wide range of circumstances. The first section is an informal examination of incentives at work in a capitalist economy. In Section 4, we look at one specific incentive issue—the ability of a trader to advantageously manipulate prices. In the last section, we look at the performance of the market system when there is a resource that is jointly and freely employed by several firms.

1

COMPETITION, PROPERTY RIGHTS, AND PROSPERITY

Competition between political parties makes it very difficult for a ruling party to conceal serious problems or to ignore them if the cat does get out of the bag.

1. Competition, Property Rights, and Prosperity

Competition between firms reduces the likelihood that a firm that attempts to profit by deceiving consumers can survive. Competition among firms in a private ownership market economy is also a key ingredient of efficiency and economic growth. We examine these claims briefly and informally in this section.

1.1 Competition and reputation

The producer of a commodity has two sets of rivals. The obvious group consists of the other producers, but in a market economy consumers are the also firm's adversaries in the sense that, unless somehow constrained, the producer can increase profit by misleading its customers—about product quality, in particular. Product quality has many dimensions—durability, versatility, performance, operating cost, maintenance cost—and there are many opportunities for the unscrupulous manufacturer to reduce production costs by sacrificing quality while keeping up a good appearance. In a command economy, where profit maximization may not be a firm's chief goal, the workers in a production plant can still gain by misleading consumers about quality. Their task is easier if they do not have to maintain high-quality output.

What prevents a producer from succumbing to the temptation to gain at the expense of consumers? Regulation by a government agency sometimes plays an important role, even in a market economy. Far more important, typically, is the discipline of competition under capitalism. Modern technology requires a heavy initial capital outlay for the production of most goods and services. Competition among the owners of capital keeps the rate of return on capital low enough so that the initial expenditure can be recovered only after many (or at least several) years of sustained production and sales. Therefore, the firm is not just interested in profit today. High current profit will lead to heavy losses in the future if present profits rest on duplicity, which is discovered and broadcast to consumers by word of mouth, by consumer research firms, and by the media. When consumers have the option of buying from a firm's competitors, the temptation to sacrifice both consumer welfare and the firm's future is held in check to a great extent. Warranties also play a role. If other firms provide a comprehensive warranty, then firm X must do the same to stay afloat. If X were to deceive customers by manufacturing low-quality items and passing them off as high-quality products, then X would suffer heavy losses in the future as items are returned for replacement or repair under the warranty.

In some cases it is not easy to discover that the consumer has been misled. Consider housing construction. Defects may not show up for years, but when they do they can be extremely costly—wiring that causes fires, leaky roofs, and so forth. It will be very difficult for consumers to determine that construction company *X* produces houses that are more defect prone than those of other firms. The same is true for the suppliers of health care. And warranties won't work well in either case. A meaningful warranty on a new house would have to be long lived, but that presents the construction firm with a hidden information problem. Warranties remain in force as long as the consumer undertakes proper maintenance. This is easy to monitor in the case of new cars, which can be

516 General Competitive Equilibrium

brought back to the dealer for periodic inspections, but is very difficult in the case of new houses. More significant is the fact that it becomes harder to define

In spite of the difficulties that we've pointed to, construction companies are not heedless of their reputation. Nevertheless, housing construction is typically regulated by the government as well as by the market. Because the cost of mistakes or fraud can be enormous, there is a case for government regulation even if defects could be eventually attributed to the producer. It is no coincidence that regulation is more stringent in Florida and California, where hurricanes and earthquakes are a constant worry. "proper maintenance" the more time has lapsed since purchase. And if maintenance requirements can't be specified, contracts cannot be written conditional on maintenance taking place. But if the home owner is not penalized for failure to keep the house in good repair then the owner has little incentive to exercise preventive care. A serious moral hazard problem prevents long-term warranties from being offered.

But for a wide range of goods and services, competition among producers forces the individual firm to be concerned with its reputation. It is the lack of competition in the manufacture of appliances in the former Soviet Union that

accounts for the explosion of thousands of television sets per year in Russia (see Milgrom and Roberts, 1992, p. 13).

Competition and reputation also have an important role to play in the political realm, of course. One reason why pollution reached grim proportions in Eastern Europe by 1990 is that the lack of political competition made it relatively easy for ministers responsible for the environment to conceal problems and thereby minimize the probability of being sacked. Politicians in multiparty, democratic countries also have an interest in covering up, but political rivals and a compet-

In the 1980s in some areas of the former East Germany 90% of children suffered respiratory disease. Pollutionrelated cancers and infant mortality soared in Czechoslovakia during the 1970s and 1980s. During the same period, 95% of Polish rivers were polluted and the leukemia rate was soaring (*Business Week*, March 19, 1990, pp. 114, 115). itive press make it much more difficult to escape detection. "It is significant that no democratic country with a relatively free press has ever experienced a major famine.... This generalization applies to poor democracies as well as to rich ones" (Sen, 1993, p. 43). A. K. Sen has demonstrated that even in severe famines there is enough food to sustain the entire population of the affected region, and that leaders who must seek reelection are far more likely to take the steps necessary to see that food is

appropriately distributed. The incentives confronting political leaders have life and death consequences for millions.

1.2 Responsiveness of competitive markets

A lot of creativity can be traced to market forces. A firm has to improve its product and come up with new products, not just to respond to a rival's innovation but to anticipate what its rivals might do. "Catch-up" is a risky game to play. A firm is much more likely to survive and prosper if it takes its rivals by surprise. Each firm in an industry wants to be first with an innovation. The greater the competition in the industry, the more desperately a firm will strive to be the front runner. Several times a day you encounter evidence of the resulting creativity of markets.

1. Competition, Property Rights, and Prosperity

You can probably quickly recall quite a number of product innovations that owe their existence as much to some firm's fear of losing ground to others as to the ingenuity of someone's mind. Here's my list:

- Major cities now have stores that sell used wedding gowns.
- Parents can obtain the services of reliable firms for educational testing and for the provision of transcripts for children educated at home.
- In the 1990s Wall Street created catastrophe bonds to provide insurance to insurance companies—against the correlated losses that can occur when, for example, a natural disaster creates widespread damage in a particular area. Catastrophe bonds pay a higher rate of interest than other corporate bonds, but the insurance company doesn't have to pay back the principal at all if, say, a hurricane does more than \$1.5 billion of damage. If the damage is between \$1 and 1.5 billion then only part of the principal has to be repaid (Wheelan, 2002, p. 123). Is this a valuable service from the standpoint of overall consumer welfare?
- A handful of parking lots have introduced bunk beds for parked cars to maximize the number of parking spots that a plot of land can support.
- Some firms have equipped trucks with document shredders, so that other firms, too small to purchase their own shredders, can avail themselves of the service at a price they can afford.
- Before the U.S. government broke AT&T's monopoly hold on telephone service, AT&T would not allow customers to use non-AT&T equipment. More significantly, they were under no pressure to provide useful services to customers—such as a telephone ringer that could be turned off when the baby is asleep. (My wife and I put thick tape around the bell in the 1970s.) Now that the suppliers of telephone services face vigorous competition, all kinds of significant features are available. (Think about how the U.S. government postal service has changed in response to competition from Federal Express and UPS.)
- Some express highway lanes charge a fee by taking a picture of your licence plate (e.g., Highway 407 in Toronto) and sending you a bill in the mail. At least one firm has gone into business selling license plate covers that are transparent to the eye—that is, to the eye of a police officer—but are opaque to the lens of a camera.
- My son-in-law Tyler Cowie owns and operates his own computer graphics business. He used to pick up jobs from the printer in a taxi, until one driver offered to pick up his printing without Tyler having to be in the cab. University bookstores now sell used textbooks, probably because of the pressure of competition from firms that can ship new textbooks at discount prices within a few days.
- A firm in Huntsville, Alabama, sells the contents of luggage that was lost by airplane travelers.
- Firms have sprung into business to provide reliable evaluations of high school and university transcripts from China, Singapore, and so forth for American universities processing applications for admission to undergrad-uate programs and graduate schools.

518 General Competitive Equilibrium

• Television stations have been known to purchase unsold tickets to a football game so that the local broadcasting blackout could be lifted.

Pre–Industrial Revolution economies of the late eighteenth century offered even to the wealthy—only goods that were also available in ancient Rome. The

Beverage cans are made from two pieces of aluminum. A very thin disk is stamped to form the bottom and sides. The top is a separate disk that is crimped onto the sides. Because the contents are under pressure and the top is scored to allow the pop top to tear open when the consumer applies force, the top must be substantially thicker than the sides, adding to the cost of producing the can. Consequently, in the 1970s the sides of the can were tapered near the top to reduce the surface area of the top disk (Petrosky, 1996, p. 102). Think about the intricate web of market forces that inspired this innovation.

rate of economic growth over the 1500 years prior to the Industrial Revolution was essentially zero. Contrast that with the impressive stream of innovations just hinted at in previous paragraphs. The flow of goods and services per capita from the world's mature capitalist economies is from several hundred percent to several thousand percent higher than 150 years ago. Why?

Capitalist, or free market, economies provide incentives that promote growth. Entrepreneurs are motivated to devote their energies to productive outlets, rather than to activities that merely transfer wealth from one pocket to another, as when an individual acquires land from a grateful monarch for services rendered. The rule of law plays a key role:

Contracts are enforced and property is protected from arbitrary expropriation.

1.3 Why not China?

Why did the Industrial Revolution start in England and a few other countries in Western Europe in the late eighteenth century and not much earlier in China? So

Most of the following Chinese inventions are from the T'ang and Sung dynasties (618–906 AD and 960–1126 AD, respectively): paper, movable type, compasses, water wheels, sophisticated water clocks, gunpowder, spinning wheels, mechanical cotton gins, hydraulic trip hammers, ship construction techniques that permitted larger and more seaworthy vessels, sternpost rudders, superior sail designs, porcelain, umbrellas, matches, toothbrushes, and playing cards (Baumol, 1993, p. 42). many fundamental technological breakthroughs are of ancient Chinese origin.

There is no question that China had sufficient technical know-how for sustained economic growth. But it lacked a political and legal environment capable of nurturing and sustaining private enterprise. The monarch claimed the right to all property, and this right was used by the monarch to raise money to solve budget difficulties or finance wars. "Private" property could be confiscated at any time. Also, the state thwarted private enterprise, and when it didn't succeed in stopping a venture that had got under way it would often confiscate the fruits of inventive activity, as it

did with paper, printing, and the bill of exchange. Consequently, any wealth that a merchant did succeed in accumulating was typically used to purchase land or invest in the enterprise of becoming a scholar-official. (English property owners led the world in extracting property right guarantees from the monarch.)

1. Competition, Property Rights, and Prosperity

The Chinese scholar-official, called a Mandarin, held a position in what we would today call the civil service. A Mandarin occupied a far more prestigious position in Chinese society than even an extremely successful merchant or industrialist. One could not become a Mandarin without passing the astonishingly difficult civil service exams, and these required years of preparation, study, and tutoring. Families devoted lavish amounts of effort and resources in the attempt to get their children through these exams. The Mandarin's pay was low, however, and this led to corruption: He often extorted money from the people under his jurisdiction to boost the return on the investment in his education.

1.4 To make a long story short

Let's suppose that we have been hired as consultants to design an economic system. Suppose also that we have never heard of the market system. The attempt to design a system that elicits enough information about individual preferences and firm production technologies to precipitate an efficient outcome will actually steer us toward the market system.

Example 1.1: Two individuals with straight-line indifference curves

Individual 1's utility function is $U_1 = 4a + b$, where *a* is the amount of good *A* and *b* is the amount of good *B* consumed by person 1. Individual 2's utility function is $U_2 = 5x + 2.5y$, where *x* and *y* denote person 2's consumption of commodities *A* and *B*, respectively. Suppose that $\Delta a = 1$. In words, person 1's consumption of *A* increases by 1 unit. Then

$$\Delta U_1 = 4a + 4 + b + \Delta b - (4a + b) = 4 + \Delta b$$

(The Greek letter Δ always denotes change.) Therefore, if $\Delta b = -4$ then $\Delta U_1 = 0$. In other words, 4 units of commodity *B* is the maximum that person 1 would be willing to sacrifice to obtain 1 more unit of *A*. Now, let $\Delta x = -1$. Then

$$\Delta U_2 = 5x - 5 + 2.5y + 2.5\Delta y - (5x + 2.5y) = -5 + 2.5\Delta y.$$

If $\Delta y = +2$ then $\Delta U_2 = 0$. We see that 2 units of commodity *B* are the minimum compensation that person 2 needs for the loss of 1 unit of *A*. Finally, suppose that

$$\Delta a = 1 = -\Delta x$$
 and $\Delta b = -3 = -\Delta y$.

In words, person 2 gives 1 unit of good *A* to person 1 in exchange for 3 units of good *B*. Then person 1 sacrifices less than the maximum that person 1 would be prepared to give up to get another unit of *A*. And person 2 receives more than the minimum amount required to compensate him or her for the loss of a unit of *A*. It follows that the utility of both increases: $\Delta U_1 > 0$ and $\Delta U_2 > 0$. We have increased the utility of both consumers without affecting the consumption of anyone else. (Production plans are unchanged—it's just a matter of two people trading.) We conclude that the original outcome was not efficient.

520 General Competitive Equilibrium

The only thing that could undermine the argument of Example 1.1 is the possibility that b < 3 or x < 1 initially. But suppose that b and x are both positive. Then for some positive number δ we have $b > 3\delta$ and $x > \delta$, and if

$$\Delta a = \delta = -\Delta x$$
 and $\Delta b = -3\delta = -\Delta y$

we still have

$$\Delta U_1 = 4\Delta a + \Delta b = \delta > 0$$
 and $\Delta U_2 = 5\Delta x + 2.5\Delta y = 2.5\delta > 0$

The assumption that both individuals have linear utility functions is quite extreme, but it is very easy to generalize the argument, as we do now.

Example 1.2: Indifference curves that are curved

The commodity bundle (10, 8) provides 10 units of commodity *A* and 8 units of *B*. Suppose that the straight line 4a + b = 48 is tangent at (10, 8) to person 1's indifference curve through (10, 8). Suppose that the equation of the tangent to individual 2's indifference curve through (7, 6) at the bundle (7, 6) itself is 5x + 2.5y = 50. Therefore, for $\delta > 0$ sufficiently small we can set

$$\Delta a = \delta = -\Delta x$$
 and $\Delta b = -3\delta = -\Delta y$

and thus $\Delta U_1 > 0$ and $\Delta U_2 > 0$. The tangent line approximates the indifference curve, but it is an extremely good approximation in a region very close to the tangency point. Therefore, we can use the argument of Example 1.1 if δ is sufficiently small. We have increased the utility of both consumers without affecting the consumption of anyone else, and hence the original outcome is not efficient.

By definition, an individual's MRS at a commodity bundle is the negative of slope of the tangent to the indifference curve at that bundle. For Examples 1.1 and 1.2, the slope of 4a + b = 48 is -4. The slope of 5x + 2.5y = 50 is -5/2.5 = -2. Therefore, person 1's MRS at (10, 8) is 4 and 2's MRS at (7, 6) is 2. Because these marginal rates are not equal we were able to engineer a trade between the two individuals that made both better off and that had no effect on anyone else's consumption. If the original allocation of goods and services is efficient, this won't be possible. Therefore, we have discovered an important necessary condition for efficiency.

If two consumers have a positive amount of each of two commodities then if their marginal rates of substitution are not equal the economy-wide allocation of goods and services is not efficient.

Why should we worry about inefficiency if individuals can get together and trade, as in Examples 1.1 or 1.2, to rectify the situation? Because there are millions of consumers, and hence an astronomical number of trades that may have to be arranged to move the economy toward an efficient outcome. Moreover, the two individuals may be two thousand miles apart. It is vital that an economic

1. Competition, Property Rights, and Prosperity

system get the allocation of commodities right the first time. The market system does: Each consumer maximizes utility subject to the budget constraint, and that results in equality between the individual's MRS and the price ratio. At individual 1's chosen consumption plan his or her MRS will equal p_A/p_B , where p_A is the price of *A* and p_B is the price of *B*. That means that when other consumers determine their own utility-maximizing consumption plans they will take individual 1's MRS into consideration because the price ratio p_A/p_B will play a central role. Because budget-constrained utility maximization implies $MRS_1 = p_A/p_B = MRS_2$, where MRS_i denotes individual *i*'s MRS, person 1's MRS plays a central role in person 2's utility-maximization exercise and vice versa. Of course, this can be said of any pair of consumers.

To summarize: The price ratio transmits information about one consumer's MRS to all other consumers, and the budget constraint gives each the incentive to use this information in his or her planning. This incentive derives from the fact that the price ratio p_A/p_B is the amount of commodity *B* that the individual could have purchased with the amount of money it takes to purchase one unit of commodity *A*. That opportunity cost will play a central role in the utility-maximizing individual's decision making. This information and incentive role of prices is one reason for the high-level performance of the privaten ownership market economy.

Now let's consider the production of goods and services. Reinterpret Examples 1.1 and 1.2 with individual 1 as *firm* 1, and individual 2 as *firm* 2. In that case, $U_1(a, b)$ is firm 1's output when it employs *a* units of input *A* and *b* units of input *B*. Similarly, $U_2(x, y)$ is firm 2's output when it employs *x* units of *A* and *y* units of *B*. (The two firms may supply very different commodities.) If the firms do not have identical marginal rates of technical substitution (RTS; i.e., their isoquants do not have identical slopes at the respective input bundles employed) then they can trade inputs in a way that increases the output of both firms. This extra output can be used to increase the utility of everyone—for instance, by dividing the extra output evenly.

If two firms employ a positive amount of each of two inputs then if their marginal rates of technical substitution (RTSs) are not equal, the economywide configuration of production and consumption activities is not efficient.

Note that cost minimization—one consequence of profit maximization by a firm in the market system implies that each firm's employment of inputs equates the input price ratio p_A/p_B to the firm's RTS. Efficiency requires $RTS_1 = RTS_2$ for any two firms 1 and 2. In the market system, $RTS_1 = p_A/p_B = RTS_2$ and thus $RTS_1 = RTS_2$. The input price ratio transmits information to each firm about the RTS of each other firm, and the profit motive—via cost minimization gives each firm the incentive to use that information. To activate the profit motive, of course, we have to give consumers ownership shares in the firms. They will then have an incentive to motivate the managers of firms to maximize profit.

Now that we have decided to use the market system, suppose that we come on the scene just after it reaches an equilibrium: Each consumer has maximized

522 General Competitive Equilibrium

utility subject to the budget constraint, each firm has maximized profit, given the prices, and demand equals supply for every good. Let's test the efficiency of the equilibrium by trying to divert resources from the production of commodity B to the production of commodity A to see if we can make some individuals better off without harming anyone else. Specifically, we'll withdraw a specific package Rof resources (inputs) from the production of B and use them to produce more A. Suppose that this results in the loss of ∇b units of *B* and an increase of Δa units of A. Now, suppose that this change in the economy-wide output of goods and services allows us to rearrange everyone's consumption in a way that increases everyone's utility. But all consumers have maximized utility subject to their budget constraints. It must be that the bundles, which yield higher utility, were not chosen because they were not affordable-that is, too expensive. If this is true for each consumer, it must be true in the aggregate. Therefore, $p_A \Delta a > p_B \nabla b$, where p_A and p_B denote the equilibrium prices of A and B, respectively. In words, the total amount of money that would have been saved by reducing B consumption is insufficient to finance the total increase in A consumption.

We know that $p_A \Delta a > p_B \nabla b$ and that the change in production was brought about by moving a set R of resources from firms that produce B to firms that produce A. This implies that firms have not maximized profit, contradicting the supposition that we started from equilibrium. Why does $p_A \Delta a > p_B \nabla b$ contradict profit maximization? If each firm maximizes profit, given the prices, then total, economy-wide profit must be maximized at equilibrium. However, the transfer of the set R of resources from B firms to A firms will not change total cost in the economy. Cost will be lower in the industry producing commodity B but will be higher by the same amount in industry A. However, revenue has increased because $p_A \Delta a > p_B \nabla b$. If we can increase total revenue in the economy without increasing total cost, we can increase total profit. But if that is possible, at least one firm must initially have fallen short of profit maximization. That contradicts the definition of equilibrium, which requires, among other things, profit maximization by each firm. (If each firm's profit is maximized then total profit is maximized.) Therefore, we have proved that, starting from an equilibrium of the market system, it is not possible to make everyone better off by changing the composition of goods and services produced.

We seem to have exhausted all the possibilities for improving on the outcome of the market system, without finding a source of inefficiency. Nothing succeeded, so the market equilibrium must be efficient. However, this subsection is far from rigorous enough to be called a proof. It does have the advantage of bringing out some of the intuition behind the remarkable performance of the private ownership market economy. The rigorous proof of efficiency comes in the next section.

Sources

Milgrom and Roberts (1992, p. 13) recount the story of exploding television sets in Russia. The paragraph on famines in Section 1.1 is based on Sen (1981). Hammons (2001) notes the existence of educational testing services for children educated at home. The discussion of economic growth in Section 1.2 is based

2. The Arrow-Debreu Economy

on Chapter 1 of Baumol (2002). Section 1.3 on China is based on pages 31–44 in Baumol (1993).

Links

See Drèze and Sen (1989) for a thorough examination of the problem of hunger. Wheelan (2002, p. 123) provides a brief introduction to catastrophe bonds. Balazs (1964) is a classic study of the Mandarin class.

Problem set

- **1.** Individual 1's utility function is $U_1 = \alpha a + \beta b$, where α and β are positive constants, *a* is the amount of the good *A* consumed by person 1, and *b* is the amount of good *B*. Individual 2's utility function is $U_2 = \rho x + \sigma y$, where ρ and σ are positive constants, and *x* and *y* denote person 2's consumption of commodities *A* and *B*, respectively.
 - **A.** Show that if α/β is not equal to ρ/σ then the two individuals have indifference curves with different slopes.
 - **B.** Show that if *a*, *b*, *x*, and *y* are all positive and the two individuals have different marginal rates of substitution then there is a trade between the two that raises the utility of each.
- **2.** The utility functions of individuals 1 and 2 are, respectively, $U_1 = ab$ and $U_2 = xy$, with *a*, *b*, *x*, and *y* as in question 1. Find the tangent to person 1's indifference curve through the commodity bundle (5, 3) at bundle (5, 3), and the tangent to person 2's indifference curve through (6, 9) at (6, 9). If you don't know calculus, a good approximation will suffice: Draw an accurate indifference curve and use a ruler to construct the tangent. Then find the equation that it represents. Use the arguments of Examples 1.1 and 1.2 to devise a trade between the two individuals that leaves both better off. Plug the new values of *a*, *b*, *x*, and *y* into the utility functions to confirm that $U_1(a, b)$ is now higher than $U_1(5, 3)$ and that $U_2(x, y)$ is now higher than $U_2(6, 9)$.
- **3.** Section 1.4 showed that if the transfer of resources from the production of commodity *A* to commodity *B* results in a change in the total production of *A* and *B* that could be used to make everyone better off, then $p_A \Delta a > p_B \nabla b$, where p_A and p_B denote the equilibrium prices of *A* and *B*, Δa is the increase in the output of *A*, and ∇b is the reduction in the output of *B*. Show that $p_A \Delta a > p_B \nabla b$ holds even if the change in production can be used to make just one consumer better off without making any other consumer worse off, assuming that before the change is made consumers have maximized utility subject to their respective budget constraints.

2 THE ARROW-DEBREU ECONOMY

This section establishes the efficiency of the model of a private ownership market economy developed by K. J. Arrow and G. Debreu, who first worked out the fundamental properties of general equilibrium.

524 General Competitive Equilibrium

2.1 The model

We begin with the pure exchange version: Production has already taken place. As a result, each consumer has an abundance of some commodities and a dearth of others. The individual supplies some goods to the market and uses the resulting income to buy other goods. This *exchange economy* is at equilibrium when the prices have adjusted to the point where supply equals demand in each market simultaneously.

DEFINITION: *n*-person exchange economy

There are *n* consumers, indexed by i = 1, 2, ..., n, and there are ℓ commodities, named $c = 1, 2, ..., \ell$. Individual *i*'s consumption plan is a vector (list) $x_i = (x_{i1}, x_{i2}, ..., x_{i\ell})$ that specifies the consumption of x_{ic} units of each commodity *c*. Individual *i* is endowed with the ℓ -vector $\omega_i = (\omega_{i1}, \omega_{i2}, ..., \omega_{i\ell})$ that specifies the amount ω_{ic} of each commodity *c* that *i* possesses before trade takes place. An *allocation x* is an assignment of a commodity bundle x_i to each individual *i*, and it is *feasible* if

$$x_{1c} + x_{2c} + \dots + x_{nc} \le \omega_{1c} + \omega_{2c} + \dots + \omega_{nc}$$

holds for each commodity c. The preference scheme of individual i is represented by a utility function U_i .

To simplify the notation we let $x_1 + x_2 + \cdots + x_n \le \omega_1 + \omega_2 + \cdots + \omega_n$ represent the statement " $x_{1c} + x_{2c} + \cdots + x_{nc} \le \omega_{1c} + \omega_{2c} + \cdots + \omega_{nc}$ for each commodity *c*."

A *price system* is an ℓ -vector $p = (p_1, p_2, ..., p_\ell)$ specifying the price p_c of each commodity *c*. A *competitive equilibrium* of this economy is a priceallocation pair (p, x) such that for each individual *i* the consumption plan specified for *i* by allocation *x* maximizes *i*'s utility subject to *i*'s budget constraint, and the total consumption of each good equals the total endowment of that good—that is, demand equals supply for each commodity. Individual *i*'s *budget constraint* is

$$p_1 x_{i1} + p_2 x_{i2} + \dots + p_\ell x_{i\ell} \le p_1 \omega_{i1} + p_2 \omega_{i2} + \dots + p_\ell \omega_{i\ell}.$$

We assume that the individual sells all of the goods in *i*'s endowment ω_i for an amount of money $p_1\omega_{i1} + p_2\omega_{i2} + \cdots + p_\ell\omega_{i\ell}$. This money is then used to buy a new collection $(x_{i1}, x_{i2}, \dots, x_{i\ell})$ of goods and services.

Example 2.1: Buying back some of your endowment

There are two commodities, so $\ell = 2$. Individual 1's endowment is $\omega_1 = (12, 0)$. That is, person 1 holds 12 units of the first good and 0 units of the second good before markets open. If $p_1 = 2$ and $p_2 = 5$ then the individual's budget constraint is $2x_{11} + 5x_{12} \le 24$. The consumption plan (7, 2) is affordable because $2 \times 7 + 5 \times 2 = 24$. If the individual demands (7, 2) we portray him or her as selling the entire endowment for \$24 and then using \$14 of that to buy back 7 units of commodity 1. The remaining \$10 is used to purchase 2 units of commodity 2. In fact, the consumer would simply sell 5 units of commodity 1 for \$10 and then use that money to buy 2 units of the second good.

DEFINITION: Competitive equilibrium of an exchange economy

The price system *p* and allocation *x* constitute a competitive equilibrium if for each individual *i* the consumption plan x_i maximizes U_i subject to *i*'s budget constraint, and $x_1 + x_2 + \cdots + x_n = \omega_1 + \omega_2 + \cdots + \omega_n$, which means that demand equals supply for each commodity.

Now let's add production to the model. There are ℓ commodities and n individual consumers (or households) as in the exchange economy. In addition there are m firms. We let $I = \{1, 2, ..., n\}$ denote the set of households and let $J = \{1, 2, ..., m\}$ denote the set of firms. An allocation assigns a consumption plan to each individual and a production plan to each firm. Each individual has an initial endowment and a preference scheme, represented by a utility function. Each firm j has a technology T_j that specifies the production plans that the firm is able to carry out.

DEFINITION: The Arrow-Debreu economy with production

There are *n* consumers and ℓ commodities. Each individual *i* has an initial endowment ω_i . Also, individual *i* may own shares in one or more firms. We let α_{ij} be the fraction of firm *j* owned by individual *i*. We let the utility function U_i represent the preference scheme of individual *i*. An allocation (x, y) specifies a consumption plan x_i for each individual *i* and a production plan y_j for each firm *j*. The production plan $y_j = (y_{j1}, y_{j2}, \ldots, y_{j\ell})$ specifies a positive, zero, or negative number y_{jc} for each commodity *c*. If $y_{jc} > 0$ then the production plan y_j requires $|y_{jc}|$ units of commodity *c* as input. This sign convention allows us to distinguish inputs and outputs, and to compute profit quite easily.

Given the price system $p = (p_1, p_2, ..., p_\ell)$ the production plan $y_j = (y_{j1}, y_{j2}, ..., y_{j\ell})$ yields a profit of

$$p_1 y_{j1} + p_2 y_{j2} + \cdots + p_\ell y_{j\ell}$$
,

which we write as py_j for short. If $y_{jc} > 0$ then $p_c y_{jc}$ is the revenue resulting from the sale of y_{jc} units of commodity *c* at price p_c . If $y_{jc} < 0$ then $|y_{jc}|$ units of commodity *c* are employed as input at a cost of $p_c \times |y_{jc}|$. When we *add* $p_c y_{jc}$ into the firm's profit calculation we are subtracting $p_c \times |y_{jc}|$ from revenue.

Example 2.2: The profit calculation

There are two commodities, and the firm's technology is simply characterized: Each unit of output of commodity 2 produced by the firm requires 2 units of commodity 1 as input. The production plan $y_1 = (-10, 5)$ employs 10 units of commodity 1 as input and produces 5 units of commodity 2 as output. If $p_1 = 3$ and $p_2 = 9$ then the firm's profit is $p_1y_{11} + p_2y_{12} = 3 \times -10 + 9 \times 5 = 15$.

Let's keep track of a particular commodity *c*. Given the allocation (*x*, *y*), the net output of *c* is $\sum_{j \in J} y_{jc}$. Because y_{jc} is a negative number when *j* uses *c* as an input, the expression $\sum_{j \in J} y_{jc}$ gives us the total output of *c* by the production sector less the total amount of *c* used as input by firms. Therefore, $\sum_{j \in J} y_{jc}$ is indeed the net output of commodity *c*. It is possible to have $\sum_{j \in J} y_{jc} < 0$. This would be inevitable if *c* were labor: All firms use labor as an input but it is not produced by any firm. Labor would be supplied by households, of course, and if ω_{ic} denotes household *i*'s endowment of labor then $\sum_{i \in I} \omega_{ic} + \sum_{j \in J} y_{jc}$ is equal to the total endowment of labor in the economy less the total amount used as input. Therefore, $\sum_{i \in I} \omega_{ic} + \sum_{j \in J} y_{jc}$ is the total amount of labor available to households for consumption (as leisure). And $\sum_{i \in I} x_{ic}$ is obviously the total amount of leisure consumed by the household sector. Therefore, any allocation (*x*, *y*) must satisfy

$$\sum_{i \in I} x_{ic} \leq \sum_{i \in I} \omega_{ic} + \sum_{j \in J} y_{jc}$$

if commodity *c* is labor.

Consider another commodity *c* for which $\sum_{j \in J} y_{jc}$ is positive. Then the allocation (*x*, *y*) leads to a net output of *c*, and this can be added to the households' endowment $\sum_{i \in I} \omega_{ic}$ of *c* (if any) to determine the total amount of *c* available for consumption. This total again is $\sum_{i \in I} \omega_{ic} + \sum_{j \in J} y_{jc}$ and again we see that an allocation must satisfy

$$\sum_{i \in I} x_{ic} \le \sum_{i \in I} \omega_{ic} + \sum_{j \in J} y_{jc}$$

We refer to this inequality as the *material feasibility* condition, and it must hold for each commodity *c*.

An allocation must also satisfy the *firm-level* feasibility conditions. Firm *j*'s technology must be capable of turning the inputs specified by y_j into the outputs specified by y_j . Recall that T_j is the set of production plans that are technologically feasible for firm *j*. If y_j belongs to T_j there is no guarantee that *j* will actually be able to obtain the inputs required by y_j . There might be excess demand for one input, making it unavailable to some firms. This will not happen if the material feasibility condition holds, but that is guaranteed to hold only at equilibrium. The individual firm cannot be expected to worry about the economy-wide material feasibility condition.
DEFINITION: Feasibility

The allocation (*x*, *y*) is feasible if

$$\sum_{i \in I} x_{ic} \leq \sum_{i \in I} \omega_{ic} + \sum_{i \in J} y_{ic}$$

holds for every commodity c (material feasibility), and in addition the production plan y_j belongs to the technology set T_j , for each firm j (firm-level feasibility).

Example 2.3: Feasibility in a simple economy

There are two commodities, one individual, and one firm. It follows that $\alpha_{ij} = 1$. (Individual 1 owns the firm and hence the rights to the firm's profit.) The individual's endowment is $\omega_1 = (48, 0)$. The firm's production technology is specified by the equation "output = $2\sqrt{\text{input.}}$ " Suppose the allocation (*x*, *y*) specifies the consumption plan $x_1 = (32, 8)$ and the production plan $y_1 = (-16, 8)$, which employs 16 units of commodity 1 as input and produces 8 units of commodity 2 as output ($8 = 2\sqrt{16}$). If $p_1 = 1$ and $p_2 = 4$ then the firm's profit is 16, and the consumer's budget constraint $x_{11} + 4x_{12} \le \omega_{11} + 4\omega_{12} + 16$ is satisfied. The allocation ((32, 8), (-16, 8)) is feasible because (-16, 8) is in the firm's technology set, and in addition

32 = 48 - 16 and 8 = 0 + 8.

At equilibrium, each firm maximizes profit, given prices and their technology sets, all consumers maximize utility given their budget constraints, and every market clears—in other words, the demand for each good equals the supply. Suppose that the prices are given by $p = (p_1, p_2, ..., p_\ell)$.

Profit maximization is easy to characterize: Recall that $p_1y_{j1} + p_2y_{j2} + \cdots + p_\ell y_{j\ell}$, which we denote by py_j , is *j*'s profit from the production plan y_j . Then the plan y_j maximizes profit if y_j belongs to T_j and $py_j \ge pz_j$ holds for all z_j in T_j . In words, the plan y_j is feasible for the firm, and no other feasible plan yields higher profit.

Derivation of the consumer's budget constraint requires a little work. If x_i is individual *i*'s consumption plan then expenditure is clearly $p_1x_{i1} + p_2x_{i2} + \cdots + p_\ell x_{i\ell}$, which we write as px_i for short. What is *i*'s income? Income from the sale of *i*'s endowment is just $p\omega_i = p_1\omega_{i1} + p_2\omega_{i2} + \cdots + p_\ell\omega_{i\ell}$ but *i* may also have profit income. Household *i* owns the fraction α_{ij} of firm *j* so *i* will receive that fraction of *j*'s profit and hence will receive $\alpha_{ij} py_j$ in total from firm *j*. If we add this term over all firms *j* we get *i*'s total profit income, namely $\sum_{j \in J} \alpha_{ij} py_j$. Therefore, *i*'s total income is $p\omega_i + \sum_{j \in J} \alpha_{ij} py_j$, and hence *i*'s budget constraint is

$$px_i \leq p\omega_i + \sum_{j \in J} \alpha_{ij} py_j.$$

DEFINITION: General competitive equilibrium

The price system p and allocation (x, y) constitute a competitive equilibrium if for each individual i the consumption plan x_i maximizes U_i subject to i's budget constraint; for each firm j the production plan y_j belongs to T_j , the set of feasible production plans for firm j; and no member of T_j gives a higher profit than y_j given the price regime p. Finally, all markets clear, which means that

$$\Sigma_{i\in I} x_{ic} = \Sigma_{i\in I} \omega_{ic} + \Sigma_{j\in J} y_{jc}$$

holds for each commodity c. We refer to this last property as market clearance.

Before presenting an example of general equilibrium we identify a property of a simple family of utility functions that is used frequently.

Cobb-Douglas utility and positive consumption

Cobb-Douglas utility functions are of the form $U(a, b) = a^{\alpha}b^{\beta}$, where *a* is the consumption of the first good, *b* is the consumption of the second good, and α and β are positive constants. If the price of either good is zero then there will be unlimited demand for that good and demand will certainly exceed supply. Therefore, both prices will be positive at equilibrium. It follows that, unless the individual's endowment is zero, he or she will have a positive income and hence can afford a positive amount of each good. Consequently, utility will be positive. But if a = 0 or b = 0 then utility is zero. (With Cobb-Douglas preferences, U(0, b) = 0 = U(a, 0).) That can't be a utility-maximizing strategy. Therefore, the consumer will demand a positive amount of each good at equilibrium. It follows that the consumer's marginal rate of substitution will equal the price ratio at the bundle that the consumer demands.

We exploit this property (equality of the MRS and the price ratio), without explicitly invoking it, every time we use a Cobb-Douglas utility function.

∂ Example 2.4: Competitive equilibrium in a simple economy with productuion

The setup is identical to that of Example 2.3: There are two commodities, one individual, and one firm. Individual 1's preference is represented by the utility function $U_1 = x_{11} \times x_{12}$, and 1's endowment is $\omega_1 = (48, 0)$. Individual 1 owns the firm, so it follows that $\alpha_{ij} = 1$. The firm's production technology is given by "output = $2\sqrt{\text{input.}}$ " We show that $p_1 = 1$ and $p_2 = 4$ are competitive equilibrium prices. The production plan $y_1 = (-16, 8)$ results in a profit of $1 \times -16 + 4 \times 8 = 16$. To confirm that (-16, 8) is profit maximizing given the

2. The Arrow-Debreu Economy

prices, let *q* denote the amount of input employed. Then the firm's output of commodity 2 is $2\sqrt{q}$ and hence the firm's profit is

$$4 \times 2\sqrt{q} - 1 \times q$$

and the first derivative of profit is

$$\frac{4}{\sqrt{q}} - 1.$$

(Note that the second derivative is negative.) When we set the first derivative equal to 0 and solve for q we get q = 16. Then no other feasible production plan gives a higher profit than (-16, 8).

The budget constraint is $x_{11} + 4x_{12} \le 1 \times 48 + 16$. (The right-hand side is the value of the individual's endowment plus profit.) The consumption plan $x_1 = (32, 8)$ maximizes U_1 subject to the budget constraint. To confirm this we set

$$MRS = \frac{x_{12}}{x_{11}} = \frac{p_1}{p_2} = \frac{1}{4}$$

Then $x_{11} = 4x_{12}$ and hence the budget equation $x_{11} + 4x_{12} = 64$ reduces to $2x_{11} = 64$, and thus the chosen consumption plan has $x_{11} = 32$. Because $x_{11} = 4x_{12}$ we have $x_{12} = 8$. (If you don't know calculus, you can take my word that the MRS is the amount of the second good divided by the amount of the first good. To use calculus to derive the MRS, begin with the utility function U = ab. Hold utility constant at ℓ . The equation of the associated indifference curve is $ab = \ell$, and we can solve this for *b*, yielding $b = \ell a^{-1}$. The derivative db/da is $-\ell a^{-2}$. Because $\ell = ab$ we can write

$$\frac{db}{da} = -\ell a^{-2} = -(ab)a^{-2} = -\frac{b}{a}.$$

The MRS is the negative of this derivative. That is, the MRS is the negative of the slope of the indifference curve. The derivative can also be computed by applying the implicit function theorem to $u(x, y) = \ell$.)

The demand for commodity 1 is $x_{11} = 32$, and the supply is $\omega_{11} + y_{11} = 48 - 16$. The demand for commodity 2 is $x_{12} = 8$, and the supply is $\omega_{12} + y_{13} = 0 + 8$. Therefore, both markets clear.

2.2 Welfare theorem for an exchange economy

We prove that every competitive equilibrium of the *n*-person exchange economy is weakly efficient. That is, there is no feasible allocation that gives everyone more utility than the equilibrium. An important implicit assumption is that for each individual *i* the utility function U_i depends only on *i*'s consumption plan $x_i = (x_{i1}, x_{i2}, ..., x_{i\ell})$ and hence is independent of any other individual's consumption of any other commodity. A concluding remark shows that we can actually claim that the equilibrium is efficient, not just weakly efficient.

A competitive equilibrium of the *n*-person exchange economy is efficient.

Proof

We begin by showing that every allocation that gives everyone more utility than the equilibrium violates the material feasibility condition $x_1 + x_2 + \cdots + x_n \le \omega_1 + \omega_2 + \cdots + \omega_n$. Let the price system *p* and the allocation *x* constitute a competitive equilibrium. Suppose that the allocation *z* satisfies $U_i(z_i) > U_i(x_i)$ for each individual *i*. Because x_i maximizes U_i subject to *i*'s budget constraint

$$p_1 x_{i1} + p_2 x_{i2} + \dots + p_\ell x_{i\ell} \le p_1 \omega_{i1} + p_2 \omega_{i2} + \dots + p_\ell \omega_{i\ell}$$

the consumption plan z_i was not affordable when x_i was chosen. That is,

$$p_1 z_{i1} + p_2 z_{i2} + \dots + p_\ell z_{i\ell} > p_1 \omega_{i1} + p_2 \omega_{i2} + \dots + p_\ell \omega_{i\ell}$$

for each individual *i*. It follows that

$$p_{1}(z_{11} + z_{21} + \dots + z_{n1}) + p_{2}(z_{12} + z_{22} + \dots + z_{n2})$$

+ \dots + p_{\ell}(z_{1\ell} + z_{2\ell} + \dots + z_{n\ell}) > p_{1}(\omega_{11} + \omega_{21} + \dots + \omega_{n1}))
+ p_{2}(\omega_{12} + \omega_{22} + \dots + \omega_{n2}) + \dots + p_{\ell}(\omega_{1\ell} + \omega_{2\ell} + \dots + \omega_{n\ell}).

In words, the total market value of consumption exceeds the total market value of the commodities available for consumption. That strict inequality is inconsistent with the following material feasibility requirement:

$$z_{1c} + z_{2c} + \cdots + z_{nc} \leq \omega_{1c} + \omega_{2c} + \cdots + \omega_{nc},$$

which must hold for each commodity c. (If the total consumption of commodity c is less than or equal to the total endowment of c, then the market value of the amount of consumption of commodity c provided by z must be less than or equal to the market value of the total endowed amount of c. In that case, the total market value of consumption at z, over all commodities, cannot exceed the total market value of all endowments.) We have proved that the competitive equilibrium allocation is weakly efficient: There is no *feasible* allocation that gives everyone more utility than the competitive equilibrium.

Suppose that we had a feasible allocation z that gave *one* individual i more utility than the equilibrium allocation x and gave everyone else at least as much utility as x. Then we can construct a feasible allocation z' that gives everyone strictly more utility than x, contradicting what we have just proved. To construct z' from z we merely take a small amount of some commodity away from individual i, but we confiscate a sufficiently small amount so that i still has more utility than at x. Then we divide the amount taken from i among the remaining individuals. The resulting allocation z' will give each more utility than at z, and hence strictly more utility than at x. And z' is feasible because it is constructed from z simply by redistributing a little bit of one of the commodities. Because we have already proved that there is no feasible allocation that makes everyone better off than he is at the competitive equilibrium we have demonstrated that the competitive equilibrium is in fact efficient, not just weakly efficient.

The previous paragraph incorporates two implicit assumptions: First, individual *i*'s preference ordering is continuous: If *i* strictly prefers z_i to x_i and we reduce z_i by some sufficiently small amount the resulting consumption plan will

2. The Arrow-Debreu Economy

still be strictly preferred to x_i . Second, any individual's utility will increase if that person's consumption of any commodity increases, however small the increase.

Note that our proof is valid even when one or more consumers is at a corner of his or her budget line, with marginal rates of substitution unequal to the equilibrium price ratio, and even when marginal rates of substitution are not defined. The proof does not even depend on the representation of individual preference by a utility function: If individual *i* prefers z_i to x_i , and x_i was chosen at equilibrium, then z_i must have been too expensive. The rest of the proof follows without modification. It is a *very* general argument. The key assumptions are that agents take prices as given and that preferences are *self-regarding*—each individual cares only about his or her own consumption.

2.3 The welfare theorem in the general model

Now we add production to the model and show that a competitive equilibrium of the Arrow-Debreu economy is efficient, provided that every commodity that affects individual utility is traded in a competitive market—the *completeness of markets* assumption. Completeness of markets means that for every possible realization of every random event there is a market in which one can purchase or sell a unit of any good contingent on that realization. Complete markets also require that for every future date there is a market in which anyone can purchase or sell a unit of any good for delivery at that date. It also entails the assumption that if individual or firm *K*'s actions affect the welfare of other agents—think of pollution—then there is a market that causes this side effect to be brought to bear on *K*'s decision making via *K*'s budget constraint (if *K* is a consumer) or *K*'s profit (if *K* is a firm).

DEFINITION: Complete set of markets

Every commodity that affects some individual's welfare is traded in a competitive market.

A complete set of markets would be astronomical in number. We never have anything close to completeness in the real world. Nevertheless, the Arrow-Debreu economy with complete markets is a valuable framework within which to study resource allocation. It is also an important benchmark case. For one thing, it helps us identify what has gone wrong when the economy is not efficient.

If markets are complete then a competitive equilibrium of the Arrow-Debreu economy is efficient.

Proof

We begin by showing that if an allocation gives everyone strictly more utility than the competitive equilibrium then it cannot be feasible. Let (p, x, y) be the competitive equilibrium. Suppose that the allocation (x', y') gives everyone

more utility. Then $U_i(x'_i) > U_i(x_i)$ for each consumer *i*. But x_i maximizes U_i subject to the budget constraint. This means that x'_i was not affordable when x_i was chosen. Therefore,

$$px_i' > p\omega_i + \sum_{j \in J} \alpha_{ij} py_j$$

This is true for each *i*, so it will remain true when we sum up over all individuals. That is,

$$\sum_{i\in I} px'_i > \sum_{i\in I} p\omega_i + \sum_{i\in I} \sum_{j\in J} \alpha_{ij} py_j.$$

Now $\sum_{i \in I} \sum_{j \in J} \alpha_{ij} py_j$ is equal to $\sum_{j \in J} py_j \sum_{i \in I} \alpha_{ij}$. But for any firm *j* the sum $\sum_{i \in I} \alpha_{ij}$ is the sum of all the ownership shares in firm *j* and that total must equal 1. Therefore, $\sum_{i \in I} \sum_{j \in J} \alpha_{ij} py_j$ equals $\sum_{j \in J} py_j$, which is the total profit in the economy at equilibrium. Therefore, we can state

$$\Sigma_{i\in I} p x_i' > \Sigma_{i\in I} p \omega_i + \Sigma_{j\in J} p y_j.$$
^[1]

If y'_j does not belong to T_j for some firm j then the allocation (x', y') is not feasible. If each y'_j does belong to T_j , then because y_j maximizes j's profit given p we must have $py_j \ge py'_j$ for each firm j. Therefore, $\sum_{j \in J} py_j \ge \sum_{j \in J} py'_j$. This inequality and [1] gives us

$$\sum_{i \in I} p x_i' > \sum_{i \in I} p \omega_i + \sum_{j \in J} p y_j'.$$
[2]

However, [2] is inconsistent with material feasibility of (x', y'): If we did have $\sum_{i \in I} x'_{ic} \leq \sum_{i \in I} \omega_{ic} + \sum_{j \in J} y'_{jc}$ for each commodity *c*, then by virtue of the fact that $p_c \geq 0$ we would have

$$p_c \Sigma_{i \in I} x'_{ic} \le p_c \Sigma_{i \in I} \omega_{ic} + p_c \Sigma_{j \in J} y'_{jc}$$

$$[3]$$

for each c. When we add each side of [3] over all commodities we get

$$\sum_{i \in I} px'_i \le \sum_{i \in I} p\omega_i + \sum_{j \in J} py'_j, \tag{4}$$

a direct contradiction of [2]. (Statements [3] and [4] are equivalent because $p\sum_{i\in I} x'_i$ is equal to $p_1\sum_{i\in I} x'_{i1} + p_2\sum_{i\in I} x'_{i2} + \cdots + p_\ell\sum_{i\in I} x'_{i_\ell}$, and the other terms of [2] can be similarly expressed.) Therefore, there can be no feasible allocation that gives everyone more utility than the competitive equilibrium.

We have shown that any allocation that gives everyone more utility than the competitive equilibrium must violate one of the feasibility conditions. In other words, there is no feasible allocation that would give everyone more utility than the market equilibrium. If there were a feasible allocation (x', y') that gave *one* individual *i* more utility than the equilibrium and gave no one any less utility, then we could construct a feasible allocation (x'', y'') from (x', y') by taking a sufficiently small amount of some commodity away from individual *i* in a way that still leaves person *i* strictly better off than at the equilibrium. If we divide the amount confiscated from *i* among the remaining individuals we will have made everyone better off than at the equilibrium, contradicting what we have just proved. Therefore, the competitive equilibrium is in fact efficient, not

2. The Arrow-Debreu Economy

just weakly efficient. (The second-to-last paragraph of Section 2.2 discloses the implicit assumptions that make this argument work.)

To complete the demonstration that the Arrow-Debreu economy performs efficiently we must prove that an equilibrium actually exists. Example 3.1 of Section 3 shows that existence cannot be taken for granted. However, if individual preferences exhibit diminishing MRS and production processes have diminishing marginal products, then a competitive equilibrium will exist. That is, *if* agents take prices as given, there will be an equilibrium and it will be efficient—assuming complete markets. The existence question is explored briefly in Section 3 and the assumption of price-taking behavior is the subject of Section 4. We conclude this section by showing what goes wrong when markets are not complete.

2.4 Externalities

A competitive equilibrium of the private ownership market economy need not be efficient if one person's consumption directly affects the welfare of someone else, which we refer to as an *externality*.

DEFINITION: Externality

An externality is present when one agent's actions affect the welfare of another agent but the market system does not provide an incentive for the former to include the effect on the latter in the former's decision making.

Let's see why externalities undermine the efficient operation of competitive markets.

Example 2.5: A two-person exchange economy with externalities

The endowments are $\omega_1 = (0, 16)$ and $\omega_2 = (16, 0)$, with

$$U_1 = x_{11} \times x_{12} - 3x_{21}$$
 and $U_2 = x_{21} \times x_{22}$.

Person 1's utility is adversely affected by the other person's consumption of the first good. Person 1 cannot control the choice of person 2 so when 1 maximizes utility he or she must take x_{21} as given. That is, x_{21} will be treated as a constant in person 1's decision making. Therefore, to find the chosen consumption plans we can employ the utility functions $U_1 = x_{11} \times x_{12}$ and $U_2 = x_{21} \times x_{22}$. The MRSs are x_{12}/x_{11} for person 1 and x_{22}/x_{21} for person 2. Each consumer will maximize utility by setting the MRS equal to the price ratio. (The MRS is derived at the end of Example 2.4. Note that individual utility would be 0 if the individual consumed 0 units of one of the goods.) Therefore we have

$$\frac{x_{12}}{x_{11}} = \frac{p_1}{p_2} = \frac{x_{22}}{x_{21}}$$

At a competitive equilibrium we have demand equaling supply, and thus

$$x_{11} + x_{21} = 16$$
 and $x_{12} + x_{22} = 16$.

These equations and $x_{12}/x_{11} = x_{22}/x_{21}$ give us

$$\frac{x_{12}}{x_{11}} = \frac{16 - x_{12}}{16 - x_{11}}$$

Cross multiplying yields $x_{11} = x_{12}$. Because $x_{12}/x_{11} = p_1/p_2$ we have $p_1/p_2 = 1$, or $p_1 = p_2$. Consumer 1's budget constraint is $p_1x_{11} + p_2x_{12} = 16p_2$ because 1's endowment consists of 16 units of good 2. Because $p_1 = p_2$ and $x_{11} = x_{12}$ the budget constraint reduces to $2p_1x_{11} = 16p_1$, the solution of which is $x_{11} = 8$. Then $x_{11} = 8 = x_{12}$. Because market clearance requires $x_{11} + x_{21} = 16$ and $x_{12} + x_{22} = 16$ we also have $x_{21} = 8 = x_{22}$. At equilibrium each individual's consumption plan is (8, 8), but $U_1 = 40$ while $U_2 = 64$.

Is the equilibrium of Example 2.5 efficient? Because 1's utility increases when 2's consumption of the first good falls, it should be possible to make both people better off than they are at the equilibrium by transferring some of the first good from person 2 to person 1 and some of the second good from person 1 to person 2. Specifically, transfer 0.5 units of the first good from person 2 to 1, and 0.6 units of the second good from person 1 to 2. The new utility levels will be

$$U_1 = 8.5 \times 7.4 - 3 \times 7.5 = 40.4$$
 and $U_2 = 7.5 \times 8.6 = 64.5$.

Both individuals have a higher level of utility than at equilibrium, proving that the equilibrium allocation was not even weakly efficient.

When there are consumption externalities the proof of the first welfare theorem breaks down right at the start. If U_i depends on the consumption of other individuals as well as on *i*'s consumption then it is no longer true that if *i* prefers the consumption plan *b'* to the equilibrium then the market value of *b'* must exceed *i*'s income. It may be changes in the consumption of others, changes that are beyond *i*'s control, that make *b'* superior to *x* in *i*'s estimation. For example at the equilibrium prices $p_1 = p_2$ the basket (8.5, 7.4) is *cheaper* for person 1 than the equilibrium basket (8, 8). But the latter was the best that person 1 could do given his or her budget constraint *and* given the choice (8, 8) of person 2. But 1 prefers (8.4, 7.4) to the equilibrium in part because of the changes in the other person's consumption.

In general, negative externalities are created when the decision of a consumer or firm imposes costs on society that are not costs to the decision maker. In most cases the market system uses prices to transmit information about social costs and benefits. But when an agent's decision imposes a cost on society that is not incorporated in the price that the agent pays, then vital information, necessary for efficiency, is not transmitted to the decision-making agent. Even if the agent receives the information from other sources, if it does not reduce

2. The Arrow-Debreu Economy

his or her spending power then the agent has no *incentive* to take those costs into consideration.

Consider what happens when social costs *are* incorporated into the price: Labor accounts for 70% to 75% of all costs of production in mature capitalist

Microevolution of antibiotic-resistant microbes has resulted from the widespread administration of antibiotics to humans. (We also indirectly consume the antibiotics that are routinely given to livestock and used in commercial food preparation.) When a physician prescribes an antibiotic the physician benefits the patient but also does an imperceptible amount of harm to the rest of the population. The total harm done, by all physicians, is of enormous significance. The excessive prescribing of antibiotics in wealthy nations has created a giant experiment with a vast number of opportunities for microbes to mutate. And they have done so with great success. The World Health Organization has warned that the high level of resistance to drugs that were once extremely effective in treating common infectious diseases will soon precipitate a global crisis. "Superbugs" could render our antibiotics useless by the year 2010 (Kmietowicz, 2000).

economies. The labor used by a firm is clearly a cost to society-if it were not used by the firm it could be productively employed elsewhere. The firm using the labor has to pay wages that are a function of its workers' potential contribution to production in general-a consequence of all firms bidding for the use of productive factors. That gives the firm incentive to economize on the use of labor. This contributes to efficiency. More strikingly, the firm has a strong incentive to reduce the size of its wage bill (more than 70% of total cost) by inventing labor-saving equipment. If capital equipment lowers the labor requirement per unit of output, it raises the total output of a given labor force. And of course, output per worker, and hence consumption per worker, has risen dramatically over the decades and centuries. And all because the firm's use of labor is a cost to society that is brought to bear on the firm's decision by means of a price-the wage rate.

If the costs of pollution could be incorporated into the prices that confront firms and consumers, the same powerful force for innovation would be unleashed. Each household would strive to avoid the social costs of its

polluting: by altering activities to reduce the amount of waste discharged into the water and air and by purchasing products that lower total social cost by incorporating pollution reduction technology. Knowing this, firms would have a strong incentive to invent pollution abatement technology and to invest in that technology when it became available. And to the extent that the activities of firms result in waste being discharged into the water and air, if this cost to society were converted to a cost paid by the polluting firms there would be a strong incentive for firms to avoid these costs by investing in technology to reduce waste by-products.

The presence of significant *positive* externalities also results in an inefficient market equilibrium. For example, when someone considers purchasing fireworks for a private independence day celebration that person doesn't take into consideration the benefit that his or her neighbors will derive from the display. But that spillover benefit should clearly be counted—by someone—as a benefit to society. But the decision makers consider only the benefit to themselves.

∂ Example 2.6: One private good and one public good

Commodity 1 is a *pure public good*—any amount of it provided for one person or group benefits everyone in the community. Commodity 2 is a conventional private good that can be consumed directly or used to produce the public good. Each unit of the public good requires 1 unit of the private good as input. If we let the price of the private good be unity then the price of the public good will also be unity in a competitive equilibrium, because the marginal cost of producing the public good is constant at 1. Therefore, the equilibrium price ratio is 1. Now, suppose that there are three consumers, each is endowed with 50 units of the private good and 0 units of the public good, and each has the utility function $6\sqrt{q} + t_i$, where q is the amount of the public good produced and t_i is the amount of the private good consumed by individual *i*. If the public good—for example, fireworks-were only available on the private market then each consumer would optimize by equating marginal benefit and the price ratio. Therefore, $3/\sqrt{q} = 1$ at equilibrium. We solve this for q = 9. The total amount of the public good purchased in the community is 9 units, and everyone benefits from each of the 9 units, whether individuals purchased much or little of the good. Suppose that each individual buys 3 units of the public good. Is this outcome efficient? Individual utility is $U_i = 6\sqrt{9} + 50 - 3 = 65$ for each *i*. However, if the community could somehow arrange for q = 81 and have each *i*'s consumption of the private good reduced by an additional $\frac{1}{3} \times 72 = 24$ to collect the input necessary to produce 72 additional units of the public good, then $U_i = 6\sqrt{81} + 50 - 27 = 77$, which is much higher for each individual than utility at the private market equilibrium. The market equilibrium is not efficient.

Sources

The foundations of modern general equilibrium theory were laid by Kenneth J. Arrow, Gerard Debreu, and Lionel McKenzie. See Arrow (1951b), Arrow and Debreu (1954), McKenzie (1954), and Debreu (1959).

Links

McKenzie (2002) provides a superb treatment of modern general equilibrium theory, from its inception in the middle of the twentieth century to the latest research. See Stiglitz (1993) for more on why markets can't even be approximately complete. For a fuller treatment of the technical side of the Arrow-Debreu model—existence of equilibrium, in particular—see Campbell (1987), especially pages pp. 39–47, Chapter 7, and Appendixes 2 and 3. To see why the completeness of markets assumption rules out externalities see Campbell (1987, pp. 56–60). Some economists claim that hidden information problems, particularly those related to contract enforcement, are so severe that the Arrow-Debreu model has nothing to teach us (see Gintis, 2000, pp. 46, 136, 140). This section does implicitly assume that contracts can be costlessly enforced. Even so, we learn that the private ownership market economy orchestrates production and consumption activities in a way that obviates the need for an astronomical number of welfare-improving trades.

2. The Arrow-Debreu Economy

Problem set

For the first six questions (*a*, *b*) denotes the bundle consumed by person 1 and (*x*, *y*) is the bundle consumed by person 2.

1. Compute the competitive equilibrium of the two-person, two-commodity exchange economy with utility functions

 $u_1(a, b) = a^2 b$ and $u_2(x, y) = xy^2$

and endowments (0, 3) for person 1 and (3, 0) for person 2.

2. The *second* theorem of welfare economics asserts that under certain modest conditions on preferences and technology sets, every efficient allocation is a competitive equilibrium allocation *for some redistribution of initial endowments—and profit shares, if the model includes production*. Verify the second welfare theorem for the following two-person, two-commodity exchange economy:

$$u_1(a, b) = ab$$
 $\omega_1 = (0, 1),$
 $u_2(x, y) = xy$ $\omega_2 = (1, 0).$

That is, identify all of the efficient allocations and then prove that each is a competitive equilibrium outcome for some distribution of the total endowment (1, 1).

- **3.** Consider the following two-person, two-commodity exchange economy: $u_1 = a + b$, $\omega_1 = (1, 1)$, $u_2 = 2x + y$, $\omega_2 = (1, 1)$.
 - **A.** Characterize the set of efficient allocations.
 - **B.** Show that the allocation that assigns (1, 2) to person 1 and (1, 0) to person 2 is efficient.
 - **C.** Show that the allocation of *B* is a competitive equilibrium outcome for some price system and some distribution of wealth.
- **4.** Consider the following simple two-person, two-commodity exchange economy:

$$u_1(a, b) = ab$$
 and $\omega_1 = (0, 1),$
 $u_2(x, y) = x + \ln y$ and $\omega_2 = (2, 2).$

(Recall that $\ln y$ is the function whose first derivative is y^{-1} .) Find the competitive equilibrium for this economy. Is the competitive equilibrium allocation efficient? Explain.

5. Solve the following two-person, two-commodity exchange economy for the competitive equilibrium:

$$u_1 = ab$$
 and $\omega_1 = (1, 1),$
 $u_2 = xy + b$ and $\omega_2 = (1, 0).$

Is the equilibrium outcome efficient? Explain briefly.

6. Consider the following simple two-person, two-commodity exchange economy:

$$u_1 = ab - x$$
 and $\omega_1 = (0, 1),$
 $u_2 = xy + a$ and $\omega_2 = (1, 0).$

- **A.** Explain carefully why $p_1 = 1 = p_2$ and $a = b = x = y = \frac{1}{2}$ define a competitive equilibrium.
- **B.** Is this competitive equilibrium allocation efficient? Explain your answer.
- **7.** Consider a simple model of an economy with two private goods and many consumers. Each individual has a utility function of the form $B_i(a_i) + b_i$, where a_i is *i*'s consumption of good 1 and b_i is *i*'s consumption of good 2. Let *P* denote the price of the second good. Choose units so that the price of the first good is \$1. Good 1 can either be consumed or used to produce good 2, and when one unit of commodity 1 is used as input exactly one-third of a unit of commodity 2 is obtained as output.
 - **A.** What is the equilibrium price of commodity 2 if the two goods are produced in a private ownership market economy with competitive firms?
 - **B.** Suppose now that the government decides to subsidize the consumption of the second good by paying \$1 for each unit consumed. That is, consumers pay P 1 for a unit of good 2 but firms receive P a unit. What will P be in this case ? Will the outcome be efficient? If so, explain why; if not, prove it with an numerical example.

3 NONCONVEX ECONOMIES

What can we say about economies in which individual preferences do not have the diminishing marginal rate of substitution property? The proof of the welfare theorem does not make any assumptions about the MRS. As long as there is a complete set of markets then a *competitive* equilibrium is efficient. *But will a competitive equilibrium exist*? Yes, if preferences have the diminishing MRS property. (Production functions also have to have a nonpositive second derivative, but we only look at exchange in the section.) The easiest way to uncover the difficulties that can arise without the diminishing MRS property is to examine a *one-person* exchange economy with two commodities.

Example 3.1: Nonexistence of equilibrium with increasing MRS

There are two commodities and one individual (and no production). The individual's consumption plan is a bundle (*a*, *b*), and the utility function is $U = a^2 + b^2$. The equation of an indifference curve is $a^2 + b^2 = \ell$, where ℓ is a given positive constant. This is the equation of a circle with center at the origin. Therefore,



Figure 10.1

the MRS increases as we slide down the indifference curve (Figure 10.1). The consumer is endowed with one unit of each good. Therefore, if demand equals supply in each market, the consumer will demand one unit of each good. It is easy to show that there is no price system at which this can happen.

Everything hinges on the fact that $(\alpha + \beta)^2 = \alpha^2 + 2\alpha\beta + \beta^2 > \alpha^2 + \beta^2$ if α and β are both positive. This has implications for demand. The consumer will want to spend all of his or her income on one of the goods. Accept the truth of this assertion for a moment. Then there can be no equilibrium at any price regime: The demand for one of the goods will be zero but there is a supply of one unit of that good—Figure 10.1. (And there will be excess demand for the commodity on which the consumer spends all of his or her income.) Therefore, our demonstration that there is no equilibrium will be complete once we show that all of the income will be spent on one good.

Suppose that $p_2 \ge p_1$. (The second good is at least as expensive as the first good.) Suppose also that b > 0. $U = a^2 + b^2$, but if the individual reduces b to zero p_2b dollars will be released that can be used to buy $\Delta a = p_2b/p_1$ additional units of good 1. Because $p_2 \ge p_1$, we have $\Delta a \ge b$, which means that the new level of consumption of the first good is at least a + b. Finally,

$$U(a + \Delta a, 0) \ge (a + b)^2 = a^2 + b^2 + 2ab > a^2 + b^2 = U(a, b).$$

Utility has increased. Similarly, if $p_1 > p_2$ and a > 0 then the consumer can increase utility by reducing expenditure on the first good to zero and using the money thereby released to buy more of the second good. Therefore, the consumer will spend all of the money on the cheaper good, and if the prices

are equal the consumer will not care which good he or she buys but will prefer either extreme to any affordable consumption plan with a positive amount of each good. The consumer will never demand a positive amount of *both* goods, but the supply of each is positive. Therefore, there can be no price system at which demand equals supply.

The problem revealed by Example 3.1 is a lot worse than the lack of existence of an equilibrium: There is no price system at which supply and demand are *approximately* equal in each market: If $p_1 < p_2$ then the consumer will spend all income on the first good. The demand for the second good will then be zero, but the supply is 1. The consumer's income is $p_1 + p_2$, so the consumer will buy $(p_1 + p_2)/p_1$ units of good 1. Therefore, the demand for the second good will exceed 2 because $p_2/p_1 > 1$. Of course, if $p_2 < p_1$ then the individual will demand 0 units of good 1 and more than 2 units of good 2. (If $p_1 = p_2$ the consumer will get 2 units of one good and 0 units of the other.) For any price system, the demand for the other good will be zero, which is well below the supply, and the demand for the other good will be at least double the supply.

It is remarkable that we can exhibit a failure of markets to clear at any price regime, even in an approximate sense, by means of a simple economy having only one consumer and two goods. It is easy to extend this example to one that has many commodities, and the reader is invited to do so. But how special is the fact that there is only one consumer? This is far from realistic. Is it essential for the nonexistence of equilibrium?

Example 3.2: Two consumers, each with increasing MRS

This exchange economy consists of two consumers, each identical to the individual of Example 3.1. If $p_1 = p_2$ then the individual's income is $p_1 + p_2 = 2p_1$. (The individual's endowment consists of 1 unit of each good.) The consumers will spend all of that income on one of the goods and will receive $2p_1/p_1 = 2$ units of that commodity, whichever it is. Then the utility of the bundle that maximizes utility subject to the budget constraint is $4 = 2^2 + 0 = 0 + 2^2$. The individuals will be indifferent between the affordable bundles (2, 0) and (0, 2). If one of the consumers demands (2, 0) and the other demands (0, 2) then the total demand for each good will be 2. The total supply of each good is also 2, because each individual is endowed with 1 unit of each good. We have a competitive equilibrium, because each individual is maximizing utility subject to the budget constraint and each market clears.

For any even number of consumers, each identical to the individual of Example 3.1, there will be a competitive equilibrium with $p_1 = p_2$. If half of the consumers spend all of their income on the first good and the other half spend all of their income on the demand will equal supply for each good.

3. Nonconvex Economies

Table	10	.1
-------	----	----

Number who buy only the first good	Demand for the first good	Demand for the second good
0	0	6
1	2	4
2	4	2
3	6	0

However, there can be no exact equilibrium with an odd number of consumers. Consider first the case of a three-person economy.

Example 3.3: Three identical consumers, each with increasing MRS

Each of the three individuals are identical to the one of Example 3.1. If $p_1 > p_2$ or $p_1 < p_2$ then all individuals will spend all of their income on the cheaper good. The demand for the other good will be 0, but its supply will be 3. The demand for the cheaper good will be greater than 6, but the supply is only 3. (For instance, if $p_1 = 2$ and $p_2 = 3$ then income is 5, so each person will demand $\sqrt[5]{2} = 2.5$ units of good 1. The demand for good 1 is $3 \times 2.5 = 7.5$.) If $p_1 = p_2$ then a consumer will either buy 2 units of good 1 and 0 units of good 2, or vice versa. Table 10.1 shows that the closest we can get to market clearance is for one consumer to spend all income on one of the goods and the other two to spend all their income on the other good. We will get an excess supply of the former of 1 and an excess demand for the latter of 1 unit. The excess demand is a smaller fraction of total demand than in the case of the one-person economy of Example 3.1. Also, excess supply is a smaller fraction of total supply than in the one-person economy, but there is still a significant gap between demand and supply.

With an odd number *n* of identical consumers, we cannot have exact market clearance with increasing MRS preferences. Here's why: Let *t* be the number of consumers who spend all of their income on good 1, and let *s* be the number who spend all of their income on good 2. If $p_1 < p_2$ then t = n and s = 0, so we are far from market clearance. Similarly, if $p_1 > p_2$ then t = 0 and s = n. If $p_1 = p_2$ then *t* and *s* can both be positive. In fact, each individual is indifferent between (2, 0) and (0, 2) so *t* and *s* can be any two integers such that t + s = n. Because *n* is odd, we cannot have t = s. However, if demand equals supply for the first good we will have

$$n =$$
supply $=$ demand $= 2t$

and thus $t = \frac{1}{2}n$, which is impossible because *n* is odd—we can't have half a consumer. Therefore, demand cannot be exactly equal to supply.

Та	bl	е	1	0	.2

Number who buy only the first good	Demand for the first good	Demand for the second good
$t + \frac{1}{2} t - \frac{1}{2}$	$2t+1\\2t-1$	2t - 1 $2t + 1$

Example 3.4: An odd but large number of consumers

Each of the *n* individuals is identical to the one of Example 3.1. Let $t = \frac{1}{2}n$. Then both $t + \frac{1}{2}$ and $t - \frac{1}{2}$ are integers. If $p_1 = p_2$ we can have $t + \frac{1}{2}$ individuals each spend all of their income on one of the goods and $t - \frac{1}{2}$ individuals each spend all of their income on the other good (Table 10.2). The supply of each good is n = 2t. Therefore, the closest we can get to market clearance involves demand exceeding supply for one of the goods by 1 and supply exceeding demand by 1 unit for the other good. When the number *n* of individuals is large, the supply *n* of each good is large, and hence a difference between demand and supply of 1 unit is trivial in the case of a large number of consumers. We have an approximate equilibrium which, for practical purposes, is sufficiently close to market clearance.

If the number of traders is large then we are guaranteed a general equilibrium in a practical sense, if not in an exact sense, regardless of the nature of individual preferences.

Links

Anderson, Kahn, and Rashid (1982) show that under very general conditions there will be an approximate equilibrium of an *n*-person exchange economy, and that the percent by which demand is less than supply or is greater than supply approaches zero as *n* increases without limit. Even for modestly large *n*, the gap between demand and supply is virtually zero (as a fraction of total supply).

Problem set

- **1.** Consider a one-person exchange economy with two goods, *A* and *B*. Let $U_1 = a^2 + b$ and $\omega_1 = (1, 1)$ be the utility function and endowment, respectively. Prove that this economy does not have a competitive equilibrium (Hint: Write down the budget equation and then use it to solve for *b* as a function of *a*. Substitute this expression for *b* in the utility function, and then draw the graph of utility as a function of *a*.)
- **2.** For each of the following one-person exchange economies determine the efficient allocation (or allocations) and see if there is *any* price

4. Efficiency and Incentive Compatibility

system and income level at which the consumer will choose an efficient outcome.

A.
$$U_1 = 2a^2 + b^2$$
 and $\omega_1 = (1, 1)$.

B. $U_1 = a^2 + b^2$ and $\omega_1 = (1, 2)$.

3. For each of the three economies of the first two questions, demonstrate that there is an approximate competitive equilibrium when there is a large number of individuals identical to person 1.

4 EFFICIENCY AND INCENTIVE COMPATIBILITY

In showing that a complete set of markets results in a competitive equilibrium of the private ownership market economy that is efficient, we assumed that each individual and firm takes the prices as given. That means that each individual or firm calculates a best response to the given prices, without considering how a different response might affect prices, perhaps in a direction favorable to the decision maker. In this section we consider the appropriateness of the price taking assumption. In Section 4.4 we establish that it is rational for each individual to take market prices as given *if* there is a large number of suppliers of each good.

In real-world market economies there are many goods that are produced by only a handful of firms, and in such cases the individual producer can profit by exploiting the fact that a change in its supply will have an effect on prices. This is easy to understand, but we nevertheless work through a specific example in Section 4.1, in part to clarify what we mean by price taking on the one hand and exploiting one's market power on the other hand.

As we see later, we can view a departure from price taking behavior as a misrepresentation of the agent's hidden characteristic—preference in the case of a consumer and the production technology in the case of a firm. Because it is possible for a supplier to profit from misrepresentation in the private ownership market economy when there are few rival producers, we are led to ask if there are alternative economic systems that are invulnerable to misrepresentation. Sections 4.2 and 4.3 are devoted to this question, which is also addressed at the end of Section 4.1.

The sensible approach would seem to be to put the spotlight on the individual firm and trace the effects of a misrepresentation of its production technology back to individual consumer welfare. For instance, in the case of the Arrow-Debreu economy we would chart how a firm's misrepresentation affected its profit, then how the incomes of the firm's shareholders were affected, and finally how that impacted shareholder utility. That makes our task much more complicated, however, so we confine attention to pure exchange. The suppliers, then, are individual households, and the supplies are the individual endowments. This gives us a much more direct connection between agents' misrepresentations— of preference schemes—and individual utility.

4.1 Dominant strategy equilibrium

Consumers' hidden (private) characteristics are their preference schemes. What does misrepresentation of preference mean in the context of the market system? Consumers are never asked to report their utility functions. They *are* asked to submit a list of demands that maximize individual utility, given the current prices. If a consumer misrepresents his or her preferences by demanding a basket of goods that is not utility maximizing, there will be no way for any overseer of the market system to detect this misrepresentation.

We can view the market system as a mechanism in which each consumer is asked to submit his or her utility function to a referee, who then computes the resulting demand functions. The individual demand functions are then fed into a computer, which then calculates the equilibrium configuration of prices. The demand functions are used once again to determine individual consumption of each good at the equilibrium prices, and the resulting baskets of goods and services are delivered to the consumers. The question we are addressing can be stated this way: Would it ever be to anyone's advantage to submit a false utility function to the referee? We are about to see that the answer is yes. An appropriately chosen false utility function will induce demands that cause prices to change in a way that leaves the individual with more utility—according to her true utility function—than she would have realized under truthful revelation. The consumer will have benefitted by deviating from the rules of the game, and no one will know that she has deviated.

Can an individual consumer in fact influence prices by altering his or her demands? The intuition is that this power is negligible when there is a realistically large number of consumers, as Section 4.4 confirms. This section shows how misrepresentation can be profitable when the number of consumers is small. This will prepare us for the large numbers case by clarifying what is meant by misrepresentation of preference.

Example 4.1: A two-person, two-commodity exchange economy

Person 1's endowment is $\omega_1 = (0, 1)$ and 2's endowment is $\omega_2 = (1, 0)$. The utility functions are

$$U_1 = ab$$
 and $U_2 = xy$,

assuming that a typical allocation assigns the consumption plan $\theta_1 = (a, b)$ to person 1 and the plan $\theta_2 = (x, y)$ to person 2. The competitive equilibrium is easy to compute: Individual 1's MRS is b/a and individual 2's MRS is y/x. (See the end of Example 2.4 for the derivation of the MRS.)

At equilibrium each individual's MRS is equal to the price ratio. Therefore, b/a equals the price ratio, which equals y/x. Therefore, b/a = y/x. But a + x = 1 and b + y = 1 at equilibrium (market clearance). Therefore,

$$\frac{b}{a} = \frac{1-b}{1-a},$$

and if we cross multiply we get a - ab = b - ab, which implies a = b. Now, (a, b) and the endowment point (0, 1) are both on 1's budget line, so the slope of the

4. Efficiency and Incentive Compatibility

line is (b-1)/(a-0) and this is the negative of the price ratio, which will equal person 1's MRS at equilibrium. Therefore,

$$\frac{b}{a} = \frac{1-b}{a},$$

which implies $b = \frac{1}{2}$, and therefore $a = \frac{1}{2}$ because a = b. Therefore, $x = y = \frac{1}{2}$. The equilibrium price ratio is equal to each MRS at equilibrium, and b/a = 1 so the price ratio equals 1 at equilibrium. That is $p_1 = p_2$. For convenience, set each price equal to unity. We have found the competitive equilibrium: The prices are equal and each individual's consumption plan is $\binom{1}{2}$, $\binom{1}{2}$. Note that each individual's income at equilibrium is $p_1 = p_2$.

DEFINITION: Price taking behavior

An individual is a price taker if at each price regime he or she demands the consumption plan that maximizes utility subject to the budget constraint, without taking into consideration how different demands might have affected prices.

By definition of a *competitive* equilibrium, each person takes the price regime as given. That is, $(\frac{1}{2}, \frac{1}{2})$ is the unique utility-maximizing consumption plan of all those plans (*a*, *b*) satisfying the budget constraint 1a + 1b = 1. But each individual has a monopoly in the supply of one of the goods, and each individual supplies half of his or her endowment at the competitive equilibrium. Individuals can be expected to know that the price of the goods that they supply will increase if they restrict the supply. That is, the individuals will surely not behave as price takers in this economy—our price taking assumption is unfounded. We offer a proof by contradiction. The argument makes use of Figure 10.2.

Figure 10.2 represents person 2's utility in terms of person 1's consumption (a, b). This means that an indifference curve for person 2 is the set of plans (a, b) such that $U_2(1 - a, 1 - b)$ is constant. In the present case, $U_2 = xy$, so an indifference curve for person 2 is the set of plans (a, b) such that $(1 - a) \times (1 - b)$ is constant.

Suppose that person 2 always acts as a price taker. This means that 2's demand vector (*x*, *y*) will always be the one that maximizes U_2 subject to the budget constraint $p_1x + p_2y = p_1$. (Person 2's income is p_1 because 2 is endowed with 1 unit of the first good.) Let's see if person 1 can profit from preference misrepresentation. The intuition is simple. If person 1 demands the vector v_1 and at that point 1 demands more of good 2, the good that 1 supplies, than the straightforward utility-maximization exercise would predict, then the price of good 2 will be kept high. This will be utility maximizing in a more sophisticated sense. Let's see why.

Suppose that $p_1 = 1$ and $p_2 = 2.25$, and person 2 demands $v_2 = (1, 1) - v_1$, where v_1 is given in Figure 10.2. (The budget line is the flatter of the two lines in the diagram.) Both markets clear because $v_1 + v_2 = (1, 1) = \omega_1 + \omega_2$. Note, however, that at the competitive demand vector θ_1 person 1 would be on a



Figure 10.2

higher indifference curve than the one through v_1 . *If person 1 could be sure of having the order* θ_1 *filled* then ordering θ_1 would be the better strategy, given the prices $p_1 = 1$ and $p_2 = 2.25$. But person 1 can be sure that the order θ_1 *won't* be filled. Markets clear when person 1 undersupplies the second good by ordering v_1 for him- or herself and because θ_1 specifies a lower demand (i.e., higher supply) for the second good, we will have excess supply of the second good when person 1 orders θ_1 and person 2 orders v_2 . The price of the second good will have to fall, and this means that individual 1 can't have θ_1 after all.

The *competitive* equilibrium allocation gives each person the basket $\binom{1}{2}, \binom{1}{2}$; Example 4.1. The associated equilibrium prices are $p_1 = 1 = p_2$. At a higher price for good 2, say $p_2 = 2.25$ person 2 will demand less of good 2. Person 2's demand under price system p = (1, 2.25) is v_2 , which is identified by the point v_1 in Figure 10.2: v_1 would be what is left over for person 1 if person 2 were given the basket v_2 . If person 1 were to demand the vector that maximizes U_1 given the price vector p = (1, 2.25) then 1 would demand θ_1 as shown in Figure 10.2. As we have seen, markets won't clear when person 1's consumption plan is θ_1 and person 2's plan is v_2 . If person 1 continued to behave competitively that is, take prices as given—then the economy would wind up back at the competitive equilibrium with each person receiving $\binom{1}{2}$, $\binom{1}{2}$. But if person 1 were to demand v_1 when $p_1 = 1$ and $p_2 = 2.25$ and person 2 demands v_2 then both markets would clear. That is because v_1 is defined as the total supply minus v_2 . The economy would be in equilibrium because there would be no tendency for prices to change: Demand and supply would be equal in both markets. And because markets clear, all consumers would be able to carry out their plans. Moreover, v_1 gives individual 1 more utility than $(\frac{1}{2}, \frac{1}{2})$ according to 1's true preferences! The next example makes this concrete by working out v_2 and θ_1 .

Example 4.2: Misrepresentation in the economy of Example 4.1

To obtain v_2 we want to maximize $U_2 = xy$ subject to 1x + 2.25y = 1, because person 2 has 1 unit of good 1 to sell at a price of \$1. Clearly, utility maximization implies

MRS =
$$\frac{y}{x} = \frac{p_1}{p_2} = \frac{1}{2.25} = \frac{4}{9}$$
.

Then y/x = 4/9, and hence 9y = 4x. We also have 1x + 2.25y = 1, which is equivalent to 4x + 9y = 4. When we substitute 4x for 9y we get 4x + 4x = 4, which implies $x = \frac{1}{2}$. Then $y = \frac{4x}{9} = \frac{2}{9}$. Therefore, $v_2 = (1/2, \frac{2}{9})$. It follows that if consumer 1 demands $v_1 = (1/2, \frac{7}{9})$ when $p_1 = 1$ and $p_2 = 2.25$ and 2 demands v_2 then both markets will clear and person 1's utility will be

$$U_1 = \frac{1}{2} \times \frac{7}{9} = \frac{7}{18},$$

which is more than

$$U_1 = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4},$$

the utility that person 1 realizes by announcing his or her true utility-maximizing demand at every turn.

Had person 1 reported the true utility-maximizing demand θ_1 when $p_1 = 1$ and $p_2 = 2.25$ then 1 would have set b/a = 4/9. This implies 9b = 4a, and because individual 1's budget constraint at this price regime is a + 2.25b = 2.25, or 4a + 9b = 9, person 1 would set 8a = 9. In that case, a = 9/8 and thus b = 1/2. The consumption plan (9/8, 1/2) yields $U_1 = 9/8 \times 1/2 = 9/16$, which is even greater than 7/18. However, consumer 1 could never realize the utility level 9/16 because markets wouldn't clear. If the consumers continued to takes prices as given, the excess demand for good 1 when person 1 demanded the bundle (9/8, 1/2) would cause the price of the first commodity to rise relative to the second, pushing the price regime toward the equilibrium where $p_1 = p_2$. But when the two prices are equal person 1's utility is considerable less than 9/16 or even 7/18. This explains why it is in person 1's interest to demand v_1 when $p_1 = 1$ and $p_2 = 2.25$, even though the indifference curve through v_1 is not tangent to the budget constraint, as Figure 10.2 shows. (There is some indifference curve tangent to the budget line at v_1 —the broken curve of Figure 10.2—so individual 1 can always claim that he or she is following the rules.)

Note that person 1 misrepresents U_1 by demanding 7/9 of a unit of commodity 2, instead of 1/2 a unit that would be demanded if person 1 were passively maximizing utility subject to the budget constraint. Because person 1 is

a monopoly supplier of good 2, by *demanding* 7/9 units person 1 is under supplying commodity 2. That causes its price to be significantly higher than the competitive level and allows person 1 to purchase a bundle that yields more utility than the competitive equilibrium would yield.

Deviating from price taking behavior can be interpreted as preference misrepresentation: By demanding v_1 , individual 1 is in effect claiming that 1's preference scheme generates the broken indifference curve in Figure 10.2 instead of the solid curve. In terms of the algebra (Example 4.2), individual 1 is in effect announcing the utility function $U_1 = (4\sqrt{2a})/9 + b$. (Confirm that a price taker with this utility function will demand $v_1 = (1/2, 7/9)$ when $p_1 = 1$ and $p_2 = 2.25$.) We have demonstrated that truthful revelation is not a dominant strategy in the case of the Arrow-Debreu model of the private ownership market economy.

Perhaps a different resource allocation mechanism would induce truthful revelation of individual preference in a two-person exchange economy. We'll be wasting our time attempting to design one. There is no allocation mechanism for which truthful revelation is a dominant strategy, equilibrium outcomes are efficient, and all individuals are guaranteed at least as much utility as they would get simply by consuming their endowments and not participating in the economy. Section 4.1 of Chapter 8 proves this for economies that include a public good, although we do not present the corresponding proof for exchange economies with private goods only. Instead, we relax our incentive compatibility requirement and simply ask for a mechanism that yields efficient outcomes at each Nash equilibrium.

4.2 Nash equilibrium

Instead of seeking a resource allocation mechanism for which truthful revelation is a dominant strategy, let's merely ask for one in which Nash equilibria always exist and Nash equilibrium outcomes are always efficient.

Example 4.3: Two individuals and two feasible outcomes

There are two players and two possible outcomes, x and y. We only need to know whether an individual prefers x to y or the converse. (For simplicity we ignore the possibility that i is indifferent between x and y. We can think of each outcome as specifying the amounts of a large number of commodities to be consumed by each person.) Then there are four possible individual preference schemes:

(x, x), (x, y), (y, x), and (y, y).

We refer to each of these as an *environment*. At (x, x) each person prefers x to y. At (x, y) person 1 prefers x to y, but 2 prefers y to x. Person 1 prefers y to x at (y, x), but 2 prefers x to y. And they both prefer y to x at (y, y).

We show how demanding is the requirement that a Nash equilibrium exists for every specification of individual preferences and that every equilibrium give rise to an efficient outcome. Consider an arbitrary mechanism, which is simply

4. Efficiency and Incentive Compatibility

a specification of the set of strategies available to each individual and a function f, which determines whether the outcome is x or y, given the chosen strategies. That is, if consumer 1 plays s_1 and 2 plays s_2 then the outcome $f(s_1, s_2)$ is either x or y. We require that for each of the four possible environments of Example 4.3 a Nash equilibrium (s_1, s_2) exists, and $f(s_1, s_2)$ is efficient with respect to the underlying preferences.

Note that efficiency is a very modest requirement in this context. Both outcomes are efficient for the environments (x, y) and (y, x). Only x is efficient for (x, x) and only y is efficient for (y, y).

The Hurwicz-Schmeidler Theorem

For the economy of Example 4.3, if a mechanism has a Nash equilibrium for every environment, and every Nash equilibrium is efficient, then it must give one individual the power to force the mechanism to produce his or her most-preferred outcome in every environment, in which case we call that person a *dictator*.

Proof

Let (s_1, s_2) be a Nash equilibrium for (x, y). Assume that $f(s_1, s_2) = x$. (We treat the case $f(s_1, s_2) = y$ last.) Because (s_1, s_2) is an equilibrium and 2 prefers y to $x = f(s_1, s_2)$ it must be the case that $f(s_1, \beta) = x$ for every strategy β available to person 2. Now, let (t_1, t_2) be a Nash equilibrium for (y, x). If $f(t_1, t_2) = y$ then we must have $f(t_1, \beta) = y$ for every strategy β available to person 2 because (t_1, t_2) is an equilibrium for (y, x) and 2 prefers x to y at (y, x). Therefore, person 1 can ensure that the outcome is x by playing s_1 and can guarantee that the outcome is y by playing t_1 , in which case 1 is a dictator.

Suppose, then, that $f(t_1, t_2) = x$. Then (s_1, s_2) is a Nash equilibrium for (x, y), with $f(s_1, s_2) = x$, and (t_1, t_2) is a Nash equilibrium for (y, x), with $f(t_1, t_2) = x$. We now show that (s_1, t_2) is an equilibrium for (y, y), and that $f(s_1, t_2) = x$, contradicting efficiency: If $f(\alpha, t_2) = y$ for some strategy α available to person 1, then (t_1, t_2) is not an equilibrium for (y, x). Therefore, we have $f(\alpha, t_2) = x$ for all α available to person 1. In particular, $f(s_1, t_2) = x$. If $f(s_1, \beta) = y$ for some strategy β available to person 2, then (s_1, s_2) is not an equilibrium for (x, y). Therefore, $f(s_1, \beta) = x$ for all β . We have established that (s_1, t_2) is an equilibrium for (y, y), although $f(s_1, t_2) = x$. Therefore, existence and efficiency of equilibrium require $f(t_1, t_2) = y$ and hence person 1 is a dictator.

We began the proof by assuming that $f(s_1, s_2) = x$ and concluded that person 1 is a dictator. There is only one other possibility: $f(s_1, s_2) = y$. In that case, an analogous argument will reveal that person 2 is a dictator.

4.3 Subgame-perfect Nash Equilibrium

We can circumvent the Hurwicz-Schmeidler Theorem of the previous section by allowing one person to threaten to punish the other player for employing a strategy that would precipitate an inefficient outcome. A threat has to be





credible if we are to take the resulting equilibrium seriously. According, we employ subgame-perfect Nash equilibrium. (See Subsection 5.7 of Chapter 1.) This allows us to demand more of a resource allocation mechanism than mere efficiency. We assume that a social choice function *g* is given. In other words, *g* specifies the socially optimal outcome for each specification of individual preferences.

Example 4.3: A very simple resource allocation problem

We assume an exchange economy with two people and two private goods. The width of the box in Figure 10.3 is the total amount of good 1 available in this economy, and the height of the box is the total amount of good 2. A point such as *B* in the box specifies the consumption of each individual: Person 1 gets an amount of good 1 equal to the distance of *B* from the left-hand edge of the box, and an amount of good 2 equal to the height of *B* from the bottom of the box. Person 2 gets everything not allocated to person 1. Specifically, person 2's consumption of good 1 is the distance of *B* from the right-hand edge of the box, and 2's consumption of good 2 is the distance of *B* from the top of the box.

The individual's preferences can be represented by standard utility functions. Suppose there are two possible scenarios, called *environments*:

Case C: Both individuals have the Cobb-Douglas utility function $U_C(x, y) = xy$.

Case L: Both individuals have the Leontief utility function $U_L(x, y) = \min\{x, y\}$.

Person 1's utility function U_C is represented in Figure 10.3 by the solid indifference curves that bend in toward the origin. U_C is higher for person 1 at any

4. Efficiency and Incentive Compatibility

point above an indifference curve than it is at any point on the curve. Person 2's utility function U_C is represented by the solid indifference curves that bend up and away from the origin. As we move northeast in the diagram, person 2's utility decreases and 1's utility increases.

 U_L for person 1 is represented in Figure 10.3 by means of the L-shaped broken-line indifference curves, with U_L increasing as we move northeast in the box. For the environment *L*, person 2's indifference curves are the broken line curves that are L-shaped when the diagram is rotated 180°. (Turn the book upside down.) With the diagram in its normal position, person 2's utility increases as we move southwest.

The social choice function g designates the outcome g(C) in Figure 10.3 as socially optimal in case C, and g(L) in case L.

The rest of this subsection examines the possibility for implementing g. That is, we wish to design an economic system—a *mechanism*—whose equilibrium for each environment E is the outcome g(E) specified by g.

DEFINITION: Implementation

A mechanism specifies a set of strategies available to each participant in the economy. The mechanism implements the social choice rule g if, for each environment E for which g specifies an outcome g(E), the equilibrium of the mechanism at environment E is precisely g(E).

We begin by showing that it is impossible to implement the rule g defined by Example 4.3 if we require each *ordinary* Nash equilibrium to precipitate g(C) in case C and g(L) in case L.

Proof

Suppose to the contrary that we have a mechanism whose Nash equilibria in cases *C* and *L* yield, respectively, outcomes g(C) and g(L). Let (s_1, s_2) be a Nash equilibrium for case *C*. Then (s_1, s_2) must be an equilibrium for case *L* as well. If to the contrary there is some strategy α available to person 1 such that (α, s_2) gives person 1 higher utility than (s_1, s_2) , that is, more utility than $U_L(g(C))$, then that must be true for U_C also, contradicting the fact that (s_1, s_2) is a Nash equilibrium for case *C*. This follows from the fact that person 1's indifference curve through g(C) for the utility function U_L (the broken curve) is strictly above person 1's indifference curve through g(C) for the utility function U_L (the order U_C also, for any consumption plan *Z* for person 1, if $U_L(Z) > U_L(g(C))$ then $U_C(Z) > U_C(g(C))$. Therefore, if (s_1, s_2) is a Nash equilibrium for environment *C* then there is no α available to person 1 such that (α, s_2) gives person 1 higher utility than (α, s_2) .

Similarly, there can't be any strategy β available to person 2 such that (*s*₁, β) results in 2 receiving more utility than $U_L(gC)$): Figure 10.3 shows that any

outcome *Z* that person 2 prefers to g(C) according to U_L is southwest of 2's broken-line indifference curve through g(C) and hence southwest of 2's solid indifference curve through g(C), and thus person 2 would prefer Z to g(C) according to U_C . If (s_1, β) gave person 2 more utility than g(C) according to U_L , then (s_1, β) would give person 2 more utility than g(C) according to U_C , contradicting the supposition that (s_1, s_2) is a Nash equilibrium for *C*. Therefore, (s_1, s_2) must be a Nash equilibrium for environment *L* as well.

We have contradicted the claim that the mechanism implements *g*: The strategy pair (s_1, s_2) precipitates the outcome g(C) and is a Nash equilibrium for *L*. But $g(C) \neq g(L)$ as shown by Figure 10.3. Therefore, there is no mechanism that implements *g* in ordinary Nash equilibrium.

We now show that g can be implemented by an *extensive form* mechanism with subgame-perfect Nash equilibrium as our equilibrium concept. Specifically, all subgame-perfect equilibria for case C precipitate the outcome g(C), and all subgame-perfect equilibria for case L precipitate the outcome g(L).

Example 4.4: <i>A</i> simple mechanism with subgame-perfect equilibria that are optimal		
Stage 1.	Person 1 announces either L or C.If person 1 announces L the outcome is g(L) and there is no further play.If person 1 announces C then player 2 makes a move at stage 2.	
Stage 2.	Person 2 agrees or disagrees with player 1.If person 2 agrees that <i>C</i> is the true state then the outcome is g(<i>C</i>) and there is no further play. If person 2 disagrees then player 1 makes the next and last move at stage 3.	
Stage 3.	<i>Person 1 chooses outcome A or B.</i> The mechanism is represented as Figure 10.4.	

Before working out the subgame-perfect Nash equilibria for each of the two environments, *L* and *C*, we point out that in either situation both people will know which is the true environment. That's because each person knows that they have identical preferences in each environment. The mechanism designer (i.e., the consultant hired by the government) can exploit the fact that in each environment each consumer knows the other's utility function. That makes it easier to design a successful mechanism. However, neither the designer nor the government will know which of the environments is the true one—because individual preference is hidden information.

Analysis of Example 4.4

First, suppose that the true state is L, which means that both have Leontief preferences. Person 1 prefers g(C) to g(L), but if 1 announces C at stage 1 then player 2 will challenge, knowing that 1 will choose B over A at stage 3 because 1





prefers B to A in case L. Because person 2 prefers B to g(C) in case L, the challenge would be 2's best response. If person 1 threatened to choose A in stage 3 then person 2 would be forced to accept g(C) and we would have a Nash equilibrium. But it is not subgame perfect—that is, it is not in person 1's interest to carry out this threat if stage 3 were actually reached. The only subgame-perfect Nash equilibrium results in g(L). It has person 1 announcing L at stage 1 and declaring that he or she would choose B if stage 3 were reached, while person 2 declares her intention to challenge in stage 2 if person 1 announces C at stage 1.

Finally, suppose that the true state is C, and hence both consumers have Cobb-Douglas preferences. Person 1 has no incentive to announce L and take g(L) unless he fears that 2 would challenge at stage 2. But then person 1 would be forced to choose between A and B, and he would pick A, which gives more utility than B according to U_C , and A is worse for 2 than g(C). Therefore, even if both A and B are worse for player 1 than g(L), a threat by 2 to challenge C at stage 2 unless 1 announces L in stage 1 would not be credible, because if push came to shove the challenge would precipitate A, which gives 2 less utility (in terms of U_{C}) than g(C). Therefore, the only subgame-perfect equilibria in this case results in g(C). This is sustained by having person 1 announce C at stage 1, declaring that he would choose A if stage 3 were reached, while 2 announces that she would accept *C* at stage 2.

There are social choice functions that can be implemented in subgameperfect Nash equilibrium but not in Nash equilibrium.

The fact that there are social choice functions that can be implemented in subgame-perfect Nash equilibrium but not in Nash equilibrium is established by Example 4.3 and the following discussion.

4.4 Price taking

Now we consider an individual consumer's ability to manipulate prices in a private ownership market economy when the total number of individuals is large. Intuition tells us that an individual's power is negligible when there is a realistically large number of consumers.

Recall the definition of price taking behavior from Subsection 4.1: A trader is a *price taker* if he or she always demands a basket that maximizes utility *given* the price regime. Section 4.1 showed that price taking behavior is not a dominant strategy. Now we show that the gain from deviating from price taking behavior is virtually zero if the number of consumers is large. Specifically, we show that the gain from deviating goes to zero as the number of traders gets arbitrarily large. We demonstrate this by assuming a large number of individuals identical to person 1 of Example 4.1, and a large number identical to person 2.

Example 4.5: Many traders of each type

There are *t* individuals *i* with $U_i = ab$ and $\omega_i = (0, 1)$ and *t* individuals *j* with $U_j = xy$ and $\omega_j = (1, 0)$. We call the former type 1 and the latter type 2. We show that the ability of a single type-1 consumer to manipulate prices becomes negligible as *t* becomes sufficiently large. We begin by determining the demand functions: For convenience, let's normalize and set the price of the second good equal to unity, with *P* denoting the price of the first good. For a type 1 individual, utility maximization implies $MRS = b/a = p_1/p_2 = P$, and thus b = Pa. (Refer back to Example 2.4 for the derivation of the MRS.) The budget constraint is Pa + b = 1 because a type-1 person is endowed with 1 unit of the second good, whose price is \$1. The solution to b = Pa and Pa + b = 1 is

$$a = \frac{1}{2P}$$
 and $b = \frac{1}{2}$

These are the demands of a type-1 person as a function of the price ratio *P*, *assuming price taking behavior*. A type-2 consumer sets MRS = y/x = P. We have y = Px. The budget constraint is Px + y = P. (A type-2 person has 1 unit of good 1 to sell at price *P*.) Solving the two equations y = Px and Px + y = P yields

$$x = \frac{1}{2}$$
 and $y = \frac{P}{2}$,

the demand functions of a price taking type-2 person.

If everyone is a price taker, and P = 1, we have $a = \frac{1}{2} = x$, and thus the total demand for commodity 1 would be $t \times \frac{1}{2} + t \times \frac{1}{2} = t$, which equals the total supply, because there are *t* individuals (the type 2s), each supplying 1 unit of good 1. And when P = 1 we have $b = \frac{1}{2} = y$. The total demand for commodity 2 would be $t \times \frac{1}{2} + t \times \frac{1}{2} = t$, which equals the total supply, because there are *t* individuals (the type 1s), each supplying 1 unit of good 2. Therefore, P = 1 is the market clearing price ratio, *assuming price taking behavior*.

Now, suppose that everyone takes price as given and announces their true utility-maximizing demand vectors *except* for one type-1 person, whom we refer

4. Efficiency and Incentive Compatibility

to as person *m*. Then the total demand for good 1 from everyone except person *m* is $(t - 1) \times (1/2P) + t \times 1/2$. The total supply of the first good is exactly *t*, so even before we add on the demand of person *m* we see that the following inequality is a necessary condition for market equilibrium:

$$\frac{t-1}{2P} + \frac{t}{2} \le t.$$
[5]

This must hold because if (t - 1)/2P + t/2 > t then demand will exceed supply, whatever *m* announces, and market forces will cause the price to change. Note that statement [5] is equivalent to

$$P \ge \frac{t-1}{t}.$$
[6]

Similarly, the aggregate demand for the second good is at least $(t - 1) \times (1/2) + t \times P/2$. Exactly *t* units of the second good are supplied, so even without the demand of person *m* we must have

$$\frac{t-1}{2} + \frac{tP}{2} \le t;$$
[7]

otherwise the demand for good 2 would exceed the supply, even without including person *m*, and the prices would change. Statement [7] is equivalent to

$$P \le \frac{t+1}{t}.$$
[8]

Putting [6] and [8] together yields

$$\frac{t-1}{t} \le P \le \frac{t+1}{t}.$$
[9]

If *t* is even reasonably large, then both (t - 1)/t and (t + 1)/t will be very close to unity. Therefore, whatever person *m*'s demands, the market price ratio *P* will be very close to the *competitive* equilibrium price.

Example 4.1 featured an exchange economy with t = 1: One type-1 consumer and one type-2 consumer. Each individual is a monopoly supplier of one of the goods and hence has the ability to affect the market clearing price to advantage. Example 4.5 shows that when there is a large number of suppliers of each commodity, no individual can manipulate the price to any significant degree. Inequality [9] shows that any utility gain to the individual who deviates from price taking will be swallowed up by the cost of learning enough about aggregate demand to be able to manipulate the price in an advantageous way.

Sources

Hurwicz (1972) showed that price taking is not a dominant strategy and proved that there is *no* allocation mechanism for which truthful revelation is a dominant strategy, equilibrium outcomes are efficient, and each individual is guaranteed at least as much utility as that individual would get simply by consuming his or

her endowment and not participating in the economy. The theorem and proof of Section 4.2 are due to Hurwicz and Schmeidler (1978). The mechanism of Section 4.3 is from Moore and Repullo (1988). The proof that a single individual's ability to manipulate prices vanishes as the number of traders increases, which we illustrated with an example in Section 4.4, is due to Roberts and Postlewaite (1976).

Links

Jackson (2001) and Maskin and Sjöström (2002) provide extensive discussions of the incentive issues that have been sketched in this section.

Problem set

- 1. Show that the ability of a *type*-2 person to manipulate prices in the economy of Example 4.5 becomes negligible as *t* becomes sufficiently large.
- **2.** Consider an exchange economy with two goods and n = 2t individuals. The first *t* individuals have the utility function $U = a^2 b$ and endowment $\omega = (1, 0)$. The other individuals have the utility function U = xy with endowment $\omega = (0, 1)$. Show that the ability of a single individual to influence the price ratio is negligible if *t* is sufficiently large.

5 COMMON PROPERTY RESOURCES

Suppose that each of *n* firms has free access to a resource from which it can extract a marketable commodity, say fish. If we assume that the amount harvested by any entrepreneur depends on the effort expended by that agent *and* also on the effort of all others, and if we assume in addition that output per unit of effort declines as the total effort of all fishers increases, then we have the classical common property resource model in which the pursuit of self-interest leads to an inefficient rate of extraction in the short run and insufficient conservation in the long run. Let's examine the short run problem.

First, we need to identify the efficient rate of extraction. Let e_i be the effort expended by firm *i* on the lake, which we take as the common resource. That is, e_i denotes the number of hours of fishing per week spent by the workers in boat *i*. Let $e = e_1 + e_2 + e_3 + \cdots + e_n$ represent the total effort. For simplicity, assume that effort is undertaken at a constant opportunity cost of *c*. That is, one hour of fishing on the lake involves the sacrifice of *c* fish that could have been obtaining by fishing in the ocean for one hour. (Alternatively, *c* is the opportunity cost of leisure.) If T(e) denotes the total number of units of output (e.g., fish harvested) from all entrepreneurs as a function of total effort then A(e) = T(e)/e is the average product, which we assume declines as *e* increases. Therefore, marginal product M(e) is less than A(e) for each level of *e*. (You don't need calculus to follow this section, but if you want to use calculus then, of course, M(e) is the first derivative of T(e).)



Figure 10.5

 ∂ Let t(x) be any real-valued function. Then the average, t(x)/x, is itself a function of *x*, which we name a(x). Take the derivative of a(x):

$$\begin{aligned} x'(x) &= x^{-1}t'(x) - t(x)x^{-2} \\ &= \frac{t'(x) - [t(x)/x]}{x} = \frac{t'(x) - a(x)}{x} \end{aligned}$$

Then the average is falling (a' is negative) if and only if t'(x) < a(x). But t' is the marginal. Therefore, the average is falling if and only if the marginal is less than the average. This is quite intuitive. If the grade that you get in your next course is higher than your GPA then your GPA will rise, but if the your next grade is lower than your GPA then your GPA will fall.

The efficient level of effort e^* is that value of e that equates marginal cost and opportunity cost. This is depicted in Figure 10.5 where the intersection of the marginal product curve M(e) and the constant opportunity cost line c identifies the efficient level of effort e^* . The efficient rate of extraction of fish is $x^* = T(e^*)$.

Proof

If *e* is the actual amount of effort expended and M(e) < c then one unit of effort transferred *from* the lake *to* the ocean will increase the community's consumption of fish by c - M(e). There won't be any increase in cost in the form of additional input employed. Therefore, everyone can be made better off by distributing the gain c - M(e) throughout the community. This means that $M(e) \ge c$ must hold if the outcome is efficient. But if we actually have M(e) > c then one unit of effort transferred from the ocean to the lake will increase fish consumption by M(e) - c, and we can use the extra output to increase everyone's utility. Therefore, efficiency implies M(e) = c.

The efficient level of effort equates marginal product and marginal opportunity cost.

Now let's determine the *actual*—that is, equilibrium—extraction rate when the resource is available to anyone without a user charge, in which case the only cost to an individual entrepreneur as a result of expending a unit of effort in extraction is the opportunity cost *c*. It turns out that the equilibrium occurs at e^+ in Figure 10.5 where *average* product equals marginal cost, and that is *above* the efficient level of input (or effort).

Proof

Suppose that the current total input level is some amount e^0 below the point where A(e) equals c. Suppose that firm i now increases its input slightly. The addition to the total catch per unit of additional input will be $M(e^0)$, the marginal product at e^0 . *Society* would be better off if that additional unit of effort were employed in fishing the ocean where the yield would be c fish per unit of effort instead of $M(e^0)$. But the individual firm owner has no incentive to maximize social welfare. In the case of a common property resource the harvest per boat tends to be the average harvest. Even though the average harvest falls as a single firm increases its input level, the total harvest tends to be shared evenly by all boat owners. If the individual boat's catch equals the average output then it pays the boat's owner to increase the input level as long the extra harvest at the margin by the particular boat, which will be close to the average output, is greater than the private cost at the margin, which is c in this case. Therefore, harvesting will continue until average product is equal to marginal cost, or A(e) = c. This defines the equilibrium e^+ in Figure 10.5.

If the resource is available for a user charge of zero then the equilibrium level of effort equates average product and marginal opportunity cost, leading to an inefficient outcome.

If the government imposed a user charge of *g* dollars per hour spent fishing on the lake then the private marginal cost to an individual boat will now be c + g and self-interest will drive entrepreneurs to harvest up to the point where average product equals c + g. In other words, A(e) = c + g characterizes the equilibrium that results when there is a user charge of *g*. If *g* is set equal to $A(e^*) - c$ then A(e) = c + g implies $e = e^*$ and the efficient outcome is attained. (Figure 10.5 again).

∂ Example 5.1: Output is proportional to the square root of effort

Let *q* denote output. The production function is $q = 20\sqrt{e}$, where *e* is total effort over all firms. The opportunity cost of a unit of effort is 1. Consumer welfare is maximized by maximizing $20\sqrt{e} - e$. The first derivative is $10/\sqrt{e} - 1$,

5. Common Property Resources

and when we set that equal to 0 we get $e^* = 100$, the efficient level of effort. (Confirm that the second derivative is negative for all e > 0.) To determine the equilibrium effort level when the user cost of the resource is 0, we equate the average product to 1. Average product $= 20\sqrt{e}/e = 20/\sqrt{e}$. When we equate this to 1 we get $e^+ = 400$. Suppose that we currently have e = 100. Consider the position of entrepreneur *F* contemplating adding 200 units of effort. That would yield 200 units of output on the ocean. When devoted to fishing the lake it causes output to increase from $20\sqrt{100}$ to $20\sqrt{300}$, an increase of only 146.4. However, *F*'s yield is determined by average product. Total output will be $20\sqrt{300} = 230.9$, which is better than the 200 available to *F* on the ocean.

The efficient outcome would be realized if the rights to the lake were held by a profit-maximizing firm, which might be owned by a single individual or perhaps the entire community. Let's find the profit-maximizing fee f per unit of effort—that is, per hour on the lake. Assume that the costs of setting and collecting the fee are independent of the intensity of economic activity on the lake. That is, the costs of the lake owner are fixed, at k, say. Then the lake owner's profit is fe - k. The functions fe and fe - k are maximized by the same value of e.

Let's determine the equilibrium value of e when the users of the lake must pay a user charge of f to the private owner: They will harvest up to the point where A(e) = c + f. Therefore, the profit-maximizing owner of the lake will set f = A(e) - c at equilibrium, and thus fe = A(e)e - ce. Because A(e)e is just total product, T(e), the owner of the lake will want to maximize T(e) - ce, which is just total product minus total variable cost. We have already examined this case. The function T(e) - ce will be maximized when marginal product equals marginal cost, at e^* . Therefore, profit maximization by the owner of the lake leads to the efficient outcome, unlike free, unrestricted use of the lake.

Notice that the social optimum can be realized by private ownership or by government directive. (Verify that f = g.) Which approach goes further toward promoting social welfare? Perhaps it doesn't matter whether society chooses the public or private remedy for correcting the inefficiency. They yield the same outcome, e^* . Is there any difference between the two regimes?

In both cases the principal delegates the crucial job to an agent. The public approach requires the state to impose a user fee. The principal is the society, and the agent is a government body that is given the job of computing the efficient fee g and enforcing it. Let's call this government agency the regulator. Acquiring the necessary information about marginal and average product is not an easy task. What assurance is there that the regulator will devote the necessary effort to this task? Who monitors the regulator? We can ask the same question of the private solution. It works only if the owner of the lake maximizes profit. But owners typically hire an agent, called a manager, to carry out this task for them. The manager's personal welfare is his or her chief concern, and that is not perfectly

aligned with the interests of the owners. But the manager is monitored by the capital market in a capitalist economy. If the manager's activities do not provide

Wild salmon feed and grow off the southwest coast of Greenland and Faroe Islands before returning to spawn in their native rivers. The salmon population in these rivers has declined precipitously as a result of ocean net fishing. When salmon are caught they don't return to the rivers to spawn. Because there are private property rights in the ocean netting of salmon off Greenland and Faroe Islands it is possible to restore the wild salmon population without impoverishing the ocean fishermen. The property rights consist of individual boat quotas, which can be bought and sold. A boat can't harvest salmon unless it holds a quota, which is an entitlement to harvest a given number of fish. The Atlantic Salmon Federation compensates fishermen for not exercising their netting rights, resulting in a much greater number of salmon returning to spawn. By contrast, the U.S. government has spent more than \$35 million to restock Atlantic coastal rivers with salmon, to little effect (Anderson and Leal, 1997, pp. 156-7)

a satisfactory rate of return on the firm's capital equipment and resources then a takeover would be profitable for someone who has access to an agent who *can* get the most out of the inputs.

Moreover, there is usually better information on the performance of agents in the private sector than in the government sector. Almost all of the relevant data can be reduced to a single number: rate of return on capital. Another reason why government agents can be harder to motivate is that their budget constraints are typically softer than those that govern firms in the private sector. (However, if a private firm anticipates that the government will come to its rescue, the discipline effect of its budgets constraint will soften. In the late 1970s the U.S. government gave massive assistance to the Chrysler Corporation when it was on the verge of bankruptcy. Similarly, the government used the "too big to fail" rationale to assist some large U.S. banks in the 1980s.) This is just a very preliminary look at the question of public versus private regulation, but it gives an indication of the kind of issues that have to be settled before one can decide whether the public or private remedy is better for society in a particular case.)

Link

Ostrom (1990) is a thorough treatment of the theory and history of common property resources.

Problem set

- 1. Confirm that the profit-maximizing fee *f* set by the monopoly owner of the common property resource is equal to the fee *g* that the government would charge to elicit an efficient outcome.
- **2.** Suppose that $16\sqrt{e}$ units of output result when a total of *e* units of effort are supplied. The opportunity cost of *e* units of effort is $1/2 e^2$. Determine the efficient level of effort and the equilibrium level of effort when the resource is freely available.
- **3.** Show that the constant opportunity cost assumption is critical to our demonstration that profit maximization will lead to efficiency.

References

- Abdulkadiroğlu, Atila, and Tayfun Sönmez. (1999). House allocation with existing tenants. *Journal of Economic Theory* 88: 233–260.
- Abdulkadiroğlu, Atila, and Tayfun Sönmez. (2003). School choice: A mechanism design approach. *American Economic Review* 93: 729–747.
- Akerlof, G. A. (1970). The market for "lemons": Qualitative uncertainty and the market mechanism. *Quarterly Journal of Economics* 84: 488–500.
- Akerlof, G. A., and P. Romer. (1994). Looting: The economic underworld of bankruptcy for profit. *Brookings Papers on Economic Activity*, Vol. 2: 1–73.
- Alchian, A., and H. D. Demsetz. (1972). Production, information costs, and economic organization. *American Economic Review* 62: 777–795.
- Allen, D. A., and D. Lueck. (2002). *The Nature of the Farm: Contracts, Risk, and Organization in Agriculture*. Cambridge: MIT Press.
- Allen, Franklin, and Gale, Douglas. (2000). Corporate governance and competition. In *Corporate Governance*, ed. Xavier Vives, 23–84. Cambridge, UK: Cambridge University Press.
- Anderson, L. R. (2001). Public choice as an experimental science. In *The Elgar Com*panion to Public Choice, ed. W. F. Shughart and L. Razzolini, 497–511. Cheltenham, UK: Edward Elgar.
- Anderson, R. M., M. A. Khan, and S. Rashid. (1982). Approximate equilibrium with bounds independent of preferences. *Review of Economic Studies* 44: 473–475.
- Anderson, T. L., and D. R. Leal. (1997). *Enviro-capitalists: Doing Good While Doing Well*. New York: Rowman and Littlefield.
- Aoki, Masahiko. (2000). Information and governance in the Silicon Valley model. In *Corporate Governance*, ed. Xavier Vives, 169–195. Cambridge, UK: Cambridge University Press.
- Arrow, K. J. (1951a). Social Choice and Individual Values. 1st ed. New York: Wiley.
- Arrow, K. J. (1951b). An extension of the basic theorems of classical welfare economics. In *Proceedings of the Second Berkeley Stigl Symposium on Mathematical Statistics and Probability*, ed. J. Neyman, 507–532. Berkeley: University of California Press.
- Arrow, K. J. (1963a). Social Choice and Individual Values. 2nd ed. New York: Wiley.
- Arrow, K. J. (1963b). Uncertainty and the welfare economics of medical care. *American Economic Review* 53: 941–973.

- Arrow, K. J. (1970). The organization of economic activity: Issues pertinent to the choice of market versus non-market allocation. In *Public Expenditures and Policy Analysis*, ed. R. H. Haveman and J. Margolis, 59–73. Chicago: Markham.
- Arrow, K. J. (1971). Essays in the Theory of Risk-Bearing. Chicago: Markham.
- Arrow, K. J. (1979). The property rights doctrine and demand revelation under incomplete information. In *Economics and Human Welfare*, ed. M. Boskin, 23–39. New York: Academic Press.
- Arrow, K. J. (1984). The economics of agency. In *Principals and Agents: The Structure of Business*, ed. J. Pratt and R. Zeckhauser, 37–51. Boston: Harvard Business School Press.
- Arrow, K. J., and G. Debreu. (1954). Existence of equilibrium for a competitive economy. *Econometrica* 22: 265–290.
- Ashenfelter, O. (1989). How auctions work for wine and art. *Journal of Economic Perspectives* 3: 23–36.
- Attiyeh, G., R. Franciosi, and M. Isaac. (2000). Experiments with the pivotal process for providing public goods. *Public Choice*, forthcoming.
- Axelrod, R. (1984). The Evolution of Cooperation. New York: Basic Books.
- Ayres, Ian, and S. D. Levitt. (1998). Measuring positive externalities from unobservable victim precaution: An empirical analysis of Lojack. *Quarterly Journal of Economics* 112: 43–77.
- Baker, George. (2002). Distortion and risk in optimal incentive contracts. *Journal of Human Resources* 37: 728–751.
- Baker, George, Robert Gibbons, and K. J. Murphy (2002). Relational contracts and the theory of the firm. *Quarterly Journal of Economics* 117: 39–84.
- Balazs, E. (1964). *Chinese Civilization and Bureaucracy*. New Haven: Yale University Press.
- Balinksi, Michel, and Tayfun Sönmez. (1999). A tale of two mechanisms: Student placement. *Journal of Economic Theory* 84: 73–94.
- Ballou, Dale, and Michael Podgursky. (1997). *Teacher Pay and Teacher Quality*. Kalamazoo, Michi.: W. E. Upjohn Institute for Employment Research.
- Ballou, Dale, and Michael Podgursky. (2001). Merit pay symposium: Let the market decide. *Education Next* 1, Spring: 16–25.
- Banerjee, A., and E. W. Eckard. (2001). Why regulate insider trading? Evidence from the first great merger wave (1897–1903). *American Economic Review* 91: 1329–1349.
- Barberà, S. (2001). An introduction to strategy proof social choice functions. *Social Choice and Welfare* 18: 619–653.
- Barberà, S. (2004). Strategy proofness. In *Handbook of Social Choice and Welfare,* Vol. 2, ed. K. J. Arrow, A. K. Sen, and K. Suzumura, Chapt. 24. Amsterdam: Elsevier.
- Bardhan, P. K., and J. E. Roemer. (1993). *Market Socialism: The Current Debate*. New York: Oxford University Press.
- Basu, Kaushik. (1994). The traveler's dilemma: Paradoxes of rationality in game theory. *American Economic Review: Papers and Proceedings* 84: 391–395.
- Baumol, W. J. (1993). *Entrepreneurship, Management, and the Structure of Payoffs*. Cambridge, Mass.: MIT Press.
- Baumol, W. J. (2002). *The Free-Market Innovation Machine*. Princeton, N.J.: Princeton University Press.
- Bebchuk, L. A., J. M. Fried, and D. I. Walker. (2002). Managerial power and rent extraction in the design of executive compensation. *University of Chicago Law Review* 69: 751–846.
- Bernstein, P. (1992). Capital Ideas. New York: The Free Press.
- Bertrand, M., and S. Mullainathan. (2000). Agents with and without principals. *American Economic Review* 90: 203–208.
- Bertrand, M., and S. Mullainathan. (2001). Are CEOs rewarded for luck? Agents without principals are. *Quarterly Journal of Economics* 116: 910–932.
- Bertrand, M., and S. Mullainathan. (2003). Enjoying the quiet life. *Journal of Political Economy* 111: 1043–1075.
- Binmore, K. (1992). Fun and Games. Lexington, Mass.: D.C. Heath.
- Binmore, K., and J. Davies. (2001). Calculus: Concepts and Methods. Cambridge, UK: Cambridge University Press.
- Binmore, K., and P. Klemperer. (2002). The biggest auction ever: The sale of British 3G telecom licenses. *Economic Journal* 112: C74–C96.
- Binmore, K., and L. Samuelson. (1992). Evolutionary stability in repeated games played by finite automata. *Journal of Economic Theory* 57: 278–305.
- Blair, R. D., and D. L. Kaserman. (1991). The economics and ethics of alternative cadaveric organ procurement policies. *Yale Journal on Regulation* 8: 403–452.
- Borda, J.-C. (1781). *Memoire sur les elections au scrutins*. Histoire de l Academie Royale des Sciences.
- Brams, S. J. (1976). *Paradoxes in Politics: An Introduction to the Nonobvious in Political Science*. New York: The Free Press.
- Brickley, J. A., J. L. Coles, and R. L. Terry. (1994). Outside directors and the adoption of poison pills. *Journal of Financial Economics* 35: 371–390.
- Calvert, R. L. (1986). *Models of Imperfect Information in Politics*. Chur, Switzerland: Harwood Academic Publishers.
- Camerer, C. F., L. Babcock, G. Lowenstein, and R. H. Thaler. (2004). Labor supply of New York City cab drivers: One day at a time. In *Advances in Behavioral Economics*, ed. C. F. Camerer, G. Lowenstein, and M. Rabin, 533–547. Princeton, N.J.: Princeton University Press.
- Cameron, S., and J. Heckman. (1993). The nonequivalence of high school equivalents. *Journal of Labor Economics* 11: 1–47.
- Campbell, D. E. (1987). *Resource Allocation Mechanisms*. New York: Cambridge University Press.
- Campbell, D. E., and J. S. Kelly. (2000). A simple characterization of majority rule. *Economic Theory* 15: 689–700.
- Campbell, D. E., and J. S. Kelly. (2002). Impossibility theorems in the Arrovian framework. In *Handbook of Social Choice and Welfare*, Vol. 1, ed. K. J. Arrow, A. K. Sen, and K. Suzumura, 35–94.
- Campbell, D. E., and J. S. Kelly. (2003). A strategy-proofness characterization of majority rule. *Economic Theory* 22: 557–568.
- Campbell, D. E., and M. Truchon. (1988). Boundary optima in the theory of public goods supply. *Journal of Public Economics* 35: 241–249.
- Carlin, Wendy, and Colin Mayer. (2000). How do financial systems affect economic performance? In *Corporate Governance*, ed. Xavier Vives, 137–160. Cambridge, UK: Cambridge University Press.

- Carmichael, H. L. (1988). Incentives in academics: Why is there tenure? *Journal of Political Economy* 96: 453–472.
- Carmichael, H. L. (1989). Self-enforcing contracts, shirking, and life-cycle incentives. *Journal of Economic Perspectives* 3: 65–83.
- Carmichael, H. L. (2001). Tenure promotes honest hiring. *Policy Options* 22: 10–14.
- Carmichael, H. L., and B. W. McLeod. (2000). Worker cooperation and the ratchet effect. *Journal of Labor Economics* 18: 1–19.
- Carr, Jack L., Frank Mathewson, and N. C. Quigley. (1995). Stability in the absence of deposit insurance: The Canadian banking system 1890–1966. *Journal of Money, Credit and Banking* 27: 1137–1158.
- Chen, Yan. (2006). Incentive-compatible mechanisms for pure public goods: A survey of experimental research. In C. R. Plott and V. Smith, eds., *The Handbook of Experimental Economics Results. Volume 1*. Amsterdam: Elsevier, in press.
- Chen, Yan, and Tayfun Sönmez. (2002). Improving efficiency of on-campus housing: An experimental study. *American Economic Review* 92: 1669–1686.
- Chen, Yan, and Tayfun Sönmez. (2003). School choice: An experimental study. Available at http://home.ku.edu.tr/~tsonmez/school_choice_experiment_ 2003-11-06.pdf. Accessed on December 15, 2005.
- Chen, Yan, and Tayfun Sönmez. (2004). An experimental study of house allocation mechanisms. *Economics Letters* 83: 137–140.
- Chiappori, P.-A. and B. Salanié. (2003). Testing contract theory: A survey of some recent work. In Advances in Economics and Econometrics: Invited Lectures to Eighth World Congress of the Econometric Society, ed. M. Dewatripont, L. Hansen, and S. Turnovsky, 115–149. Cambridge, UK: Cambridge University Press.
- Chichilnisky, G., and G. Heal. (1993). Global environmental risks. *Journal of Economic Perspectives* 7: 65–86.
- Chichilnisky, G., and G. Heal. (1999). Catastrophe futures: Financial markets for unknown risks. In *Markets, Information, and Uncertainty: Essays in Economic Theory in Honor of Kenneth J. Arrow,* ed. G. Chichilnisky, 120–140. New York: Cambridge University Press.
- Chou, Y. K. (2000). Testing alternative models of labor supply: Evidence from taxi drivers in Singapore. University of Melbourne Department of Economics. Available at http://www.economics.unimelb.edu.au/research/768/htm.
- Clarke, E. H. (1971). Multipart pricing of public goods. Public Choice 8: 19-73.
- Coase, R. H. (1960). The problem of social cost. *Journal of Law and Economics* 3: 144–171.
- Coase, R. H. (1974). The lighthouse in economics. *Journal of Law and Economics* 17: 357–376.
- Comment, Robert, and G. W. Schwert. (1995). Poison or placebo: Evidence on the deterrence and wealth effects of modern antitakeover measures. *Journal of Financial Economics* 39: 3–43.
- Condorcet, Marquis de. (1785). *Essai sur la'application de l'analyse à la possibilité des décisions rendues à la pluralité des voix*. Paris: Imprimerie Royale. (Reprinted in 1972 by Chelsea Publishing Co., New York.)
- Conley, J., and D. Diamantaris. (1996). Generalized Samuelson and welfare theorems for nonsmooth economies. *Journal of Public Economics* 59: 137–152.

- Cook, C. (1990). A survey of perestroika. The Economist April p. 6.
- Cooter, R. D., and T. S. Ullen. (1994). *Law and Economics*. 2nd ed. Glenview, Il.: Scott, Foresman.
- Corchón, L. C. (1996). *The Theory of Implementation of Socially Optimal Decisions in Economics*. London: Macmillan Press.

Cornes, R., and T. Sandler. (1996). *The Theory of Externalities, Public Goods, and Club Goods.* 2nd ed. New York: Cambridge University Press.

- Costa, D. L. (1998). The Evolution of Retirement. Chicago: University of Chicago Press.
- Cramton, P., R. Gibbons, and P. Klemperer. (1987). Dissolving a partnership efficiently. *Econometrica* 55: 615–632.
- Cramton, P., and S. Kerr. (1999). The distributional effects of carbon regulation. In *The Market and the Environment*, ed. Thomas Sterner, Chapt. 12. Cheltenham, UK: Edward Elgar.

Cramton, P., and S. Kerr. (2002). Tradeable carbon permit auctions. *Energy Policy* 30: 333–345.

- Cramton, P., and J. A. Schwartz. (2000). Collusive bidding: Lessons from the FCC spectrum auctions. *Journal of Regulatory Economics* 17: 229–252.
- d'Aspremont, C., and L.-A. Gerard-Varet. (1979). Incentives and incomplete information. *Journal of Public Economics* 11: 25–45.
- Dana, J. D. (1998). Advance-purchase discounts and price discrimination in competitive markets. *Journal of Political Economy*. 106: 395–422.

Danilov, V. I., and A. I. Sotskov. (2002). Social Choice Mechanisms. Berlin: Springer.

- Debreu, G. (1959). Theory of Value. New York: Wiley.
- Demsetz, H. (1968). Why regulate utilities? *Journal of Law and Economics* 11: 55–65.
- Demirgüç-Kunt, A., and E. J. Kane. (2002). Deposit insurance around the globe. Where does it work? *Journal of Economic Perspectives* 16: 175–195.
- Dewatripont, Mathias, and Jean Tirole. (1994). *The Prudent Regulation of Banks*. Cambridge, Mass.: MIT Press.
- Dewatripont, Mathias, L. Hansen, and S. Turnovsky. (2003). *Advances in Economics and Econometrics: Invited Lectures to Eighth World Congress of the Econometric Society*, Vol. 1. Cambridge, UK: Cambridge University Press.
- Diamond, Peter. (1992). Organizing the health insurance market. *Econometrica* 60: 1233–1254.
- Dixit, A. K. (2004). *Lawlessness and Economics: Alternative Modes of Governance*. Princeton, N.J.: Princeton University Press.
- Downs, A. (1957). An Economic Theory of Democracy. New York: Harper & Row.
- Dranove, David. (2000). *The Economic Evolution of American Health Care*. Princeton, N.J.: Princeton University Press.
- Dray, Phillip. (2002). *At the Hands of Persons Unknown: The Lynching of Black America*. New York: Random House.
- Drèze, J., and A. K. Sen. (1989). *Hunger and Public Action*. Oxford, UK: Oxford University Press.
- Dutta, P. K. (2000). Strategies and Games. Cambridge, Mass.: MIT Press.
- Dyer, D., and J. H. Kagel. (2002). Bidding in common value auctions: How the commercial construction industry corrects for the winner's curse. Pp. 349–393 in Kagel and Levin (2002a).

- Easterbrook, F. H. (1984). Insider trading as an agency problem. In *Principals and Agents: The Structure of Business*, ed. J. Pratt and R. Zeckhauser. Boston, Mass.: Harvard Business School Press, 81–100.
- Easterbrook, F. H. (1986). Manager's discretion and investor welfare: Theories and evidence. *Delaware Journal of Corporate Law* 9: 540–567.
- Ehrenberg, R. G. (2000). *Tuition Rising: Why College Costs So Much*. Cambridge, Mass.: Harvard University Press.

Ergin, Haluk, and Tayfun Sönmez. (2004). Games of school choice among Boston students. Massachusetts Institute of Technology, Department of Economics working paper.

- Farber, H. S. (2003). Is tomorrow another day? The labor supply of New York City cab drivers. National Bureau of Economic Research, working paper no. 9706. Available at http://www.nber.org/papers/w9706. (Also available as working paper no. 743, Industrial Relations Section, Princeton University.)
- Farrell, J. (1987). Information and the Coase theorem. *Journal of Public Economics* 1: 113–129.
- Fay, S., E. Hurst, and M. J. White, (2002). The household bankruptcy decision. American Economic Review 92: 706–718.
- Foley, D. (1970). Lindahl's solution and the core of an economy with public goods. *Econometrica* 38: 66–72.
- Frank, R. H. (2004). *What Price the Moral High Ground? Ethical Dilemmas in Competitive Environments*. Princeton, N.J.: Princeton University Press.
- Freeman, R. B., and M. M. Kleiner. (1998). The last American shoe manufacturer: Changing the method of pay to survive foreign competition. National Bureau of Economic Research, Cambridge, Mass. Working Paper No. 6750.
- Freixas, Xavier, and J.-C. Rochet. (1997). *The Microeconomcs of Banking*. Cambridge, Mass.: MIT Press.
- Friedman, B. M., and F. H. Hahn. (1990). *Handbook of Monetary Economics*, Vol. 2. Amsterdam: Elsevier Science.
- Friedman, J. (1971). A non-cooperative equilibrium for supergames. *Review of Economic Studies* 38: 1–12.
- Fudenberg, D., and E. S. Maskin. (1986). The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54: 533–554.
- Gale, D., and L. S. Shapley. (1962). College admissions and the stability of marriage. *American Mathematical Monthly* 69: 9–15.
- Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica* 40: 587–602.
- Gibbons, R. (1992). *Game Theory for Applied Economists*. Princeton, N.J.: Princeton University Press.
- Gintis, Herbert. (2000). *Game Theory Evolving: A Problem-Centered Introduction to Modeling Strategic Interaction*. Princeton, N.J.: Princeton University Press.
- Goeree and Holt. (2001). Ten little treasures of game theory, and ten intuitive contradictions. *American Economic Review* 91: 1402–1422.
- Green, J., and J.-J. Laffont. (1977). Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica* 45: 427–438.
- Green, J., and J.-J. Laffont. (1979). *Incentives in Public Decision Making*. Amsterdam: Elsevier.

- Gribbin, John, and Mary Gribbin. (1997). *Richard Feynman: A Life in Science*. New York: Dutton.
- Grossman, S. J., and O. Hart. (1980). Take-over bids, the free-rider problem, and the theory of the corporation. *Bell Journal of Economics* 11: 42–64.
- Grossman, S. J., and O. Hart. (1982). Corporate financial structure and managerial incentives. In *The Economics of Information and Uncertainty*, ed. J. McCall. Chicago: University of Chicago Press, 109–140.
- Groves, T. (1973). Incentives in teams. *Econometrica* 41: 617–631.
- Gruber, J., and M. Owings. (1996). Physician financial incentives and Caesarian section delivery. *RAND Journal of Economics* 27: 99–123.
- Hall, B. H. (1988). The effect of takeover activity on corporate research and development. In *Corporate Takeovers: Causes and Consequences*, ed. A. J. Auerbach, 69–96. Chicago: University of Chicago Press.
- Hall, B. J. (2000). Regulatory free cash flow and the high cost of insurance company failures. *Journal of Risk and Insurance* 67: 415–438.
- Hall, J., and J. Liebman. (1998). Are CEOs really paid like bureaucrats? *Quarterly Journal of Economics* 113: 653–691.
- Hall, B. J., and K. J. Murphy. (2000). Optimal exercise prices for risk averse executives. *American Economic Review* 90: 209–214.
- Hall, B. J., and K. J. Murphy. (2002). Stock options for undiversified executives. *Journal* of Accounting and Economics 33: 3–42.
- Hall, B. J., and K. J. Murphy. (2003). The trouble with stock options. *Journal of Economic Perspectives* 17: 49–70.
- Hammons, C. W. (2001). School at home. Education Next 1: 48-55.
- Hansmann, Henry. (1996). *The Ownership of Enterprise*. Cambridge, Mass.: Cambridge University Press.
- Hanushek, E. A. (2002). The seeds of growth. Education Next 2, Fall: 10-17.
- Harrison, B. (1993). Taking the high road to growth. Technology Review October: 68.
- Haubrich, Joseph. (1994). Risk aversion, performance pay, and the principal-agent problem. *Journal of Political Economy* 102: 258–276.
- Hendel, I., and Alessandro Lizzeri. (1999). Adverse selection in durable goods markets. *American Economic Review* 89: 1097–1115.
- Herstein, I. N., and J. Milnor. (1953). An axiomatic approach to measurable utility. *Econometrica* 21: 291–297.
- Hicks, J. R. (1939). Value and Capital. London: Clarendon Press.
- Hill, Christopher. (1986). The political dilemma for Western governments. In *Terrorism and International Order*, ed. Lawrence Freedman, Christopher Hill, Adam Roberts, R. J. Vincent, Paul Wilkinson, and Philip Windsor. London: Routledge and Kegan Paul.
- Hirshleifer, J., and J. G. Riley. (1992). *The Analytics of Uncertainty and Information*. Cambridge, UK: Cambridge University Press.
- Holmström, B. (1979a). Moral hazard and observability. *Bell Journal of Economics* 10: 74–91.
- Holmström, B. (1979b). Groves' schemes on restricted domains. *Econometrica* 47: 1137–1144.
- Holmström, B., and S. N. Kaplan. (2001). Corporate governance and merger activity in the U.S.: Making sense of the 1980s and 1990s. *Journal of Economic Perspectives* 15: 121–144.

- Holmström, B., and S. N. Kaplan. (2003). The state of U.S. corporate governance: What's right and what's wrong. *Journal of Applied Corporate Finance* 16: 8–20.
- Holt, C. A., and J. Goeree. (2001). Ten little treasures of game theory, and ten intuitive contradictions. *American Economic Review* 91: 1402–1422.
- Hoxby, C. M. (2001a). Changing the profession: How would school choice affect teachers? *Education Next* 1, Spring: 57–63.
- Hoxby, C. M. (2001b). Rising tide: New evidence on competition and the public schools. *Education Next* 1, Winter: 68–74.
- Hoxby, C. M. (2002). Would school choice change the teaching profession? *Journal of Human Resources* 38: 846–891.
- Hsiao, W. C., P. Braun, P. Dunn, and E. R. Becker. (1988). Resource-based relative values: An overview. *Journal of the American Medical Association* 260 (16): 2418–2424.
- Hurwicz, L. (1972). On informationally decentralized systems. In *Decision and Organization*, ed. C. B. McGuire and R. Radner, 297–336. Amsterdam: Elsevier.
- Hurwicz, L., and D. Schmeidler. (1978). Construction of outcome functions guaranteeing existence and Pareto optimality of Nash equilibria. *Econometrica* 46: 1447–1474.
- Hurwicz, L., and M. Walker. (1990). On the article nonoptimality of dominantstrategy allocation mechanisms: A general theorem that includes pure exchange economies. *Econometrica* 58: 683–704.
- Illing, G., and U. Klüh. (2003). Spectrum Auctions and Competition in Telecommunications. Cambridge, Mass.: MIT Press.
- Jackson, M. (2001). A crash course in implementation theory. *Social Choice and Welfare* 18: 655–708.
- Jaffee, Dwight, and Joseph Stiglitz. (1990). Credit rationing. Chapter 1 (pp. 837–888) in Friedman and Hahn (1990).
- Jarrell, G. A., J. A. Brickley, and J. M. Netter. (1988). The market for corporate control: The empirical evidence since 1980. *Journal of Economic Perspectives* 2: 49–68.
- Jensen, M. C. (1986). Agency costs of free cash flow, corporate finance, and takeovers. *American Economic Review* 76: 323–329.
- Jensen, M. C. (1988). Takeovers: Their causes and consequences. *Journal of Economic Perspectives* 2: 1–48.
- Jensen, M. C. (1993). The modern industrial revolution, exit, and the failure of internal control systems. *Journal of Finance* 48: 831–880.
- Jensen, M. C., and Kevin J. Murphy. (1990a). CEO incentives: It's not how much you pay, but how. *Harvard Business Review* 68: 138–153.
- Jensen, M. C., and Kevin J. Murphy. (1990b). Performance pay and top-management incentives. *Journal of Political Economy* 98: 225–264.
- Joskow, P. L., Schmalensee, Richard, and Bailey, E. M. (1998). The market for sulfur dioxide emissions. *American Economic Review* 88: 669–685.
- Kagel, J. H., and D. Levin. (2002a). *Common Value Auctions and the Winner's Curse*. Princeton, N.J.: Princeton University Press.
- Kagel, J. H., and D. Levin. (2002b). Bidding in common-value auctions: A survey of experimental research. Pp. 1–84 in Kagel and Levin (2002a).
- Kahn, C., and G. Huberman. (1988). Two-sided uncertainty and up-or-out contracts. *Journal of Labor Economics* 6: 423–445.

- Kandel, E., and E. P. Lazear. (1992). Peer pressure and partnerships. *Journal of Political Economy* 100: 801–817.
- Kane, E. J. (1989). *The S & L Insurance Mess: How Did It Happen?* Washington, D.C.: Urban Institute Press.
- Kanter, R. M. (1989). When Giants Learn to Dance. New York: Simon and Schuster.
- Katzner, D. W. (1970). Static Demand Theory. New York: Macmillan.
- Kelly, J. S. (1988). Social Choice Theory. Berlin: Springer-Verlag.
- Kirby, K. N., M. Santiesteban, and G. C. Whinston. (2003). Impatience and grades: Delay-discount rates correlate negatively with college GPA. DP–63, Williams Project on the Economics of Higher Education, Department of Economics, Williams College. Available at http://www.williams.edu/wpehe/downloads.html. Accessed on December 15, 2005.
- Klemperer, Paul. (1998). Auctions with almost common values. *European Economic Reviews* 42: 757–769.
- Klemperer, Paul. (1999). Auction theory: A guide to the literature. *Journal of Economic Surveys* 13: 227–286. Reprinted as Chapter 1 in Klemperer (2004).
- Klemperer, Paul. (2002a). What really matters in auction design. *Journal of Economic Perspectives* 16: 169–189. Reprinted as Chapter 3 in Klemperer (2004).
- Klemperer, Paul. (2002b). How (not) to run auctions: The European 3G telecom auctions. *European Economic Review* 46: 829–845. Reprinted as Chapter 7 in Illing and Klüh (2003).
- Klemperer, Paul. (2003). Why every economist should learn some auction theory. Pp. 25–55 in Dewatripont, Hansen, and Turnovsky (2003). Reprinted as Chapter 2 in Klemperer (2004).
- Klemperer, Paul. (2004). *Auctions: Theory and Practice*. Princeton, N.J.: Princeton University Press.
- Kmietowicz, Zosia. (2000). WHO warns of threat of "superbugs." *British Medical Journal* 320: 1624.
- Kohler, Heinz. (1977). *Scarcity and Freedom: An Introduction to Economics*. Lexington, Mass.: D.C. Heath.
- Koopmans, T. C. (1957). *Three Essays on the State of Economic Science*. New York: McGraw–Hill.
- Kotlikoff, L. J., and S. Burns. (2004). *The Coming Generational Storm: What You Need to Know About America's Economic Future*. Cambridge, Mass.: MIT Press.
- Kotowitz, Y. (1989). Moral hazard. In *Allocation, Information, and Markets*, ed. J. Eatwell, M. Milgate, and P. Newman, 207–213. New York: Norton.
- Krauss, L. M. (1993). The Fear of Physics. New York: Basic Books.
- Kreps, D. M. (1988). Notes on the Theory of Choice. Boulder, Colo.: Westview Press.
- Kreps, D. M. (1990). *Game Theory and Economic Modeling*. Oxford, UK: Oxford University Press.
- Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory* 27: 245–252.
 Krishna, Vijay. (2002). *Auction Theory*. San Diego, Calif.: Academic Press.
- Krueger, A. B. (1993). How computers have changed the wage structure: Evidence from microdata, 1984–1989. *Quarterly Journal of Economics* 107: 33–60.
- Krueger, A. B., and A. Mas. (2004). Strikes, scabs, and tread separation: Labor strife and the production of defective Bridgestone/Firestone tires. *Journal of Political Economy* 112: 253–289.

- Kuhn, H. W., and A. W. Tucker. (1950). Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. Jerzy Neyman, 481–492. Berkeley: University of California Press.
- Laffont, J.-J. (1994). The new economics of regulation ten years after. *Econometrica* 62: 507–537.
- Laffont, J.-J., and D. Martimort. (2002). *The Theory of Incentives: The Principal–Agent Model*. Princeton, N.J.: Princeton University Press.
- Laffont, J.-J., and J. Tirole. (1987). Auctioning incentive contracts. *Journal of Political Economy* 95: 921–937.
- Laffont, J.-J., and J. Tirole. (1993). A Theory of Incentives in Procurement and Regulation. Cambridge, Mass.: MIT Press.
- Lazear, E. P. (1979). Why is there mandatory retirement? *Journal of Political Economy* 87: 1261–1284.
- Lazear, E. P. (1992). Compensation, productivity, and the new economics of personnel. In *Research Frontiers in Industrial Relations*, ed. D. Lewin, O. Mitchell, and P. Sherer, 341–380. Madison, Wis.: Industrial Relations Research Association.
- Lazear, E. P. (2000). Pay, performance, and productivity. *American Economic Review* 90: 1346–1361.
- Lazear, E. P. (2003). Teachers for the new century. Hoover Digest 1: 20-23.
- Lee, D. R. (1988). Free riding and paid riding in the fight against terrorism. *American Economic Review* 78: 22–26.
- Leland, H. E. (1992). Insider trading: Should it be prohibited? *Journal of Political Economy* 100: 859–887.
- Leibenstein, Harvey. (1966). Allocative efficiency vs. X-efficiency. *American Economic Review* 56: 392–415.
- Leibenstein, Harvey. (1976). *Beyond Economic Man*. Cambridge, Mass.: Harvard University Press.
- Leontief, W. W. (1936). Composite commodities and the problem of index numbers. *Econometrica* 4: 39–59.
- Levin, Jonathan, and Tadelis, Steven. (2002). A theory of partnerships. Stanford University Economic Department Working Paper. Available at http:// www.stanford.edu/~jdlevin/research.htm. Accessed on December 15, 2005.
- Lichtenberg, F. R. (1992). *Corporate Takeovers and Productivity*. Cambridge, Mass.: MIT Press.
- Litan, R. E. (1991). Comment on T. Romer and B. R. Weingast's article, "Political foundations of the thrift debacle." In *Politics and Economics in the Eighties*, ed. Alberto Alesina, Chicago: University of Chicago Press, 209–214.
- Lucking-Reiley, David. (2000). Vickrey auctions in practice: From nineteenth-century philately to twenty-first century e-commerce. *Journal of Economic Perspectives* 14: 183–192.
- Makowski, L., and J. M. Ostroy. (1987). Vickrey-Clarke-Groves mechanisms and perfect competition. *Journal of Economic Theory* 42: 244–261.
- Makowski, L., and J. M. Ostroy. (1991). The margin of appropriation and an extension of the first theorem of welfare economics. Los Angeles, Calif.: Department of Economics, U.C.L.A. working paper.
- Makowski, L., and J. M. Ostroy. (1993). General equilibrium and market socialism: Clarifying the logic of competitive markets. In *Market Socialism: The Current*

Debate, ed. P. K. Bardhan and J. E. Roemer, 69–88. New York: Oxford University Press.

- Malatesta, P., and R. Walking. (1988). Poison pill securities: Stockholder wealth, profitability, and ownership structure. *Journal of Financial Economics* 20: 347–376.
- Malkiel, B. G. (2003). A Random Walk Down Wall Street. 8th ed. New York: Norton.
- Mann, D. P., and J. P. Wissink. (1988). Money-back contracts with double moral hazard. *Rand Journal of Economics* 19: 285–292.
- Mann, D. P., and J. P. Wissink. (1990a). Hidden actions and hidden characteristics in warranty markets. *International Journal of Industrial Organization* 8: 53–71.
- Mann, D. P., and J. P. Wissink. (1990b). Money-back warranties vs. replacement warranties: A simple comparison. *American Economic Review* 80: 432–436.
- Manne, H. G. (1965). Mergers and the market for corporate control. *Journal of Political Economy* 73: 110–120.
- Manne, H. G. (1966). Insider Trading and the Stock Market. New York: The Free Press.
- Mas-Colell, A., M. D. Whinston, and J. R. Green. (1995). *Microeconomic Theory*. New York: Oxford University Press.
- Maskin, E. S. (1994). The invisible hand and externalities. *American Economic Review* 84: 333–337.
- Maskin, E. S. (2003). Auctions and efficiency. Pp. 1–24 in Dewatripont, Hansen, and Turnovsky (2003).
- Maskin, E., and T. Sjöström. (2002). Implementation theory. In *Handbook of Social Choice and Welfare,* Vol. 1, ed. K. J. Arrow, A. K. Sen, and K. Suzumura, Chapt. 5. Amsterdam: Elsevier.
- McAfee, R. P. (2002). *Competitive Solutions*. Princeton, N.J.: Princeton University Press.
- McAfee, R. P., and John McMillan. (1988). *Incentives in Government Contracting*. Toronto, Canada: University of Toronto Press.
- McConnell, J. J., and C. J. Muscarella. (1986). Corporate capital expenditure decisions and the market value of the firm. *Journal of Financial Economics* 14: 399–422.
- McKelvey, R. D., and T. R. Palfrey. (1992). An experimental study of the centipede game. *Econometrica* 60: 803–836.
- McKenzie, L. W. (1954). On equilibrium in Graham's model of world trade and other competitive systems. *Econometrica* 22: 147–161.
- McKenzie, L. W. (2002). *Classical General Equilibrium Theory*. Cambridge, Mass.: MIT Press.
- McMillan, John. (1992). *Games, Strategies, and Managers*. Oxford, UK: Oxford University Press.
- McMillan, John. (1994). Selling spectrum rights. *Journal of Economic Perspectives* 8: 13–29.
- McMillan, John. (1997). Markets in transition. In *Advances in Economics and Econometrics: Theory and Applications*, Vol. 2, ed. D. M. Kreps and K. F. Wallis. Cambridge, UK: Cambridge University Press, 210–239.
- McMillan, John. (2002). *Reinventing the Bazaar: A Natural History of Markets*. New York: W.W. Norton.
- McPhee, J. (1966). Oranges. New York: Farrar, Straus, and Giroux.
- Megginson, W. L., and J. M. Netter. (2001). From state to market: A survey of empirical studies on privatization. *Journal of Economic Literature* 39: 321–389.

- Milgrom, P. R. (1987). Auction theory. In *Advances in Economic Theory Fifth Worlds Congress*, ed. T. Bewley. Cambridge, UK: Cambridge University Press, 1–32.
- Milgrom, P. R. (1989). Auctions and bidding: A primer. *Journal of Economic Perspectives* 3: 3–32.
- Milgrom, P. R. (2004). *Putting Auction Theory to Work*. Cambridge, UK: Cambridge University Press.
- Milgrom, P. R., and J. Roberts. (1992). *Economics, Organization, and Management*. Englewood Cliffs, N.J.: Prentice Hall.
- Milleron, J.-C. (1972). Theory of value with public goods: A survey article. *Journal of Economic Theory* 5: 419–477.
- Mirlees, J. (1971). An exploration in the theory of optimal income taxation. *Review* of *Economic Studies* 38: 175–208.
- Mirlees, J. (1974). Notes on welfare economics, information, and uncertainty. In *Essays in Economic Behavior under Uncertainty*, ed. M. Balch, D. McFadden, and S. Wu, 243–258. Amsterdam: North Holland.
- Mirlees, J. (1976). The optimal structure of incentives and authority within an organization. *Bell Journal of Economics* 7: 105–131.
- Mirlees, J. (1999). The theory of moral hazard with unobservable behavior, part 1. *Review of Economic Studies* 66: 3–22.
- Mishel, L., and J. Bernstein. (1992). Declining wages for high school and college graduates. Economic Policy Institute working paper. Washington, D.C.: Economic Policy Institute.
- Mishkin, F. S. (1992). *Money, Banking, and Financial Markets*. New York: Harper Collins.
- Moe, Terry. (2001). Hidden demand. Education Next 1, Spring: 48–55.
- Molho, Ian. (1997). *The Economics of Information: Lying and Cheating in Markets and Organizations*. Oxford, UK: Blackwell Publishers.
- Moore, J., and R. Repullo. (1988). Subgame perfect implementation. *Econometrica* 56: 1191–1220.
- Murphy, K. J. (1999). Executive compensation. In *Handbook of Labor Economics*, ed. O. Ashenfelter and D. Card. Amsterdam: Elsevier, 2485–2563.
- Murphy, K. J. (2000). Explaining executive compensation: Managerial power versus the perceived cost of stock options. *University of Chicago Law Review* 69: 847–869.
- Myerson, R. B. (1981). Optimal auction design. *Mathematics of Operations Research* 6: 58–73.
- Myerson, R. B. (1999). Nash equilibrium and the history of economic theory. *Journal of Economic Literature* 37: 1067–1082.
- Nalebuff, Barry, and Ian Ayres. (2003). *Why Not?* Boston, Mass.: Harvard Business School Press.
- Nasar, Sylvia. (1998). A Beautiful Mind. New York: Simon and Schuster.
- Neufeld, J. L., W. J. Hausman, and R. B. Rapoport. (1994). A paradox of voting: Cyclical majorities and the case of Muscle Shoals. *Political Research Quarterly* 47: 423–438.

Novshek, W. (1993). Mathematics for Economists. San Diego, Calif.: Academic Press.

Olson, Mancur. (1993). Why is economic performance even worse after communism is abandoned? Unpublished mimeographed material. Fairfax, Va.: George Mason University, Center for the Study of Public Choice.

- O'Neil, Thomas. (1997). Australia's dog fence. National Geographic 191 (4): 18-37.
- Osborne, M. J. (2004). *An Introduction to Game Theory*. New York: Oxford University Press.
- Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press.
- Perry, T., and M. Zenner. (2000). CEO compensation in the 1990s: Shareholder alignment or shareholder expropriation? *Wake Forest Law Review* 35: 142–144.
- Petrosky, Henry. (1992). *To Engineer Is Human: The Role of Failure in Successful Design*. New York: Vintage Books.
- Petrosky, Henry. (1996). *Invention by Design: How Engineers Get from Thought to Thing*. Cambridge, Mass.: Harvard University Press.
- Phlips, Louis. (1981). *The Economics of Price Discrimination*. New York: Cambridge University Press.
- Poundstone, W. (1992). Prisoner's Dilemma. New York: Doubleday.
- Prendergast, Canice. (1999). The provision of incentives in firms. *Journal of Economic Literature* 37: 7–63.
- Prowse, Stephen. (1995). Corporate governance in an international perspective: A survey of corporate control mechanisms among large firms in the U.S., U.K., Japan, and Germany. *Financial Markets, Institutions and Instruments* 4: 1–63.
- Radner, R. (1991). Dynamic games in organization theory. *Journal of Economic Behavior and Organization* 16: 217–260.
- Radner, R. (1992). Hierarchy: The economics of managing. *Journal of Economic Literature* 30: 1382–415.
- Rajan, R. G., and Zingales, L. (2000). The governance of the new enterprise. In *Corporate Governance*, ed. Xavier Vives, 201–227. Cambridge, UK: Cambridge University Press.
- Rapoport, A. (1989). Prisoner's dilemma. In *Game Theory*, ed. J. Eatwell, M. Milgate, and P. Newman, 199–204. New York: Norton.
- Reinhardt, U. E., P. S. Hussey, and G. F. Anderson. (2004). U.S. health care spending in an international context. *Health Affairs* 23: 10–25.
- Repullo, R. (1987). A simple proof of Maskin's theorem on Nash implementation. *Social Choice and Welfare* 4: 39–41.
- Riley, J. G. (1989). Signalling. In Allocation, Information, and Markets, ed. J. Eatwell, M. Milgate, and P. Newman, 287–294. New York: Norton.
- Riley, J. G. (2000). Silver signals: Twenty-five years of screening and signaling. *Journal* of *Economic Literature* 39: 432–478.
- Riley, J. G., and W. Samuelson. (1981). Optimal auctions. *American Economic Review* 71: 381–392.
- Rob, R. (1982). Asymptotic efficiency of the demand revealing mechanism. *Journal* of *Economic Theory* 28: 207–220.
- Roberts, D. J. (1979). Incentives in planning procedures for the provision of public goods. *Review of Economic Studies* 46: 283–292.
- Roberts, J., and A. Postlewaite. (1976). The incentive for price-taking behavior in large exchange economies. *Econometrica* 44: 115–127.
- Romer, T., and B. R. Weingast. (1991). Political foundations of the thrift debacle. In *Politics and Economics in the Eighties*, ed. Alberto Alesina and G. Carliner. Chicago: University of Chicago Press, 175–209.

- Rosenthal, R. (1981). Games of perfect information, predatory pricing, and the chain store paradox. *Journal of Economic Theory* 25: 92–100.
- Rosovsky, Henry. (1990). The University: An Owner's Manual. New York: Norton.
- Ross, S. (1973). The economic theory of agency: The principal's problem. *American Economic Review* 63: 134–139.
- Roth, A. E. (1982). The economics of matching: Stability and incentives. *Mathematics* of Operations Research 7: 617–628.
- Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy* 92: 991–1026.
- Roth, A. E. (1990). New physicians: A natural experiment in market organization. *Science*, December 14: 1524–1528.
- Roth, A. E., and E. Peranson. (1999). The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review* 89: 748–780.
- Roth, A. E., T. Sönmez, and U. Ünver. (2004). Kidney exchange. *Quarterly Journal of Economics* 119: 457–488.
- Roth, A. E., and M. A. O. Sotomayor. (1990). *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*. Cambridge, UK: Cambridge University Press.
- Roth, A. E., and X. Xing. (1994). Jumping the gun: Imperfections and institutions relating to the timing of market transactions. *American Economic Review* 84: 992–1044.
- Rothschild, M., and J. E. Stiglitz. (1976). Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *Quarterly Journal of Economics* 80: 629–649.
- Rubinstein, A. (1986). Finite automata play the repeated prisoners' dilemma. *Journal* of *Economic Theory* 39: 83–96.
- Rubinstein, A. (1998). Modeling Bounded Rationality. Cambridge, Mass.: MIT Press.
- Ryngaert, M. (1988). The effect of poison pill securities on shareholder wealth. *Journal of Financial Economics* 20: 377–417.
- Saari, D. G. (1994). Geometry of Voting. Berlin: Springer–Verlag.
- Samuelson, P. A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics* 36: 387–389.
- Samuleson, W. (1984). Bargaining under asymmetric information. *Econometrica* 52: 995–1005.
- Samuelson, W. (1985). A comment on the Coase conjecture. In *Game-Theoretic Models of Bargaining*, ed. A. Roth. Cambridge, UK: Cambridge University Press, 321–339.
- Sandler, T. (1992). *Collective Action: Theory and Applications*. Ann Arbor: University of Michigan Press.
- Sappington, D. E. M. (1991). Incentives in principal–agent relationships. *Journal of Economic Perspectives* 5: 45–66.
- Sappington, D. E. M. (1993). Designing incentive regulation. University of Florida, College of Business Administration working paper no. 93-94-10.
- Sappington, D. E. M., and D. L. Weisman. (1996). *Designing Incentive Regulation for the Telecommunications Industry*. Cambridge, Mass.: MIT Press.
- Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory* 10: 187–217.

- Scherer, F. M. (1988). Corporate takeovers: The efficiency arguments. Journal of Economic Perspectives 2: 69–82.
- Schmalensee, Richard, P. L. Joskow, A. D. Ellerman, J. P. Montero, and E. M. Bailey. (1998). An interim evaluation of sulphur dioxide emissions trading. *Journal of Economic Perspectives* 12: 53–68.
- Schmalensee, Richard, T. M. Stoker, and R. A. Judson. (1998). World carbon dioxide emissions: 1950–2050. *Review of Economics and Statistics* 80: 15–27.
- Scully, G. W. (1989). *The Business of Major League Baseball*. Chicago: University of Chicago Press.
- Sen, A. K. (1966). A possibility theorem on majority decisions. *Econometrica* 34: 491–499.
- Sen, A. K. (1981). Poverty and Famines: An Essay on Entitlement and Deprivation. Oxford, UK: Oxford University Press.
- Sen, A. K. (1993). The economics of life and death. *Scientific American* May: 40–47.
- Sen, Arunava. (2001). Another direct proof of the Gibbard–Satterthwaite theorem. *Economics Letters* 70: 381–385.
- Sheffrin, S. M. (1993). Markets and Majorities. New York: The Free Press.
- Shiller, R., and A. N. Weiss. (1994). Home equity insurance. Cowles Foundation discussion paper no. 1074, Yale University. Available at http://cowles.econ.yale. edu/P/ab/a10/a1074.htm
- Shleifer, Andrei. (2000). Discussion. In *Corporate Governance*, ed. Xavier Vives, 134– 136. Cambridge, UK: Cambridge University Press.
- Shleifer, A., and R. W. Vishney. (1986). Large shareholders and corporate control. *Journal of Political Economy* 294: 461–481.
- Shleifer, A., and R. W. Vishney. (1988). Value maximization and the acquisition process. *Journal of Economic Perspectives* 2: 7–20.
- Shleifer, A., and R. W. Vishney. (1997). A survey of corporate governance. *Journal of Finance* 52: 737–783.
- Shoven, J. B., S. B. Smart, and J. Waldfogel. (1992). Real interest rates and the savings and loan crisis: The moral hazard premium. *Journal of Economic Perspectives* 6: 155–167.
- Singh, Simon. (1999). *The Code Book: The Science of Secrecy from Ancient Egypt to Quantum Cryptography*. New York: Random House.
- Sönmez, Tayfun, and M. U. Ünver. (2003). Course bidding at business schools. Available at http://www.stanford.edu/group/SITE/Unver.pdf. Accessed on December 15, 2005.
- Spence, A. M. (1973). *Market Signalling: Information Transfer in Hiring and Related Processes*. Cambridge, Mass.: Harvard University Press.
- Stavins, R. N. (1998). What can we learn from the grand policy experiment? Lessons from SO₂ allowance trading. *Journal of Economic Perspectives* 12: 69–88.
- Stern, S., and P. Todd. (1992). A test of Lazear's mandatory retirement model. Discussion paper no. 236. Charlottesville: University of Virginia, The Thomas Jefferson Center for Political Economy.
- Sternberg, T. (1991). FDIC auctions of failed banks: One way to measure the acquiror's surplus. Berkeley: University of California, Haas School of Business.
- Stiglitz, J. E. (1974). Incentives and risk sharing in sharecropping. *Review of Economic Studies* 41: 219–255.

- Stiglitz, J. E. (1975). Incentives, risk, and information: Notes toward a theory of hierarchy. *Bell Journal of Economics* 6: 552–579.
- Stiglitz, J. E. (1993). Market socialism and neoclassical economics. In *Market Socialism: The Current Debate*, ed. P. K. Bardhan and J. E. Roemer, 21–41. New York: Oxford University Press.
- Stiglitz, J. E. (2000). The contributions of the economics of information to twentiethcentury economics. *Quarterly Journal of Economics* 115: 144–1178.
- Stiglitz, J. E., and A. Weiss. (1981). Credit rationing in markets with imperfect information. *American Economic Review* 71: 393–410.
- Strang, Gilbert. (1991). Calculus. Wellesley, Mass.: Wellesley-Cambridge Press.
- Su, Xuanming, and S. A. Zenios. (2003). Mechanism design for kidney allocation. Available at http://pages.stern.nyu.edu/~varaman/Xuanming_paper.pdf. Accessed on December 15, 2005.
- Sutton, J. (2000). *Marshall's Tendencies: What Can Economists Know?* Cambridge, Mass.: MIT Press.
- Thaler, R. H. (1992). *The Winner's Curse: Paradoxes and Anomalies of Economic Life*. New York: The Free Press.
- Thompson, D. N. (1971). *Franchise Operations and Antitrust*. Lexington, Mass.: D.C. Heath.
- Tideman, T. N., and G. Tullock. (1976). A new and superior principle for collective choice. *Journal of Political Economy* 84: 1145–1159.
- Tirole, J. (1988). The Theory of Industrial Organization. Cambridge, Mass.: MIT Press.
- Ullen, T. S. (1994). Rational victims—rational injurers: Cognition and the economic analysis of tort law. Unpublished mimeographed material, Department of Economics, University of Illinois at Urbana–Champaign.
- Varian, H. R. (2000). Differential pricing and efficiency. First Monday June: 1-6.
- Vickrey, W. (1945). Measuring marginal utility by reactions to risk. *Econometrica* 13: 319–333.
- Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance* 16: 8–37.
- Vives, Xavier. (2000). Corporate governance: Does it matter? In *Corporate Governance*, ed. Xavier Vives, 1–21. Cambridge, UK: Cambridge University Press.
- Von Neumann, J., and O. Morgenstern. (1944). *Theory of Games and Economic Behavior*. Princeton, N.J.: Princeton University Press.
- Votruba, M. E. (2003). *Essays on the allocation of public resources*. Doctoral dissertation, Department of Economics, Princeton University.
- Waldrup, M. M. (1992). Complexity. New York: Simon and Schuster.
- Waldspurger, C. A., T. Hogg, B. A. Huberman, J. O. Kephart, and S. Stornetta. (1992). Spawn: A distributed computational economy. *IEEE Transactions on Software Engineering* 18: 103–117.
- Walker, M. (1978). A note on the characterization of mechanisms for the revelation of preferences. *Econometrica* 46: 147–152.
- Walker, M. (1980). On the nonexistence of a dominant strategy mechanism for making optimal public decisions. *Econometrica* 48: 1521–1540.
- Walker, M. (1981). A simple incentive compatible scheme for attaining Lindahl allocations. *Econometrica* 49: 65–71.
- Weinberg, Samantha. (2003). Pointing from the Grave. New York: Hyperion.

- Weintraub, E. R. (1982). *Mathematics for Economists*. Cambridge, UK: Cambridge University Press.
- Wen, Quan. (1994). The "folk theorem" for repeated games with complete information. *Econometrica* 62: 949–954.
- Wheelan, Charles. (2002). *Naked Economics: Undressing the Dismal Science*. New York: W. W. Norton.
- White, Lawrence J. (1991). *The S & L Debacle: Public Policy Lessons for Bank and Thrift Regulation*. New York: Oxford University Press.
- White, M. J. (1999). What's wrong with U.S. personal bankruptcy law and how to fix it. *Regulation* 22. Available at http://www.cato.org/pubs/regulation/regv22n3/reg22n3.html. Accessed on December 15, 2005.
- Williams, S. R., and R. Radner. (1995). Efficiency in partnership when the joint output is uncertain. In *The Economics of Informational Decentralization*, ed. J. Ledyard, 79–99. Boston: Kluwer.
- Woessman, Ludger. (2001). Why students in some countries do better. *Education Next* 1, Summer: 67–74.
- Wolak, F. A. (2002). Making sense of the Enron nonsense. Stanford Institute for Economic Policy Research policy brief, Stanford University. Available at http://stanford.edu/papers/briefs/policybrief_may02.html. Accessed on December 15, 2005.
- Yermack, David. (1997). Good timing: The CEO stock option awards and company news announcements. *Journal of Finance* 52: 449–476.
- Zeckhauser, R. (1970). Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives. *Journal of Economic Theory* 2: 10–26.
- Zhou, Lin (1990). On a conjecture by Gale on one-sided matching problems. *Journal* of *Economic Theory* 52: 123–135.
- Zimbalist, Andrew. (2004). *May the Best Team Win: Baseball Economics and Public Policy*. Washington, D.C.: Brookings Institution Press.

Author Index

Abdulkadiroğlu, Atila, 494, 502, 509, 561 Akerlof, George A., 15, 163, 224, 271, 302, 322, 561 Alchian, A., 206, 212, 561 Allen, D. A., 7, 561 Allen, Franklin, 198, 218, 225, 561 Anderson, G. F., 185, 194, 573 Anderson, L. R., 436, 561 Anderson, R. M., 542, 561 Aoki, Masahiko, 206, 561 Arrow, Kenneth J., 7, 15, 289, 334, 400, 436, 523, 536, 561-562 Ashenfelter, O., 347, 562 Attiveh, G., 450, 562 Axelrod, Robert, 55, 69, 562 Ayres, Ian, 194, 226, 334, 562, 572 Babcock, L., 18, 563 Bailey, E. M., 151, 568, 575 Baker, George, 251, 562 Balazs, E., 523, 562 Balinski, Michel, 487, 494, 562 Ballou, Dale, 179, 562 Banerjee, A., 221, 562 Barberà, S., 409, 562 Basu, Kaushik, 44, 562 Baumol, W. J., 9, 43, 198, 518, 523, 562 Bebchuk, L. A., 218, 220, 224, 229, 231, 563 Bernstein, J., 297, 572 Bernstein, P., 230, 297, 563 Bertrand, M., 226, 229, 230, 563 Binmore, K., 44, 51, 69, 86, 97, 333, 563 Blair, R. D., 512, 563 Borda, J.-C., 394, 399, 563 Brams, S. J., 390, 399, 563 Brickley, J. A., 218, 220, 222, 230, 563, 568 Burns, S., 194, 569

Calvert, R. L., 69, 563 Camerer, C. F., 18, 563 Cameron, S., 302, 563 Campbell, D. E., 112, 400, 536, 563 Carlin, Wendy, 198, 563 Carmichael, H. L., 173, 176-177, 231, 564 Carr, Jack L., 158, 564 Chen, Yan, 465, 494, 509, 564 Chiappori, P.-A., 7, 564 Chichilnisky, G., 151, 564 Chou, Y. K., 18, 564 Clarke, E. H., 451, 564 Coase, R. H., 151, 430, 564 Coles, J. L., 218, 563 Comment, Robert, 223, 564 Condorcet, Marquis de, 399, 564 Conley, J., 112, 564 Cook, C., 244, 565 Cooter, R. D., 347, 565 Corchón, L. C., 465, 565 Cornes, R., 436, 565 Costa, D. L., 177, 565 Cramton, P., 44, 148, 151, 374, 565 d'Aspremont, C., 334, 565 Dana, J. D., 259, 565 Danilov, V. I., 268, 457, 565 Davies, J., 86, 97, 563 Debreu, Gerard, 523, 536, 565 Demirgüç-Kunt, A., 162, 565 Demsetz, H. D., 206, 212, 334, 565 Dewatripont, Mathias, 162, 163, 565 Diamantaris, D., 112, 564 Diamond, Peter, 194, 565 Dixit, A. K., 251, 565 Downs, A., 51, 565 Dranove, David, 183, 194, 565 Dray, Phillip, 7, 565 Drèze, J., 523, 565 Dutta, P. K., 194, 223, 565 Dyer, D., 381, 565

Easterbrook, F. H., 219, 230, 566 Eckard, E. W., 221, 562 Ehrenberg, R. G., 177, 566 Ellerman, A. D., 151, 575 Ergin, Haluk, 494, 566 Farber, H. S., 18, 566 Farrell, J., 268, 566 Fay, S., 165, 566 Foley. D., 465, 566 Franciosi, R., 450, 562 Frank, R. H., 44, 51, 566 Freeman, R. B., 227, 566 Freixas, Xavier, 279, 566 Fried, J. M., 218, 220, 224, 229, 231, 563 Friedman, J., 69, 566 Fudenberg, D., 69, 566 Gale, Douglas, 198, 218, 225, 479, 494, 561, 566 Gerard-Varet, L.-A., 334, 565 Gibbard, Alan, 409, 566 Gibbons, Robert, 44, 69, 251, 562, 565, 566 Gintis, Herbert, 536, 566 Goeree, J., 44, 566, 568 Green, J. R., 115, 347, 451, 457, 566, 571 Gribbin, John, 6, 566 Gribbin, Mary, 6, 566 Grossman, S. J., 219, 220, 567 Groves, T., 451, 457, 567 Gruber, J., 182, 567 Hahn, F. H., 566 Hall, B. H., 230, 567 Hall, B. J., 227, 448, 567 Hall, J., 217, 567 Hammons, C. W., 522, 567 Hansen, L., 565 Hansmann, Henry, 219, 567 Hanushek, E. A., 179, 567 Harrison, B., 230, 567 Hart, O., 219, 220, 567 Haubrich, Joseph, 217, 567 Hausman, W. J., 399, 572 Heal, G., 151, 564 Heckman, J., 302, 563 Hendel, I., 272, 567 Herstein, I. N., 115, 567 Hicks, J. R., 102, 567 Hill, Christopher, 34, 567 Hirshleifer, J., 322, 567 Hogg, T., 576 Holmström, B., 220, 228, 251, 347, 567 Holt, C. A., 44, 568 Hoxby, Caroline M., 179, 568 Hsaio, W. C., 182, 568 Huberman, G., 176, 568, 576 Hurwicz, Leonid, 347, 465, 555-556, 568 Hussey, P. S., 185, 194, 573

Illing, G., 372, 568 Isaac, M., 450, 562 Jackson, M., 409, 556, 568 Jaffee, Dwight, 279, 568 Jarrell, G. A., 220, 222, 230, 568 Jensen, M. C., 217, 219, 220, 226, 230, 568 Joskow, P. L., 151, 568, 575 Judson, R. A., 151, 575 Kagel, J. H., 381, 383, 565, 568 Kahn, C., 176, 568 Kahn, M. A., 542, 561 Kandel, E., 206, 568 Kane, E. J., 162, 565, 569 Kanter, R. M., 230, 569 Kaplan, S. N., 220, 228, 567 Kaserman, D. L., 512, 563 Katzner, D. W., 112, 569 Kelly, J. S., 400, 409, 563, 569 Kephart, J. O., 576 Kerr, S., 148, 151, 565 Kirby, K. N., 334, 569 Kleiner, M. M., 227, 566 Klemperer, Paul, 44, 333, 372, 374, 377, 563, 565, 569 Klüh, U., 372, 568 Kmietowicz, Zosia, 535, 569 Kohler, Heinz, 199, 569 Koopmans, T. C., 97, 142, 569 Kotlikoff, L. J., 194, 569 Kotowitz, Y., 230, 569 Krauss, L. M., 8, 569 Kreps, D. M., 44, 69, 115, 569 Krishna, Vijay, 357, 569 Krueger, A. B., 227, 297, 569 Kuhn, H. W., 97, 570 Laffont, J.-J., 7, 9, 15, 251, 334, 347, 451, 457, 566, 570 Lazear, E. P., 173, 179, 194, 206, 227, 570 Lee, D. R., 43, 570 Leibenstein, Harvey, 155, 570 Leland, H. E., 230, 570 Levin, D., 383, 568 Levin, Jonathan, 206, 570 Levitt, S. D., 194, 562 Lichtenberg, F. R., 220, 570 Liebman, J., 217, 567 Litan, R. E., 162, 570 Lizzeri, Alessandro, 272, 567 Lowenstein, G., 18, 563 Lucking-Reiley, David, 357, 570 Lueck, D., 7, 561 Makowski, L., 347, 570

Malatesta, P., 223, 571 Malkiel, B.G., 221, 228, 571 Mann, D. P., 257, 571

Author Index

Manne, H. G., 221, 571 Martimort, D., 7, 15, 251, 570 Mas. A., 227, 569 Mas-Colell, A. M. D., 115, 571 Maskin, E. S., 69, 268, 333, 383, 409, 556, 566.571 Mathewson, Frank, 158, 564 Mayer, Colin, 198, 563 McAfee, R. P., 7, 302, 378, 383, 571 McConnell, J. J., 226, 571 McKelvey, R. D., 38, 571 McKenzie, Lionel W., 536, 571 McLeod, B. W., 231, 564 McMillan, John, 7, 9, 224, 251, 333, 334, 377, 571 McPhee, J., 236, 571 Megginson, W. L., 333, 571 Milgrom, Paul R., 6, 160, 162, 206, 216, 224, 225-226, 244, 312, 333, 347, 377, 522, 569, 572 Milleron, J.-C., 465, 572 Milnor, J., 115, 567 Mirlees, James, 15, 251, 347, 572 Mishel, L., 297, 572 Mishkin, F. S., 162, 163, 572 Moe, Terry, 177, 572 Molho, Ian, 272, 322, 572 Montero, J. P., 151, 575 Moore, J., 556, 572 Morgenstern, O., 115, 576 Mullainathan, S., 226, 229, 230, 563 Murphy, Kevin J., 217, 227, 230, 231, 251, 562, 567, 568, 572 Muscarella, C. J., 226, 571 Myerson, R. B., 43, 372, 572

Nalebuff, Barry, 194, 226, 334, 572 Nasar, Sylvia, 43, 572 Netter, J. M., 220, 222, 230, 333, 568, 571 Neufeld, J. L., 399, 572 Novshek, W., 86, 97, 572

O'Neil, Thomas, 423, 573 Olson, Mancur, 251, 572 Osborne, M. J., 44, 51, 69, 573 Ostrom, E., 560, 573 Ostroy, J. M., 347, 570 Owings, M., 182, 567

Palfrey, T. R., 38, 571 Peranson, E., 499, 574 Perry, T., 217, 573 Petrosky, Henry, 3, 573 Phlips, Louis, 259, 573 Podgursky, Michael, 179, 562 Postlewaite, A., 556, 573 Poundstone, W., 51, 573 Prendergast, Canice, 251, 573 Prowse, Stephen, 220, 573 Quigley, N. C., 158, 564 Radner, R., 206, 230, 573, 577 Rajan, R. G., 227, 573 Rapoport, Anatol, 55, 69, 573 Rapoport, R. B., 399, 572 Rashid, S., 542, 561 Reinhardt, U. E., 185, 194, 573 Repullo, R., 465, 556, 572, 573 Riley, J. G., 302, 322, 372, 567, 573 Rob, R., 451, 573 Roberts, D. J., 6, 160, 162, 206, 216, 224, 225-226, 244, 312, 347, 465, 522, 569, 572, 573 Rochet, J.-C., 279, 566 Roemer, J.E., 562 Romer, T., 162, 163, 224, 561, 573 Rosenthal, Robert, 43, 574 Rosovsky, Henry, 177, 574 Ross, S., 15, 251, 574 Roth, A. E., 333, 479, 494, 499, 512, 574 Rothschild, M., 322, 574 Rubinstein, A., 69, 574 Ryngaert, M., 51, 223, 574 Saari, D. G., 409, 574 Salanié, B., 7, 564 Samuelson, L., 436, 563 Samuelson, P.A., 436, 574 Samuelson, W., 69, 268, 372, 573, 574 Sandler, T., 436, 565, 574 Santiesteban, M., 334, 569 Sappington, D. E. M., 9, 155, 251, 574 Satterthwaite, Mark A., 409, 574 Scherer, F. M., 220, 575 Schmalensee, Richard, 151, 568, 575 Schmeidler, D., 556, 568 Schwartz, J. A., 374, 565 Schwert, G. W., 223, 564 Scully, G. W., 575 Sen, A. K., 397, 399, 516, 522, 523, 565, 575 Sen, Arunava, 417, 575 Shapley, L. S., 479, 494, 566 Sheffrin, S. M., 122, 575 Shleifer, Andrei, 214, 218, 220, 223, 230, 575 Shoven, J. B., 163, 575 Singh, Simon, 257, 575 Sjöström, T., 409, 556, 571 Smart, S. B., 163, 575 Sönmez, Tayfun, 333, 485, 487, 494, 502, 509, 512, 561, 562, 564, 566, 574, 575 Sotomayor, M. A. O., 479, 574 Sotskov, A. I., 268, 457, 565 Spence, A. Michael, 15, 271, 302, 322, 575 Stavins, R. N., 151, 575

Stern, S., 173, 575

Sternberg, T., 347, 374, 575

Stiglitz, Joseph E., 9, 15, 138, 143, 251, 271, 279, 302, 322, 536, 568, 574, 575–576

Stoker, T. M., 151, 575 Stornetta, S., 576 Strang, Gilbert, 86, 576 Su, Xuanming, 512, 576 Sutton, J., 18, 576

Tadelis, Steven, 206, 570 Terry, R. L., 218, 563 Thaler, R. H., 18, 381, 563, 576 Thompson, D. N., 243, 576 Tideman, T. N., 451, 576 Tirole, Jean, 162, 163, 251, 334, 565, 570, 576 Todd, P., 173, 575 Truchon, M., 112, 563 Tucker, Albert W., 51, 97, 570 Tullock, G., 451, 576 Turnovsky, S., 565

Ullen. T. S., 347, 565, 576 Ünver, M. U., 333, 485, 512, 574, 575

Varian, H. R., 259, 576 Vickrey, W., 15, 21, 347, 372, 576 Vishny, R. W., 218, 220, 223, 230, 575 Vives, Xavier, 198, 224, 576 Von Neumann, J., 115, 576 Votruba, M. E., 512, 576

Waldfogel, J., 163, 575 Waldrup, M. M., 347, 576 Waldspurger, C. A., 340, 347, 576 Walker, D. I., 218, 220, 224, 229, 231, 563 Walker, M., 347, 457, 465, 568, 576 Walking, R., 223, 571 Weinberg, Samantha, 14, 576 Weingast, B. R., 162, 573 Weintraub, E. R., 97, 577 Weisman, D. L., 9, 155, 574 Weiss, A. N., 279, 575, 576 Wen, Quan, 69, 577 Wheelan, Charles, 13, 184, 214, 523, 577 Whinston, A. M. D., 115, 571 Whinston, G. C., 334, 569 White, Lawrence J., 577 White, M. J., 162, 566, 577 Williams, S. R., 206, 577 Wilson, R., 569 Wissink, J. P., 257, 571 Woessman, Ludger, 179, 577 Wolak, F. A., 228, 577

Xing, X., 494, 574

Zeckhauser, R., 194, 577 Zenios, S. A., 512, 576 Zenner, M., 217, 573 Zhou, Lin, 509, 577 Zimbalist, Andrew, 12, 378, 577 Zingales, L., 227, 573

abstract models, 7-8 accidents, 180, 184, 194, 304-5, 307; probability of, 131-2, 133, 303, 313. See also insurance acid rain, 18-22, 327. See also sulfur dioxide adjustment costs, 137, 454; pollution abatement and, 18-21, 144-6 adverse selection, 13, 14-15, 162, 178, 184 advertising, 244, 292 African Americans, lynching of, 7 Age Discrimination in Employment Act (1978), 165agency problem, 197, 213. See also principal-agent problems agency theory, 231-55; basic model, 237-9; diversification and, 231-2; effort supply function and, 238-9; optimal contract and, 237, 239–44; participation constraint and, 237; risk-neutral managers, 239-41 agents, 370; in matching problems, 468, 472 agricultural reform, in China, 7, 235 airplane travel, 3, 13, 34, 180, 258 allocations, 26-8, 35-6, 104-5, 328; of dormitory rooms, 499-510; efficient, 424, 426-7, 430, 501-2, 503-5, 507; resource, 12, 139-42, 338-9, 454-5, 550. See also matching problems amusement park pricing, 235 Anglo-Dutch auction, 374, 377 antibiotics, resistance to, 535 arbitrage, 227, 265 arithmetric progression, 371 Arrow-Debreu economy, 523-38, 543, 548; budget constraint in, 524, 527, 529, 530; endowment and, 524-5; externalities in, 533-5; feasibility in, 526-7, 530, 532; profit calculation in, 526; welfare theorem and, 529-31,534 Arrow-Pratt measure, 124

asset, 11, 112-13, 157, 379, 448, 455; bankruptcy and, 164-5; corporate, 212; value of, 266-8 asset efficiency, 260, 262-3, 264; in auctions, 328, 333, 341, 342, 345, 378, 379 assignment problem, 468, 469, 499. See also matching problems asymmetric information, 10-15, 267, 272, 285-9, 297-8, 302; encryption and, 257; in insurance, 312–20, 322; in used-car market, 269, 271 athletes and athletics. See sports Atlantic Salmon Federation, 560 atomic bomb research, 6 AT&T, 153, 517 auctions, 148, 326–34; all-pay, 364; bargaining in, 331-3; Dutch auction, 329, 349, 350, 352–3, 356, 361, 374; efficiency and, 329-33, 341; English (oral) auction, 335, 349-51, 361, 374-5, 377, 383; equilibrium in, 335-7, 353-6, 359, 363-4; expected revenue for, 362-3; general mechanisms, 344, 345; interdependent values, 377-83; multistage, 375; outcome equivalence in, 351-2; probability in, 354-5, 363; procurement, 376, 377; radio spectrum, 327, 333, 372, 374, 375, 377; reservation values in, 326, 330-2; revenue equivalence in, 358-77, 379-80; standard auction mechanism, 359, 375-6; ten significant, 327-9; winner's curse in, 380-3. See also first-price, sealed-bid auction; Vickrey auction Australia, 226, 423 automobiles, 5, 48, 214, 225, 280, 340; auctions of, 377; drunk driving, 8; insurance for, 12, 162, 180, 183, 303-4; repair of, 6; taxis, 16-18, 517; theft of, 186; used-car market, 269-72; Volkswagen, 287, 328; warranty for, 138 average cost taxation, 432-3, 464-5

583

balance of trade, 140 bankruptcy, 164-5, 218, 227, 228 banks and banking, 158, 197, 221, 226; auction of assets, 328, 329, 333-4, 374; savings and loan debacle, 155-63 barbers, and haircut prices and, 258 bargaining, 223, 261-3, 265-7, 331-3, 453-7 baseball, 12, 40, 378, 381 beggar-thy-neighbor policies, 49 benefit taxation, 433-6, 446 Bethlehem Steel, 165 beverage cans, 518 bidder signal, 378. See also auctions binary choice model, 188, 244, 454 boards of directors, 198, 213, 218, 223, 229 bonds, 161, 166, 171, 517, 523 bonus, 198, 215-16, 217, 222, 225, 230 Borda, J.-C., 394 Borda (rank order) rule, 394-6 borrowing, 155, 164-5, 274-7 Boston, 178, 186 Boston mechanism, 494 British Telecom, 153 budget balance, 9, 265, 457, 462 budget constraint, 5, 72-3, 521, 522, 547, 560; in Arrow-Debreu economy, 524, 527, 529, 530; consumer choice and, 73, 76, 79, 80; preferences and, 99; quasi-linear utility and, 106, 108; retirement and, 171-3; worker contract and, 209 budget surplus, 446, 447 bundling, 318; quality and, 280-90; self-selection constraints and, 285, 286-7 burglary, 113; insurance against, 185-6 Bush administration, 9 buyer-seller relations. See two-person exchange

Caesarian section, 182 calculus, 126, 292, 371-2; overview of, 76-86 California, 209, 516 Canada, 158, 182, 197, 198 capitalist markets, 518 capital market, 213, 219-21 capital outlay, 515, 535 carbon dioxide (CO₂) emissions, 12, 148, 421. See also pollution Cardozo Law School, 14 cars. See automobiles cartels, 49-50, 375 casualty insurance companies, 448 catastrophe bonds, 517, 523 celebrity, 224, 292 centipede game, 37-40, 43 CEO (chief executive officer), 198, 215, 217, 225, 228, 229; compensation for, 214; golden parachute for, 220, 226; self-dealing by, 223-4. See also owner-management relationship

chain rule, 85-6 charter schools, 178-79 Chicago, 45, 327 China, 7, 235, 328, 518-19 Chrysler Corporation, 560 civil litigation, 340-1, 376 coalitions, 399, 447-9 Cobb-Douglas utility, 528, 550, 553 college admissions, 438, 480-96, 517; fairness in, 488–90; single-test placement, 490-4; stable matching in, 480-2, 492-3; student optimal deferred acceptance (SODA), 481-5, 486, 487, 490, 492-3; test scores and, 486, 487-91, 494 college optimal deferred acceptance (CODA), 483, 485-6, 487, 490, 492-4, 499 colleges and universities, 291; dormitory room allocation, 499-510; tenure and performance of professors, 174-6 collusion, 377 command-and-control regulation, 5, 143 common property, 556-60. See also public goods common values, 334-5, 377-9, 380 community benefit, 423, 556-7. See also social welfare community wealth, 132-3 compensation, 214, 249, 291; agency theory and, 232, 237, 239-40; golden parachutes, 220, 226; pay-for-performance, 177-9; performance bonus, 198, 215-16, 217, 222, 230; piece rates, 226, 236-7; retirement and, 166, 167, 170-1; stock options, 216-19, 224, 227. See also wages competition, 153, 178, 179, 339, 514-23; in insurance market, 130-3, 303-24; market responsiveness, 516-18; in product market, 215; reputation and, 6, 515-16; social benefits of, 50 competitive equilibrium, 338-9, 428-31, 433, 524, 528, 531-3; efficiency and, 543, 545-6; nonconvex economies and, 538-42. See also Arrow-Debreu economy; equilibrium composite commodity theorem, 98-102 compromise, 392-3 computers, 297, 340, 347 confessions, prisoner dilemma and, 45-6 constrained maximization, 75-6, 80-1, 86, 88.243-4 constrained optimization, 339-40 constraints, 10, 86-7, 88, 455; resource, 455; self-selection, 285, 286-7, 288, 296, 300, 315-16. See also budget constraint; participation constraint consumer, 5, 48, 287-9, 423, 520-1; franchise standards and, 244; surplus and, 108-12, 259; warranties and, 257, 515; welfare of,

11, 148, 153, 154

consumer choice/preference, 73, 87, 107, 285, 292; in auctions, 332-3; chain rule and, 86; maximization and, 76, 79; monopolists and, 280-3; nonnegative consumption and, 81-2; tangency and, 85 consumption plan, 139-42, 172, 528, 538-40, 545 continuum dilemma, 62-4, 122 continuum of effort supply, 190 contracts, 7, 47, 186, 213, 237, 340-1; doctor-hospital, 496, 498; insurance, 164, 165, 232, 303, 307-9, 310, 314-15; manager, 216, 218, 238, 239-44, 247-8; up-or-out, 174, 176; worker, 167, 208-9, 235-7 cooperation: in cartels, 49-50; equilibrium theories and, 46-7, 55, 56-8, 66, 69 copayment, 183 corporations: assets of, 212; hostile takeovers and, 50-1; international overview, 197-8. See also partnership; shareholders cost minimization, 521 Cowie, Tyler, 517 credit rationing, 272-80; borrower and, 274-7; lender and, 277-9 crime, 1, 14, 185-6, 422-3 Cuba, 199 currency devaluation, 49 Czechoslovakia, 516 Czech Republic, 329 dead hand pills, 223 decentralization, 34, 140, 225 decision making, 385; uncertainty and, 112-24; unilateral, 432, 440, 450. See also preference revelation; voting schemes deductibles, 183 deferred acceptance algorithm (DAA), 472-6, 479 demand, 80, 182; price-taking behavior and, 545-8, 554-6; quasi-linear utility and, 106-7, 109-11 democracy, 14, 402, 516 deposit insurance, 155-6, 158, 162 Depository Institutions Act (1982), 155, 158, 162 deregulation, 156-7, 329. See also regulation dictatorial rule, 391-2, 405, 412, 414, 416, 549 dilution provision, 220, 223 disarmament, 49 Walt Disney Corporation, 216, 223 diversification, benefits of, 231-2 division of a cake, 26-8, 35-6. See also allocations DNA testing, 14 doctors/physicians, 7, 182, 185, 535; matched with hospitals, 496-9 dominant strategy, 34, 419; equilibrium, 31-2, 260-6, 458-61, 544-5; in prisoner's

dilemma, 46, 47, 54; truthful revelation as, 345, 443, 446-7, 450, 458, 503, 506 dormitory room allocation, 499-510 Dresher, M., 51 drug-resistant microbes, 535 drunk driving, 8 Dutch auction, 329, 349, 350, 352-3, 356, 361, 374 DWG (firm), 223 Earth Day (Central Park), 432 Eastern Europe, 329, 516 East Germany, 214, 516 economic sanctions, 47-8 economies of scale, 10 The Economist, 9, 35, 46, 48, 184, 186, 223 education and training, 166, 291, 292-5, 296-7, 302. See also public schools; colleges and universities efficiency/efficient outcome, 23-30, 422, 535, 548, 559; in allocations, 26-8, 139-40; asset, 260, 262-3, 264, 333, 341, 342, 378, 379; in auctions, 329-33, 346; of competitive equilibrium, 321-2; dominant strategy equilibrium and, 458-61; economic system design, 519-20; fairness and, 24, 379, 489; incentive compatibility and, 457-66, 543-6; in insurance markets, 132-3, 134, 179; of majority rule, 389; mandatory retirement and, 166-7, 168; in matching, 468, 471; in public goods/projects, 424-8, 430, 442, 455-6, 457-66; quasi-linear utility and, 103-5; in room allocations, 501-2, 503-5, 507; Samuelson efficiency condition, 427-8, 430-1, 432, 434-5, 449, 459; takeovers and, 220. See also asset efficiency effort supply function, 238-9 Eisner, Michael J., 216, 223 electric utilities, 18, 144-5, 179-80, 281, 327, 329; arbitrage and, 227-8 Electrofax, 258 encryption, 257 English Crown, 430 English (oral) auction, 335, 349-51, 361, 374-5, 377, 383 Enron, 224, 227-8 Environmental Protection Agency, 49, 327-8 environments (scenarios), 548, 550 equilibrium, 30-40, 170, 268; Arrow-Debreu economy, 523-38; asymmetric information, 285-9, 302, 312-20; in auctions, 335-7, 353-6, 359, 363-4; centipede game, 37-40; dominant strategy, 31-2, 260-6, 458-61, 544-5; efficiency of, 321-2; full information, 283-4, 297-8, 309-12; insurance and, 131, 307-9; in labor market, 294-6, 297-302; in partnerships, 202; pollution rights and, 146, 147; pooling equilibrium (cont.) equilibria, 294-5, 298-9, 312, 314; repetition and, 53-70; in resource allocation, 139-42; separating equilibria, 294, 295-6, 299-302, 313, 316. See also competitive equilibrium; Nash equilibrium equity insurance, 181. See also fairness Europe, 182, 230, 327, 328, 372, 377 European Union, 214, 329 exchange economy, 524, 525, 529-31, 544-5; nonconvex, 538-42. See also two-person exchange expected monetary value (EMV), 112-13, 116, 117-19, 128-9, 249, 274 expected utility, 114, 249, 266, 500, 502 extensive form (of a game), 41 externalities, 421, 533-5 fairness, 24, 379, 488-90 fair odds, 120-1, 123-4, 128, 129-30, 193-4 famine, food distribution and, 516 Faroe Islands, 560 feasibility/feasible outcome, 25, 389, 463, 526-7, 530, 532. Federal Reserve, 158 fertilizer plant, 214 FHLBB, 158, 160 Field Museum of Natural History, 327 fire and panic, 45 fire insurance, 184–5 Firestone Tires, 227 first-price, sealed-bid auction, 349, 352-6, 374, 377, 381, 383; misrepresentation in, 365-7; revenue equivalence and, 359-60, 361, 362-3, 364 fishing, 556-9, 560 Flood, M., 51 Ford, Henry, 224-5 Ford Motor Company, 224-5 France, 197, 198, 329 franchises, 12, 218–19, 232, 243, 244, 341 fraud, 162, 224 free market economies, 518 free-rider problem, 220, 229 FSLIC (Federal Savings and Loan Corporation), 157, 159, 160 full information equilbrium, 283-4, 297-8, 309 - 12GAAP (generally accepted accounting

practices), 228 "gambling for resurrection," 157–61 GBM. *See* Groves bargaining mechanism General Motors, 225 General Tire, 225 Germany, 197, 198, 214, 328, 374 Gibbard-Satterthwaite theorem, 391, 396, 404–11, 422; general proof of, 411–18;

pivotal mechanism and, 450-3; strategy proofness of, 404-9, 419 Glass-Steagall Act (1999), 197 global maximum, 73-4, 83, 460 global warming, 12, 454 golden parachutes, 220, 226 Google, IPO of, 329 government, 288-9, 321-2, 448, 560; auction by, 326, 328, 330, 334; economic model and, 423, 430 government regulation. See regulation GPA (grade point average), 501, 557 Great Britain, 197, 198, 327, 328, 329 Greenland, 560 greenmail, 223 grim trigger strategy, 56-7, 59-60, 63-4 Groves bargaining mechanism (GBM), 261-3, 265-66, 453-7

H. J. Heinz corporation, 222 haircut prices, 258 Hammer, Armand, 224 WR Hambrecht & Co., 329 health care, 515 health insurance, 182-4, 185, 232 hidden action, 10, 13, 136-95, 236, 448; in corporations, 197; mandatory retirement, 165-74; pollution rights, 143-52; resource allocation, 139-42; savings and loan debacle, 155-63; telecommunications regulation, 152-5 hidden characteristics, 11-13, 137, 256-324, 331, 421, 514; in auctions, 326, 328; credit rationing, 272-80; price discrimination, 257-9; in product quality and bundling, 280–90; retirement and, 167; in two-person exchange, 259-69; used-car problem, 269-72. See also signaling, in job market; insurance markets hidden information, 11, 13, 137, 236, 257, 514, 536, 552. See also asymmetric information HMOs (health maintenance organizations), 185, 194 Hollywood film industry, 186 home mortgages, 155, 156, 164, 167 homeowners insurance, 180, 181, 185-6, 194 homestead exemption, 164-5 Hope Scholarships, 13 hospitals, 182, 183; doctors and, 496-9 household endowment, 526 household preferences, 423 House of Representatives, U.S., 390 housing construction, 515–16 Hurwicz-Schmeidler theorem, 549

IBM punch cards, 259 implementation, 551

impossibility theorem, 459 incentive compatibility, 141, 260, 263-4, 315; in auctions, 336, 343-4; constraints, 285, 296, 300, 458-9, 460, 461; efficiency and, 457-66, 543-6. See also self-selection constraints income, 106, 108, 172-3. See also compensation; wages India, fertilizer plant in, 214 indifference curve, 84-5, 253, 305, 519-20, 545-6, 550-1 individual preference, 9, 113, 385, 433; in partnership, 200; ranking of, 389, 394-6, 402. See also consumer choice; preference revelation individual rights, 469 individual security, 61 individual strategy, 41, 53 individual welfare, 2, 3-4; social cost pricing and, 4. See also self-interest Indonesia, 4 information: asymmetric, 10-15, 267, 272, 285-9, 297-8, 302, 312-20, 322; full information equilibrium, 283-4, 297-8, 309-12; inside, 4-5; transmission of, 2, 5-6, 141, 229. See also hidden information Innocence Project (Cardozo Law School), 14 innovation, 209, 535; competition and, 153, 516-18 insider trading, 221-2 insurance, 12, 14-15, 120, 124-35; asymmetric information and, 312-20, 322; automobile, 12, 162, 180, 183, 303-4; catastrophe, 517; competitive markets, 130-3, 303-24; contracts, 164, 165, 232, 303, 307-9, 310, 314-15; efficiency in, 132-3, 134; effort supply continuum, 189, 190-4; fair odds in, 121, 128, 129-30, 193-4; formal model, 186-8; full information, 309-12; health, 182-4, 185, 232; homeowners, 180, 181, 185-6, 194; incomplete, 192-3; individual variables and parameters, 187, 306; legal-defense, 14-15; model, 303-7; moral hazard and, 179–95, 312; policy, 308; preventive care and, 179, 180-1, 182, 187-8, 190; risk aversion and, 119 interdependent values, 377-83 interest rates, 165, 171, 517; credit rationing and, 272-3, 274, 275-8; savings and loan debacle and, 156, 157, 161, 163 Internet auctions, 328, 329, 349, 377 investment, 120. See also shareholders investment banks, 329, 333-4 invisible hand, 33-4 IPO (initial public offering), 329 Iraq, 9; economic sanctions against, 47-8 Israel, 34, 35 Italy, 224

Japan, 8, 197, 198, 230, 328 Johnson, E. Ross, 224 joint ventures, 36 Keating, Charles H., 163 kidney transplants, 328, 510-12 Korean conglomerates, 224 Kuhn-Tucker multiplier, 97, 339-40. See also Lagrangian multiplier labor costs, 535. See also wages labor market signaling, 5, 290-303; education and, 291, 292-5, 296-7, 302; equilibria and, 294-6, 297-302. See also workers Lagrangian multiplier, 86-97, 339-40; linear functions, 89-96; planning and, 87-8; Pythagorean theorem and, 91, 96-8 large numbers, law of, 306 Latin America, loans to, 158 legal-defense insurance, 14-15 legal profession, 175 legal system, 214-15. See also litigation leisure consumption, 7, 233, 243, 292, 526; opportunity costs and, 207-8; in partnerships, 202, 203, 205, 209; retirement and, 166, 168, 169 lender, credit and, 277-9. See also loans Leontief utility, 102, 550, 552 licensing fees, 218-19 limited liability corporation, 212 Lindahl equilibrium, 465 linear functions, 89-96. See also quasi-linear utility linear pricing schedule, 280 linear taxi fares, 16-18 litigation, 376; tort damages, 340-1 loans, 13, 158, 162; credit rationing, 272-80; home mortgages, 155, 156, 164, 167 loop, in algorithms, 506, 507, 511, 512 looting, 162 lynching of African Americans, 7 maintenance, 515-16 majority rule, 386-91, 397-9, 400, 402 manager, 328; contracts for, 216, 218, 238, 239-44, 247-8; incentive for, 216-19, 231-2; profit maximization and, 341; as residual claimant, 137, 207, 208, 218, 235, 244, 251, 341; risk aversion and, 231, 244-54. See also agency theory; CEOs; owner-manager relationship Mandarins, in China, 519 mandatory retirement, 165-74; formal argument, 167–71; intertemporal budget constraint, 171-3 manipulation, 404, 447-8, 469; of voting, 391, 392, 398-9

Mannesman (firm), 374

marginal rate of substitution (MRS), 84-5, 107, 139, 202, 305, 313-14, 339, 514; Arrow-Debreu economy and, 520-1, 529, 531; in nonconvex economies, 538-41 market: capital, 213, 219-21; completeness of, 531; discipline in, 229; mechanism, 448; opportunity in, 119-20, 127. See also Arrow-Debreu economy marriage model, 468, 469-70, 484. See also matching, stable matching, 467-512; college admissions, 480-96; deferred acceptance algorithym (DAA), 472-6, 479; dormitory room allocation, 499-510; hospital-doctor, 496-9; kidney transplants, 510-12; marriage model, 468, 469-70, 484; one-sided, 499; optimality theorem, 475, 485; preference ordering in, 475, 477-8, 482, 485, 493-94, 496, 506; stable, 470-1, 475, 476-7, 478, 480-2, 492-3, 496-7; student-advisor, 469-80 material feasibility condition, 526-7 maximization, 100, 450; constrained, 75-6, 80-1, 86, 88, 243; global, 73-4, 83, 460; of quadratic, 72-6; unconstrained, 72-5, 76-8. See also profit maximization maximum incentive, for managers, 231-2 medical malpractice, 212 Mexico, 158 minimization, 84, 521 misrepresentation, 365-7, 491, 497-8, 543-4, 547 monitoring costs, 210, 229 monopolies, 153, 329, 330, 334, 430; price discrimination and, 258-9, 280-1; profit maximization and, 282-3, 285-6; telecommunications, 517 monotonic preferences, 26 moral hazard, 13, 177-8; binary choice model, 188-90; formal model, 186-8; insurance and, 179-95, 312 movie producers, insurance and, 186 mutual funds, 230 naive mechanism, 18, 21 naked call options, 160 Nash equilibrium, 32-3, 36, 54-5, 355, 383, 432, 458, 548; benefit taxation and, 435; centipede game and, 38-9; exchange mechanism and, 266-8; grim trigger strategy and, 56-7, 63; in partnerships, 205–6; rationality assumption and, 39, 55, 64; subgame-perfect, 40-4, 60, 62, 549-53; Walker mechanism and, 461-5

National Education Association (NEA), 177, 178, 179

National Intern Matching Program (NIMP), 497, 499

nonconvex economies, 538-42 nonimposition, 409 normal form (of a game), 40 O'Brien, Ken, 251 oil industry, 47, 48, 157, 224, 226, 328; auctions in, 378, 380, 381 opportunity costs, 207-8, 521 optimality theorem, 475, 484, 485. See also efficiency optimization, constrained, 339-40 oral auction. See English (oral) auction owner-employee relationship, 199, 207-12 owner-manager relationship, 212-31, 331, 559-60; capital market and, 213, 219-21; manager discipline in, 213-14; managerial shirking, 222-7; performance bonuses, 215–16, 217, 222, 230. See also agency theory; managers ownership shares, 36, 197-8. See also shareholders

participation constraint, 260, 262, 265–6, 282, 286, 461; in auctions, 342, 345–6; dominant strategy equilibrium and, 458–9, 460; in room assignment, 500 partnership, 174, 198–207, 209, 212, 228; dissolving of, 35–7, 265; leisure consumption in, 202, 203, 205, 209; model, 199–200; reputation and interactions, 204–5; two-person, 202–4 pay-for-performance, teachers', 177–9 performance bonus, 198, 215–16, 217, 222, 230 physicians. *See* doctors/physicians

piece rates, 226, 236-7

pivotal mechanism, 341, 440–53; continuum of options, 449–50; defects of, 446–9; Gibbard-Satterthwaite theorem and, 450–3; surtax imposed by, 442, 443–7; theorem proof, 444–6

plurality rule, 393–4

poison pill, 223

Poland, 328, 516

Polaroid Corporation, 259

policy options, 402-3

pollution, 4, 18–22, 48–9, 516, 535; CO₂ emissions, 12, 148, 421; marketable rights, 143–54; profit maximization and, 146–50 pollution permits, 327–8

pooling equilibrium, 294–5, 298–9, 312, 314

Posner, Victor, 223

power sharing, 392

preference revelation, 396–418, 423, 461; dictatorial rule, 391–2, 405, 412, 414, 416, 549; in general, 402–11; Gibbard-Satterthwaite theorem, 404–11; principle of, 418–19; social choice rule,

403-4, 405, 410, 418, 419; strategy proofness and, 404-6, 413, 415, 419; unanimity lemma, 406-7, 412, 413; value-restricted, 396-9. See also truthful revelation preferences: in matching problems, 475, 477, 482, 485, 493-4, 496, 506; misrepresentation of, 497-8, 543-4, 547; ordering of, 389, 394-6; quasi-linear, 102-12, 424, 426, 450-1, 458; self-regarding, 26, 531; voting schemes, 385-401. See also consumer choice/preferences President's Council of Economic Advisors, 157 preventive care, 179, 180-1, 182, 187-8, 190 price discrimination, 257-9, 280, 281 prices, 5, 524, 544, 547; cap regulation, 153-5; competition, 156; manipulation of, 514; stock repricing, 216-17. See also social cost pricing price taking, 545-8, 554-6 principal-agent problem, 6, 10-11, 12-13, 17, 137, 176, 232-3. See also agency theory prisoner's dilemma, 32, 33, 42, 45-53, 161-2, 432; examples of, 47–51, 64–9; infinitely-repeated, 55-6, 60; repeated, with terminal date, 54-5 private goods, 142, 404, 536; public goods and, 421-2, 424, 428, 429, 441, 449, 460 private information, 11, 448. See also hidden information private ownership, 339, 521, 543 private values, 334-5, 378-9 privatization, 328-9 probability, 121-2, 306, 447; in auctions, 354-5, 363, 382; insurance and, 187, 313 procurement auction, 376, 377 production, 525, 535 productivity, 176; education and, 293, 295, 296 - 7product markets, 214, 215 professionals, 177-8, 202, 212 professors: research and, 174, 176, 177; as student advisors, 468, 469-79; tenure and, 174 - 6profit maximization, 282-3, 315, 515, 521-2; agency theory and, 233-4, 239-40, 249-50; feasibility and, 527; information equilibrium and, 283, 285-7; owner-manager relations and, 213, 219, 222, 226, 230, 341; pollution abatement and, 146-50 progressive piece rates, 236-7 property rights, 518, 560 public goods/projects, 9, 420-66, 536; auction of, 328-9; competitive market equilibrium, 428-31; continuum of, 423-4;

economic model and, 422–40; efficiency in, 424–8, 430, 442, 455–6, 457–65; Groves mechanism and, 453–7; pivotal mechanism and, 440–53; private goods and, 421–2, 424, 428, 429, 441, 449, 536; pure, 421–2, 536; Samuelson efficiency condition and, 427–8, 430–1, 432, 435; taxation and, 432–6, 442, 443–7, 453–4, 456, 464–5; voluntary contributions, 431–2 public opinion, 48 public school, 177–9, 390, 494 publishing, 176, 287 purchasing power, 172 Pythagorean theorem, 91, 96–8 quadratic, maximizing of, 72–6

quadratic profit functions, 149-50 quality, 6, 175, 244, 290, 292, 515; bundling and, 280-90; information equilibrium and, 283-9; self-selection constraints and, 285, 286-7, 288; in used-car market, 270, 271 quasi-linear preferences, 102–12, 342, 424, 426, 450-1, 458; consumer surplus and, 108-12; demand and, 106-7; efficiency and, 103-5 quotas, for fishing, 560 radio spectrum auction, 327, 333, 372, 374, 375, 377 RAND corporation, 183 random effects, 6, 137-8, 502 rank order voting, 394-6 rate of technical substitution (RTS), 521 rationality assumption, 39, 55, 64 rationing, 141 regulation, 4-5, 33, 516, 559; deregulation, 156-7, 329; managerial, 213-14; of savings and loans, 156-7, 158, 161-2; of telecommunications industry, 152-5 repetition, equilibrium and, 53-70 reputation, 6, 204-5, 515-16 research, 174, 176, 177 research and development (R&D), 198, 225, 230 reservation values, 260-4, 269, 270, 326, 330-2, 350-1; private vs. common values, 334-5; revenue equivalence and, 361-3, 365, 366-70; in Vickrey auctions, 334-5, 336, 342, 440-1 residual claimant, manager as, 137, 207, 208, 218, 235, 244, 251, 341 resource allocation, 12, 139-42, 454-5, 550; in Vickrey auction, 338-9 resources: common property, 556-60; utilization, 89 responsibility system, 7

retaliation strategy, 34–5

retirement. See mandatory retirement

revelation principle, 418-19. See also preference revelation: truthful revelation revenue equivalence, in auctions, 358-77, 379-80; applications of, 374-7; equilibrium payoffs, 363-4; integral calculus proof, 371-2; proof of, 365-71; theorem, 359-60, 380 risk, 14; degree of, 274; excessive, 226; insurance and, 183-4, 303-4, 312. See also insurance risk aversion, 117-18, 125, 212, 305; in managers, 231, 244-51; negative second derivative and, 118-19 risk neutrality, 117-18, 137; managerial, 235, 239-41, 242, 243, 244, 252-4 risky assets, 113 risky investments, 157-61 RJR Nabisco, 224 room assignment problems, 499-510 royalty payments, 244 Russia, 224, 329, 516 Safelite Glass Corporation, 226 sales receipts, 14 salmon fishing, 560 Salomon Brothers (firm), 225 Samuelson efficiency condition, 427-8, 430-1, 432, 434-5, 449, 459 savings and loan crisis, 155-63; risky investments and, 157-61; zombie institutions, 157-8, 160 Scholes, Myron, 225 schools. See education and training; public schools sealed-bid auction, 121, 349, 353-6, 359, 363, 374, 377, 379-80. See also first price, sealed bid auction second-order condition, 426 second-price auction. See Vickrey auction Securities and Exchange Commission, U.S., 221 self-dealing, by managers, 223-5 self-interest, 2, 6-7, 8, 33-4, 331, 432; prisoner's dilemma and, 32, 46; in shirking, 3,17 self-regarding preferences, 26, 531 self-selection constraints, 285, 286-7, 288, 296, 300, 315-16 separating equilibrium, 294, 295-6, 299-302, 313, 316 serial choice: with guarantee (SCG), 504-10, 511, 512; with no guarantee (SCNG), 500-4 serial choice algorithm (SCA), 487, 488, 490 - 4sharecropping, 137-8 shareholders, 197-8, 212-13, 225, 543; dilution provision and, 220; insider trading and, 221-2; managers and, 214, 215, 217,

218, 232; shirking and, 224, 228-30; takeovers and, 50-1, 222 shirking, 3, 16, 17, 138, 166-7, 212, 244; managerial, 222-30; by workers, 236 signaling, 270; in auctions, 378, 380; education and, 291, 292-5, 296-7, 302; in job market, 290-303; wages and, 5, 291, 294, 299-302 Silicon Valley (California), 209 single-test placement, 490-4 Sloane, Alfred P., 225 Smith, Adam, 33, 43 social benefits, 186, 535. See also public goods; social welfare social choice rule, 402, 403-4, 410, 418, 451, 551, 553; strategy-proof, 404, 405, 414, 419 social cost, 182, 183, 291 social cost pricing, 4, 7, 89, 96, 166, 534–5; Groves mechanisms and, 453, 456; public goods and, 440, 446, 449, 514; in Vickrey auction, 337-41, 345 social welfare, 12, 50, 96, 142, 161, 558 Southwestern Bell, 333 Soviet Union, 6, 222, 329, 516. See also Russia Spain, 327 spectrum auction. See radio spectrum auction spillover benefit, 2, 421, 428, 431, 448, 535 sports, 175, 251, 340, 518; baseball, 12, 40, 378, 381 stable matching. See under student matching Stalin, Joseph, 7 standard auction mechanism, 359, 375-6 standards, in franchises, 244 status quo, in voting schemes, 390-1, 402, 454 stock options, 216-19, 224, 227, 228 strategy, 34-5, 40, 41, 53; grim trigger, 56-7, 59-60; tit-for-tat, 55, 58-9, 64-5, 68. See also dominant strategy strategy proofness, 404-9, 413, 415, 419, 476; in room allocations, 503, 506 strictly concave functions, 82-4 student loans program, 13 student matching: acceptability in, 472-6, 480; college admissions, 480-96; dormitory room allocation, 499-510; marriage model and, 469-70; placement model, 486-7; stable matching, 470–1, 475, 476–7, 478, 480-2, 492-3, 496-7; student-advisor, 469-79. See also colleges and universities student optimal deferred acceptance (SODA), 481-5, 486, 487, 490, 492-3, 499 subgame, 42. See also Nash equilibrium, sub-game perfect suicide, insurance and, 181, 312 sulphur dioxide (SO₂), 18-22, 144-52, 327-8. See also pollution

surtax, 442, 443–7, 453–4, 456 Sweden, 327 Switzerland, 327

take-it-or-leave-it offer, 267-8, 284, 327 takeovers, 214, 219-20, 560; hostile, 50-1, 213, 219, 222-3, 229 tangency approach, 84-5 taxation, 15, 243, 341, 430; average cost, 432-3, 464-5; benefit, 433-6, 446; Groves mechanism and, 453-4, 456; pivotal mechanism surtax, 442, 443-7 taxis, 16-18, 517 teachers: college professors, 174-6, 468, 469-79; in public schools, 177-9 teamwork, 199 technology: innovation and, 209, 298; investment in, 535; standards for, 375-6 telecommunications, 152-5, 333, 375, 517 television, 6, 144, 378, 516, 518 tenure, for professors, 174-6 terrorists, harboring of, 35 tit-for-tat strategy, 55, 58-9, 64-5, 68 T-Mobil, 374 tort damages, 340-1 total derivative, 85 trade embargo, 47-8 transfer payments, 461, 462-3 T-Rex skeleton auction, 327 truthful revelation, 9, 261-2, 398, 402, 422, 440; in auctions, 326; as dominant strategy, 345, 443, 446-7, 450, 458, 503, 506; in matching problems, 478, 485, 503; tax threat and, 434-5, 445, 453; Vickrey mechanism and, 19-21; in voting, 391. See also preference revelation Turkey, 487 two-person exchange, 259-69; dominant strategy equilibrium and, 260-6; Groves bargaining mechanism, 261-3, 265-6 unanimity lemma, 406-7, 412, 413 uncertainty, 138; decision making and, 112-24 unemployment insurance, 13-14 uniform probability distribution, 121-2, 354-5 unilateral decisiveness, 432, 440, 450 unique majority winner, 397-9 United Kingdom, 197, 226, 499 United Nations, Iraq and, 47, 48 United States, 34, 197-8, 214, 215, 230, 329; Congress of, 156, 158, 161-2; Defense Department, 143; Department of Interior, 378; health care in, 182, 185, 194; House of

Representatives, 390; takeovers in, 220

universities. See colleges and universities up-or-out contract, 174, 176 used-car market, 269-72 utility function, 100, 111, 503, 521, 522. See also expected utility; quasi-linear perferences utility-of-wealth function, 113, 114, 118, 122-3, 305; insurance and, 125, 126-7, 128-9; marginal, 116, 117 value-restricted preferences, 36-7, 396-9 values, interdependent, 377-83 veto power, 393 Vickrey auction, 334-49; constrained optimization in, 339-40; direct auction mechanism, 342, 345; English auction compared, 350-1; equilibrium bids in, 335-7; resource allocation in, 338-9; revenue equivalence in, 361, 362-3; social cost pricing and, 337-41, 345, 440-1; uniqueness of, 342-5 Vickrey mechanism, 19–21, 22 Volkswagen, 287, 328 voluntary contributions, 431-2 voting schemes, 385-401; compromise in, 392-3; dictatorial, 391-2, 405; majority rule, 386–91, 397–9, 400, 402; manipulation in, 391, 392, 398-9; plurality, 393-4; power sharing, 392; rank order, 394-6; status quo in, 390-1; variable numbers in, 387-92. See also preference revelation voucher auction, 329 voucher system, for schools, 178-79 wages, 166, 209, 226, 535; signaling and, 5,

291, 294, 297-302. See also compensation Walker mechanism, 461-5 Wall Street, 227, 228 warranty, 5-6, 138, 257, 515-16 Washington Post, 13 wealth, 187, 228, 304. See also utility-of-wealth function welfare theorem, 529-31, 534 Western Europe, 214. See also Europe winner's curse, in auctions, 380-3 worker contracts, 167, 208-9, 235-7 workers, 235-7; job-market signaling, 290-303; mandatory retirement, 165-74; motivation of, 226-7; owner-employee relationship, 207-12 World Health Organization, 535

Xerox Corporation, 258, 259, 340

zero-profit conditions, 308–9, 312, 313, 316 zeroth good, 98